GIMIS: Gaze Input with Motor Imagery Selection

Baosheng James Hou
Technical University of Denmark
Lyngby, Denmark

Per Bækgaard
Technical University of Denmark
Lyngby, Denmark

I. Scott MacKenzie
York University
Toronto, Canada

John Paulin Hansen
Technical University of Denmark
Lyngby, Denmark

Sadasivan Puthusserypady
Technical University of Denmark
Lyngby, Denmark

ABSTRACT

A hybrid gaze and brain-computer interface (BCI) was developed to accomplish target selection in a Fitts' law experiment. The method, GIMIS, uses gaze input to steer the computer cursor for target pointing and motor imagery (MI) via the BCI to execute a click for target selection. An experiment (n = 15) compared three motor imagery selection methods: using the left-hand only, using the legs, and using either the left-hand or legs. The latter selection method ("either") had the highest throughput (0.59 bps), the fastest selection time (2650 ms), and an error rate of 14.6%. Pupil size significantly increased with increased target width. We recommend the use of large targets, which significantly reduced error rate, and the "either" option for BCI selection, which significantly increased throughput. BCI selection is slower compared to dwell time selection, but if gaze control is deteriorating, for example in a late stage of the ALS disease, GIMIS may be a way to gradually introduce BCI.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**.

KEYWORDS

Fitts' law, brain-computer interaction, gaze interaction, Hybrid BCI, amyotrophic lateral sclerosis, augmentative and alternative communication, pupillometry, pupil size,

1 INTRODUCTION

Our sensory and motor functions are generated and communicated by a complex, essential, yet fragile, central nervous system (CNS). Spinal cord injuries, strokes, and autoimmune diseases can damage

the CNS and impair or disable speech and movement abilities, and, in severe cases, lead to locked-in syndrome and tetraplegia. For example, the adult-onset motor neuron disease, amyotrophic lateral sclerosis (ALS), causes patients to gradually lose the ability to move, speak, swallow, and breathe. Globally, there are over 200,000 cases of ALS in 2015, and this is expected to increase by 69% to over 300,000 by 2040, mostly due to an aging population [Arthur et al. 2016].

Augmentative and alternative communication (AAC) technologies can help these patients communicate with their environments and improve social interaction and self-esteem. AAC technologies used by people with ALS include printed alphabets on a transparent sheet, eye tracking, brain-computer-interfaces (BCI), electromyography (EMG), electrooculography (EOG), etc. Eye tracking is easy and fast to use, but it requires well-functioning control of eye movements. BCIs can be used without moving the eyes, but the interaction is slower and more tiring than eye tracking. EMG can help patients who still have motor control, but cannot help severely paralysed patients who lack even basic control of muscle movements. Both eye tracking and BCI are accepted within the ALS patient community, with BCI particularly welcomed in the late stages of ALS when all movement control is lost.

In this paper we examine a hybrid system as a way to ease the transition from eye tracking to a BCI system. There are potential applications in the late stages of ALS where eye control is becoming weaker but when a full BCI system is not yet required. Also, compared to standard dwell-time target selection, as used by most eye tracking systems, a BCI click function for target selection has potential advantages. A BCI selection is an intentional act, while dwell selection can occur inadvertently just when looking. This is known as the "Midas touch" problem [Jacob 1991]. As well, different types of BCI clicks can be mapped to different actions, such as single click, double click, left click, click-and-drag, and so on. It thus is important to know how the two systems will work when combined in a hybrid setup: What throughput should we expect? What are the error-rates? How mentally demanding is the interaction? Can BCI support different types of clicks when used with eye tracking?

Gaze from an eye tracker is a fast replacement for cursor pointing, while selection is usually activated using a dwell time [Mateo et al. 2008]. However, most dwell-based systems do not allow users to rest their gaze, and the duration of dwell fixations needs to be optimized for different tasks. This causes unintentional activations which can be annoying. A hypothesis is that a hybrid system can mitigate the Midas touch problem associated with dwell-based systems. Previous studies combining eye tracking with motor imagery (MI), EMG, and mouse input have been successful (e.g., [Mateo et al. 2008]);

however, not all studies use the same experimental protocols, tasks and conditions, and this makes comparisons between the systems difficult.

The efficiency of new methods for pointing and selecting is commonly evaluated using a Fitts' law experiment for target acquisition. The next section reviews some of work in the field of eye tracking and hybrid brain-computer interfaces. Our Fitts' law experiment is then described.

2 RELATED WORK

Eye tracking has been used by people with ALS to access the Internet, to use social networking, and to communicate by texting and voice synthesis. Eye tracking is welcomed by ALS patients due to its non-invasiveness, ease of set up and use, and the availability of customer support. A successful eye tracking system achieves spatial and temporal accuracy and is tolerant to changes in ambient light levels and to patient use of glasses and contact lenses. Other issues to consider are reliability, robustness, the comfort of the mounted device, freedom to move the head, portability, battery life, wireless use, camera angles, etc. Sometimes eye tracking systems are bulky or noisy, and the customer may not be satisfied by the speed of interaction, such as when selecting a letter or word in a spell-checker. Users can also get tired due to gaze fatigue [Majaranta and Bulling 2014; Spataro et al. 2014].

The most common eye tracking technique used for communication is videooculography (VOG) with infrared radiation. The pupil position is computed and mapped to a cursor position on the computer screen; however, such systems are sensitive to light conditions and have limited temporal resolution [Lupu et al. 2013]. Electrooculography (EOG) is more tolerant of light conditions and can measure eye activities even when the user is squinting or laughing, but is less comfortable to wear for everyday use. VOG and EOG have been combined to work in challenging outdoor environments. Eye tracking data can also assist in determining the user's cognitive and emotional state, by combining measurements, such as gaze patterns, pupil size, and microsaccades, with other physiological measurements, such as facial expressions and heart rate [Majaranta and Bulling 2014].

The main downside of dwell-based gaze selection is the Midas touch problem. Normally, a selection is activated if the user fixates on a selectable target for pre-determined threshold of time. However, an unintentional selection may be triggered, for example, if the user overly extends a gaze fixation due to fatigue or prolonged mental processing of visual information. This causes frustration: Users cannot rest their gaze "just anywhere", and may need to adopt an avoidance strategy, such as deliberately looking away from a point of interest. This can be annoying and can reduce reading speed [Zander et al. 2010].

A solution to the Midas touch problem is to use another channel of information for selection. In one example, gaze was coupled with mouse and facial EMG in a point-and-select experiment [Mateo et al. 2008]. Pointing methods (gaze and mouse), and selection methods (mouse and EMG) were compared. Gaze pointing was faster than mouse pointing, but no significant difference between the selection methods was reported. For paralyzed users who cannot control a mouse or facial muscles, BCI could serve as the selection method.

The combination of BCI with eye tracking (or EMG, a mouse, etc.) is an example of a hybrid system. Hybrids have the advantage that the combinations complement each other's strengths and limitations, and can increase accuracy, specificity, sensitivity, and may overcome the zero-class problem (a non-intentional condition of 'rest', during which the user does not wish to send a command) [Rupp 2014].

Hybrid systems involving eye tracking have been evaluated and shown some feasibility. Käthner et al. [Käthner et al. 2015] compared corneal reflection eye tracking, EOG, and an auditory BCI in terms of ease of use, speed, and comfort by ALS patients. It was reported that the BCI system was the easiest to use, followed by eye tracking, then EOG. However, eye tracking was fastest and least tiring, compared to EOG and BCI. If the speed and comfort of eye tracking can be combined with the ease of BCI, such a hybrid system, by combining and complementing the advantages of eye tracking and BCI, has the potential to improve the patient experience [Käthner et al. 2015].

Pfurtscheller et al. reported that 90% of their subjects preferred a gaze-BCI hybrid over dwell-time alone [Pfurtscheller et al. 2010]. Although the hybrid was slower, it was more accurate in difficult tasks. Hybrid BCIs can improve accuracy and reduce false activation if the systems are used to confirm each other – a command is only issued if both systems are activated. In Vilimek and Zander's work [Vilimek and Zander 2009], a hybrid gaze-BCI system was built to improve a word selector. A word was selected only if both an eye tracking fixation and a BCI hand movement command were detected. In the latter case, the "movement" was detected through motor imagery signal detection in the BCI. The hybrid system was slower than dwell-time alone, but had higher user preference and did not increase cognitive demand significantly. The challenges in using MI in a BCI are to filter out the noise caused by eye movements and to reduce the time required to set up and train the BCI binary classifier for the MI signals.

Lim et al. [Lim et al. 2015] combined eye tracking with a steady-state visually-evoked potential (SSVEP) to improve the accuracy and speed (due to fewer corrections) of a speller. Letters on the speller keyboard flashed at different frequencies, and template-matching eye tracking was used to detect which part of the keyboard (left, right, middle) the user was looking at. A letter was selected only if the SSVEP frequency matched the gaze location. The hybrid improved the accuracy by avoiding an average of 16.6 typos per 68 characters.

Zander et al. [Zander et al. 2010] present a gaze-MI hybrid. The method combined gaze with MI signal detection for "selection". This was compared to long and short dwell-time selection in a search-and-select experiment. It was found that the hybrid method was slowest, but the authors argued that the difference was not of practically significance. The gaze-MI hybrid had accuracy >78%, and was as accurate as long dwell-time selection and more accurate than short-dwell time selection, especially in more difficult tasks. The overall task load index (TLX) results showed no difference in workload between selection techniques, and a strong preference for selection using gaze combined with MI, with significantly lower frustration than either dwell-time selection method. It was not reported how long it took to train users to select using MI signal detection.

Another example of a gaze-MI hybrid is the control of a mechanical arm described by Frisoli et al. [Frisoli et al. 2012]. Test subjects used an eye tracker to choose a color-coded target to reach. A mechanical arm was activated using right-arm MI, with the movement of the mechanical arm automatically controlled. The entire scene was tracked using a Kinect-based vision system. The system had about 89% accuracy, however, the speed was not reported.

Finally, Lee et al. [Lee et al. 2010] describe a gaze-MI hybrid combing eye tracking and MI in a system for 3D selection. Because tracking a single eye provides no depth information, an imaginary arm reaching movement was used to access the depth axis. A hand imagery grabbing movement was used to confirm and activate the selection. Concentration was measured using pupil accommodation speed, which was used as a switch between gaze and motor imagery modes, thus reducing the BCI training time from 5 minutes to 30 seconds. The results showed satisfactory accuracy in distinguishing grabbing and reaching movements.

From previous studies we have seen that gaze is well suited to pointing and that adding a MI BCI is intuitive and offers more freedom to users when selecting. Most studies focused on BCI's application, and only a few reported throughout using a Fitts' law experiment (e.g., [Kim and Jo 2015; Nappenfeld and Gießing 2018]). This makes it difficult to compare BCI with other HCI paradigms. Therefore, our experiment focused on implementing and evaluating a gaze-MI hybrid for pointing at targets using gaze and selecting the targets using MI. We call the method GIMIS, for Gaze Input with Motor Imagery Selection.

3 METHOD

We conducted a study to evaluate target acquisition on a computer screen using GIMIS – our hybrid system that combines eye tracking (i.e., gaze) for pointing and a BCI using MI for selection. Normally, when we select a target on a computer screen, there are two steps: we first steer the mouse cursor over the target (pointing), then we click the target (selection). For disabled users, mouse control may not be possible. Gaze can be used as an alternative to steer the mouse cursor to acquire the target. A click for selection is then executed when a brain signal generated by imagining a specific movement is detected by the electroencephalogram (EEG) classifier in the BCI.

A target acquisition (point-and-select) experiment was used to test GIMIS. The computer cursor was positioned by gaze through the eye tracking system. This was combined with three motor imagery selection methods:

- Left hand (LH) – activate the click selection by imagining closing and opening the left fist.
- Legs – activate the click selection by imagining lower legs extension.
- Either – activate the click selection by imagining either LH or legs extension. The user is free to choose which motor imagery to use and can switch between them at anytime.

Tang et al. reported that the left hand vs. both feet MI paradigm was more accurately classified than the left hand vs. right hand paradigm [Tang et al. 2016], therefore we used left hand and legs for MI. Other combinations of MI were not investigated.
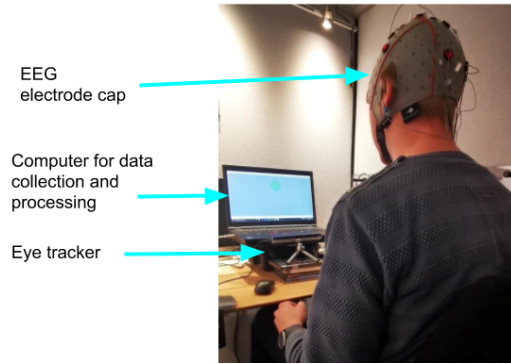

Figure 1: Subject wearing the BCI with EEG sensors siting on a chair with limbs resting on the armrests. The eye tracker is seen below the computer screen which displays the Fitts' law experiment task.

We used a Fitts' law testing protocol, as this is widely used to evaluate interaction involving rapidly pointing to targets and selecting them [MacKenzie 2013]. The target acquisition time (AT) is the time to acquire a target; it includes the time to move the cursor to the target and the time to select the target.

3.1 Participants

A total of 19 able-bodied participants were recruited from the local university and student dormitory. Fifteen completed the experiment. (Four subjects left early or cancelled due to defective EEG electrodes.) Of the 15 participants who completed the experiment, three were female. Ages were between 22-34 years with a mean of 25 years. Three of the participants were left-handed; four wore vision correction lenses. Six had participated in previous studies collecting brain data, and six had previously participated in eye tracking studies. A 400 DKK gift card was given for completing the experiment. All participants were informed about the procedures of the experiment and its purpose, and all participants signed a waiver form for informed consent.

3.2 Apparatus

EEG data were collected using the g.tec USBamp amplifier which samples at 512 Hz. The 12 AgCl electrode channels used were FC1, C3, C1, CP1, FCz, Cz, FC2, C2, C4, CP2, CPz, Pz. Conductive gel was used to reduce the impedance; all subjects had impedance lower than 4 kΩ.

The computer used to process the signal data had an Intel(R) Xeon(R) E_2186M CPU @ 2.90 GHz processor, 31.7 GB of usable RAM, and ran the Windows 10 operating system. The resolution of the screen was 3840 × 2160 pixels. The Fitts' law experiment used the GoFitts software[1].

A picture of the subject wearing the setup is shown in Figure 1.

The eye tracker used was developed by The Eye Tribe[2], and is placed below and approximately 25 cm away from the computer screen, as suggested by the developers. The eye tracker sampling

---

[1] http://www.yorku.ca/mack/FittsLawSoftware/
[2] https://theeyetribe.com

rate was 30 Hz. The subjects were seated 50-80 cm in front of the computer screen. The eye tracker is reported to have an accuracy of 0.5° to 1° visual angle [eye 2014]. All participants were calibrated using a 16-point calibration with the manufacturer's software, achieving > 3 stars accuracy. The accuracy data from vendors should be viewed as approximate, and under ideal conditions. The EyeTribe device's performance in fixation and pupilometry analysis was reported to be comparable to EyeLink 1000 [Dalmaijer 2014].

The eye tracker recorded the coordinate of the gaze and also the pupil size. Pupil size is an indication of cognitive load and has been found to increase with increased mental demand [Richer and Beatty 1985]. Mental demand, in the present setup, could depend on the task difficulty, in terms of target amplitude or target width, or the selection method. To investigate the effect of task difficulty and selection method on pupil size, the pupil size was recorded under a constant light intensity. When the colors are similar in brightness, pupil size will not be affected when the subject looks at the target, or when it changes color as feedback. Constant and equal lightness (relative luminance) values were used throughout the experiment for the target color, mouse-over color, target border color, and background color. See Table 1.

Table 1: The colors used in the Fitts' law task had approximately the same lux values.

	background	target border	target (unclicked)	button-down	mouse-over
LAB	50,0,0	50,0,56	50,0,56	50,-14,15	50,-14,15
sRGB	119,119,119	142,117,0	142,117,0	105,125,94	105,125,94
Illuminance (lux) near the eye	6.6	6.9	6.9	7.1	7.1
Illuminance (lux) 30cm from screen	22.0	22.2	22.2	21.3	21.3

The equiluminance was confirmed with an RS PRO IM720 light meter. Illuminance was measured at two locations, next to the eye and 30 cm away from the computer screen. The color displayed on the screen was the only light source. The experiment was conducted in a room without windows, thus ensuring the ambient light was fully controlled and constant.

3.3 Procedure

There were two parts to the experiment. First, there was a session where the EEG data from the BCI were acquired, processed, and used to train the classifier. The second part was to undertake the target selection task using the trained system with the Fitts' law experiment software.

A questionnaire was completed before testing and again after testing. The participant sat in a chair and was asked to rest their limbs comfortably on the armrest or table. During the experiment, they were asked to limit movement for the quality of the EEG signal and so the eye tracking remained stable.

3.3.1 Collecting training data for BCI. The computer provided cues for the participant to imagine hand or leg movement or to rest. See Figure 2. The computer screen shows a cue in a left-to-right sequence. The initial white cross indicates that attention is needed. It appears for 1 second then a picture of either the left hand or lower legs appears for 4 seconds, during which the participant imagines the corresponding motor image. This is followed by a black screen
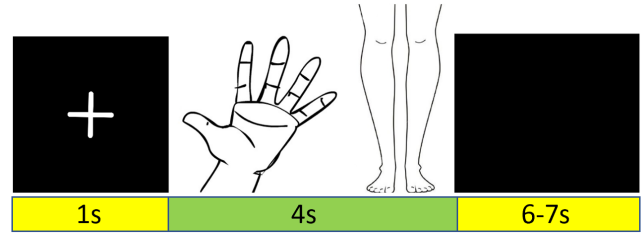

Figure 2: EEG data collection cues, shown in sequence from left to right. Each cue's timing information is labeled. Only one of the images (hand or legs) is shown in each sequence.

that prompts the user to stop imagining and become idle (i.e., rest). After 6-7 seconds, the cross reappears and the next trial starts.

A total of seven runs were conducted, each run consisting of 24 motor imagery trials appearing in random order, 12 for the left hand and 12 for the legs. For evaluation of the BCI accuracy, five of the runs were used for training and the rest were used for evaluation. The training session took about 90 minutes.

An variation of the filter bank common spatial pattern (FBCSP) paired with the naive Bayesian Parzen window (NBPW) algorithm [Ang et al. 2012] was used to process and classify the EEG signal. After successful training, the user can use imagined left-hand, legs, or either (left-hand or legs) movements to activate a mouse click and thereby select a target.

3.3.2 The Fitts' law experiment. Using the GoFitts software, we tested three motor imagery selection methods: left hand, legs, left hand or legs ("either"). These were combined with two target amplitudes, A (200, 400 pixels) and two target widths, W (100, 200 pixels). Spatial hysteresis of the target was set to 4: Once the cursor entered a target, the target was acquired until the cursor exited a region 4× the width of the target. Five targets were presented per sequence, there are four sequences per block (four combinations of amplitude and width) and three blocks (selection methods). The A-W combinations appeared in random order.

The order of the selection methods was assigned to the participant using a Latin square design. The number of targets was limited to five in order to keep the duration of the experiment comfortable. The experiment itself (not including the training period) took approximately 30 minutes per participant.

An example screenshot of the experiment task is shown in Figure 3. The target is the yellow circle, see Figure 3a; each new trial has a new target on the opposite side and then the target advances clockwise by one position. Each target has width W and the distance between the center of the current target and the next target is the amplitude A. The participant steers the cursor to the target using gaze. Once the cursor enters the target, visual feedback is given by turning the target green. See Figure 3b. With the spatial hysteresis, the target remains green even if the gaze point moves outside the target circle but is within 4× the target width. A click outside the circle was registered as an error. Audio feedback was given to the participant with different sounds signalling an error or correct selection.

(a) The yellow circle is the target for the current trial.

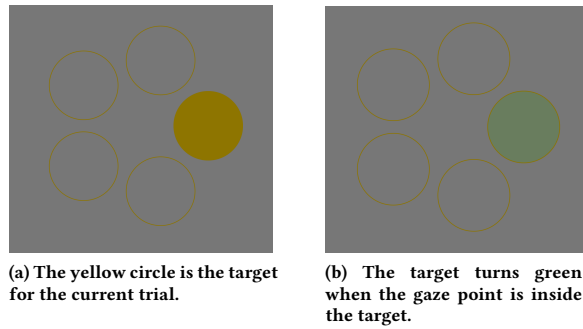(b) The target turns green when the gaze point is inside the target.

Figure 3: Target acquisition task interface. The yellow circle is the target for the current trial. The target turns green when the gaze point is inside the target. For this example, A = 400 pixels and W = 200 pixels.

3.4 Design

The experiment used a $3 \times 2 \times 2$ within-subjects design. The independent variables and levels were as follows:

- Selection method (left hand, legs, left hand or legs)
- Target amplitude (A = 200, 400 pixels)
- Target width (W = 100, 200 pixels)

The following dependant variables were used:

- Target acquisition time (AT, ms)
- Throughput (TP, bits/s)
- Error rate (%)
- Pupil size (PS, pixels)

Note that target acquisition time has two components, the movement time (MT) to acquire the target via gaze and the selection time (ST) to effect a click selection via the BCI. Selection time began when the cursor first entered the target. No trials had an acquisition time (AT) greater than 2 standard deviations from the sequence mean.

The pupil size is a measurement of cognitive load. The pupil dilates with higher cognitive loads, coincident with more complex tasks or greater movement planning [Jiang et al. 2014, 2015; Richer and Beatty 1985]. On the other hand, pupil size may decrease when the demand of precision dominates the task [Fletcher et al. 2017]; see also [Bækgaard et al. 2019]. Pupil data were normalized to each participant's baseline; that is, the mean of all the blocks of the experiment. Data points with either eye's pupil size ≤ 0 pixels were omitted from calculation of this baseline. Furthermore, data outside ±1 standard deviation of the mean were omitted from calculation of the baseline.

4 RESULTS

A three-factor repeated-measures ANOVA was used to investigate the effects of the independent variables on the dependent variables: acquisition time (AT), throughput (TP), error rate (ER) and pupil size (PS). The independent variables were selection method, target amplitude, and target width. After the experiment, the participants were given a questionnaire to rate their mental and physical state and their preference for the selection methods.

4.1 Target acquisition time

The grand mean for target acquisition time (AT) was 4257 ms. From fastest to slowest, in terms of selection method, the means were 2650 ms (either), 4157 ms (left hand) and 5963 ms (legs); in terms of target amplitude, the means were 3893 ms (400 pixels), and 4620 ms (200 pixels), and in terms of target width, the means were 4076 ms (200 pixels) and 4437 ms (100 pixels).

The effect on AT was significant for selection method, $F(2, 22)$ = 7.429, $p < .01$. A post hoc analysis shows that the target acquisition time for the selection method "either" (left hand or legs) was significantly lower than for the left-hand alone and legs alone selection methods. See Figure 4.
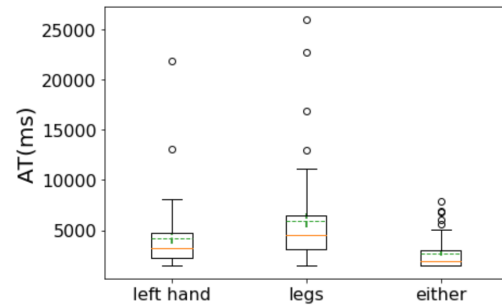

Figure 4: Effect of selection method on target acquisition time (AT). "Either" as the selection method had significantly lower movement time than legs. The green line shows the mean; the red line shows the median. Error bars show one standard error of mean.

4.2 Movement time

The grand mean for movement time (MT) was 1303 ms. From fastest to slowest, in terms of selection method, the means were 800 ms (either), 1306 ms (left hand) and 1802 ms (legs); in terms of target amplitude, the means were 1176 ms (200 pixels), and 1429 ms (400 pixels), and in terms of target width, the means were 813 ms (200 pixels) and 1793 ms (100 pixels).

The effect on MT was significant for selection method, $F(2, 20)$ = 3.794, $p < .05$, and target width, $F(1, 10)$ = 71.132, $p < .0001$. A post hoc analysis shows that the movement time for the selection method "either" (left hand or legs) was significantly lower than for the left-hand alone and legs alone selection methods. MT for the larger target width was significantly lower than for the smaller target width.

There were no other significant main or interaction effects on movement time ($p > .05$).

4.3 Selection time

The grand mean for selection time (ST) was 3198 ms. From fastest to slowest, in terms of selection method, the means were 2044 ms (either), 3152 ms (left hand) and 4396 ms (legs); in terms of target amplitude, the means were 2766 ms (400 pixels), and 3629 ms (200 pixels), and in terms of target width, the means were 2882 ms (100 pixels) and 3513 ms (200 pixels).

The effect on *ST* was significant for selection method, $F(2, 20) = 4.899$, $p < .05$. A post hoc analysis shows that the target acquisition time for the selection method "either" (left hand or legs) was significantly lower than for the left-hand alone and legs alone selection methods. See Figure 5.
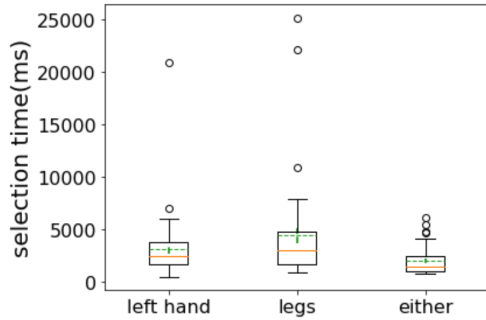

Figure 5: Effect of selection method on selection time (*ST*). "Either" as the selection method had significantly lower movement time than legs. The green line shows the mean; the red line shows the median. Error bars show one standard error of mean.

4.4 Throughput

The grand mean for throughput (*TP*) was 0.438 bits/s. From highest to lowest, in terms of selection method, the means were 0.593 bits/s (either), 0.436 bits/s (left-hand) and 0.284 bits/s (legs). In terms of target amplitude, the means were 0.542 bits/s (400 pixels) and 0.333 bits/s (200 pixels), and in terms of target width the means were 0.448 bits/s (100 pixels) and 0.427 bits/s (200 pixels).

The effect on *TP* was significant for selection method, $F(2, 22) = 13.896$, $p < .01$, and target amplitude, $F(1, 11) = 35.183$, $p < .0001$. A post hoc analysis showed that the selection method "either" had significantly higher throughput than the legs and left-hand selections methods. *TP* was significantly higher at the larger target amplitude.

The interaction effects were significant for selection method by target amplitude, $F(2, 22) = 4.966$, $p < .05$. The larger amplitude and faster selection method increased the throughput, as shown in Figure 6.

There were no other significant main or interaction effects on *TP* ($p > .05$).

4.5 Error rate

The grand mean for error rate (*ER*) was 11.81%. From lowest to highest, in terms of clicking method, the means were 7.1% (left-hand), 13.8% (legs) and 14.9% (either). In terms of target amplitude, the means were 10.3% (400 pixels) and 13.3% (200 pixels), and in terms of target width, the means were 4.4% (200 pixels) and 19.2% (100 pixels). The effect on *ER* was significant for target width only, $F(1, 11) = 13.076$, $p < .01$.

The polar plots of *ER* for target widths 100 pixels and 200 pixels are shown in Figure 7. No obvious skewness pattern in terms of target coordinates was observed.
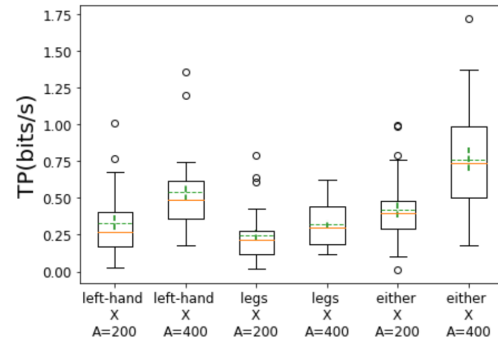

Figure 6: Interaction effect of selection method by target amplitude. "Either" by 400 px target amplitude has the highest throughput. The green line shows the mean; the red line shows the median. Error bars show one standard error of the mean.
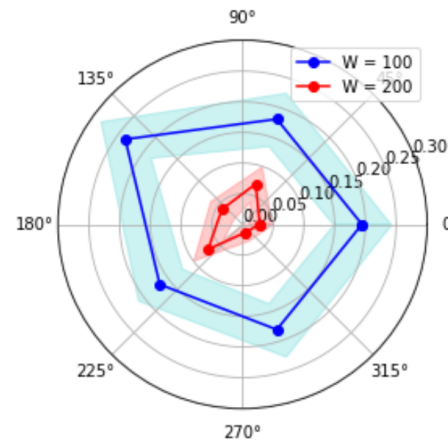

Figure 7: Polar plot showing error rate (*ER*) for targets at their screen coordinates. One standard error of mean is shaded over the mean.

4.6 Pupil size

The pupil data for participant 19 is shown in Figure 8 as an example, with the raw pupil data on the left. The black horizontal lines on the left graph show the data within 1 standard deviation of the mean. On the right, the pupil data of the left-hand block are shown. Each part marked by a yellow column is a sequence of trials. The parts that are not shaded are the rest periods between sequences, where the subject is idle or relaxed. Fluctuations at various frequencies reminiscent of pupillary unrest/hippus [Bouma and Baghuis 1971; Ohtsuka et al. 1988; Stark et al. 1958] appear to be visible in the recorded pupil data, but generally the pupil size appear smaller during the rests between the active task sequences.

The grand mean of the baseline was 20.037 pixels. The grand mean of *PS* during active task sequences was 0.00134 pixels, close to the baseline as expected. From smallest to largest, in terms of selection methods, the means were -0.9462 pixels (left hand), -0.3027
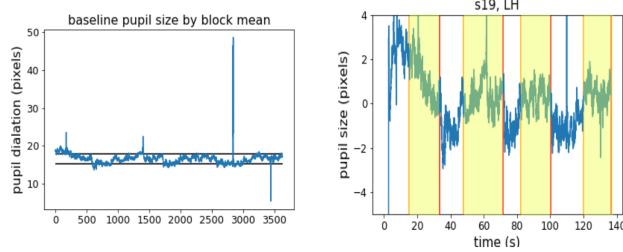
Figure 8: Participant 19's left-hand block pupil data. The figure on the left shows the threshold for accepting data points. The thresholds are plotted as two black horizontal lines on the raw pupil data. On the right, the baseline-corrected pupil data for a block are shown. Areas shaded in green are active task sequences. The participant rests between the sequences.

pixels (legs) and 0.129 pixels (either). In terms of target amplitude, the means were -0.011 pixels (400 pixels) and 0.137 (200 pixels), and in terms of target width, the means were -0.171 pixels (100 pixels) and 0.174 pixels (200 pixels).

The main effect on PS was significant for target width, $F(1, 11)$ = 15.557, $p < .01$. The larger the target width, the bigger the pupil dilates.

There were no other significant main or interaction effects.

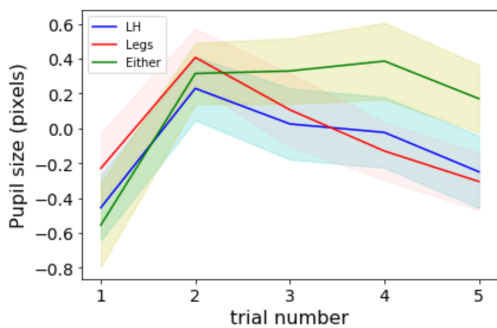The mean pupil size over time in our experiment are plotted in Figure 9.

Figure 9: Mean pupil size by trial index for each of the selection methods. One standard error of mean is shaded over the mean.

4.7 BCI MI classification

For the three-class classification of left hand vs. both legs vs. idle, the algorithm's percentage accuracy was 60.80% ± 13.41 with a kappa of 0.43 ± 0.18. For the binary classification of MI vs. idle (i.e., 'Either' selection condition), the algorithm's percentage accuracy was 74.27% ± 18.15, and kappa value was 0.49 ± 0.22. For the binary classification of left hand vs. both legs, the algorithm's percentage accuracy was 73.97% ± 10.62 with a kappa of 0.47 ± 0.36. Testing the algorithm using the left hand and feet data extracted from BCI

competition IV 2a [Brunner et al. 2008], the percentage accuracy was 83.36% ± 10.11 with a kappa of 0.73 ± 0.20, which were higher than the right hand vs. left hand paradigm from the same data set, the percentage accuracy was 74.62% ± 26.02 with a kappa of 0.49 ± 0.28, in congruence with the findings in Tang et al. [2016].

5 USER EXPERIENCE

Participants compled a post-experiment evaluation questionnaire. They were asked to rate the mental demand for each selection method on a scale of 1 to 10, with 1 being "not at all" and 10 being "extremely". From lowest to highest, the means were 5.71 ± 1.98 (either), 5.93 ± 2.13 (legs), and 6.07 ± 2.09 (left-hand). They were also asked to rate the physical demand of each selection method. In this case, the means were 2.5 ± 1.65 for left-hand and legs, and 2.43 ± 1.65 for the "either" condition. There was no significant difference between the mental or physical demand ratings for the selection methods.

Participants indicated they expended the most effort with left-hand selection at 6.64 ± 1.45, followed by legs at 6.07 ± 1.33, and "either" at 5.86 ± 1.51. The left-hand effort was significantly higher than legs and "either" ($p < .01$). The frustration level was highest for left-hand 4.86 ± 2.32, followed by 4.57 ± 2.24 for legs, and 4.07 ± 1.54 for "either". There was no significant difference in frustration ratings ($p > .05$).

On a scale of 1 to 3, 1 being most preferred, participants rated the legs (1.79 ± 0.89) the most preferred selection method, followed by "either" (2.07 ± 0.62) and left-hand (2.14 ± 0.95). There was no significant difference in preference for selection method ($p > .05$).

Subjects commented "it was difficult to completely concentrate my thoughts. Also thinking about avoiding movement at the same time was difficult. Every time I failed it got increasingly difficult to click, so I had to try to clear my mind all the time before I could click again." It was observed that some participants' bodies cramped up slightly when they were trying to activate a click. Their breathing became heavier, and they even started fidgeting due to not being able to succeed. This made it more challenging in turn, since their mental state was different from training, since the effects of breathing, tensing of muscles, and eye movements were absent in the training data – they were calm and almost "meditating" during EEG recording. When they were told "don't try harder, but try more similar to the training imagination", they were able to activate selections better.

6 DISCUSSION

When the selection method is "either", the BCI classification is reduced to a binary classification between rest and motor imagery – whether the imagery is left-hand or legs is not considered. This makes the classification simpler. Participants commented that it was "nice to have the option to use 'either' because this "made it less frustrating as it was possible to change strategy when 'stuck'". This is possibly the reason why the "either" clicking method scored the highest on throughput. Using the "either" method for clicking, however, would rule out the possibility to have different actions mapped to left-hand or legs motor imagery, for instance a single click and a double click.

The computed throughput values reported herein are quite low: less than 1 bit/s for all three selection methods. Values in the literature for throughput are generally higher – about 4-5 bits/s for mouse input (see [Soukoreff and MacKenzie 2004], Table 4). Other devices generally fair poorer with values of about 1-3 bits/s for the touchpad or joystick. However, throughputs <1 bit/s have been observed when testing unusual cursor control schemes or when engaging participants with motor disabilities [Cuaresma and MacKenzie 2017; Felzer et al. 2016; Hassan et al. 2019; MacKenzie and Teather 2012; Roig-Maimó et al. 2017].

The throughput in most BCI literature is reported as information transfer rate (ITR) [Wolpaw et al. 2002]. As Choi et al. [Choi et al. 2017] note, only about 32% of literature reports ITR; mostly task accuracy is used as a metric in BCI studies. BCI paradigms commonly report throughput <1 bit/s. For example, 0.5 bits/s [Volosyak et al. 2010], 0.15 bits/s [Pasqualotto et al. 2015], 0.2 bits/s [Riccio et al. 2015], and 0.05-1.44 bits/s [Holz et al. 2013]. A high throughput speller by Chen et al. [Chen et al. 2014] achieved 1.75 bits/s. However, most of reviewed studies used P300 or SSVEP paradigms as they are commonly used for speller design. Obermaier et al. [Obermaier et al. 2001] used hand MI, but reported ITR as bits/trial, if we assume a trial to be 1.25 s long, it gives a maximum throughput around 0.6 bits/s. GIMIS's grand mean of 0.438 bits/s is higher than most BCI literature reviewed, likely due to the gaze input.

The error rate (grand mean 11.8%) was much higher than reported by Bækgaard et al. [Bækgaard et al. 2019], where target widths of 50 pixels and 100 pixels with click selection were used and the error rates for gaze, head, and mouse pointing were 1.53%, 0.26%, and 0.24%, respectively. Table 4 in Stark et al. [Soukoreff and MacKenzie 2004] surveyed nine studies with error rates ranging from 1.6% for RemotePoint to 32% for laser pointer. Most paradigms report error rates below 10%.

The BCI may produce false-positive clicks, if a selection happens inside a target before the user begins to imagine the movement. A larger target increases the probability of such false-positive selections being inside the correct target. Another contributor of the error is that, as some subjects commented, it is difficult to keep the eyes fixated while imagining the movements, a larger target therefore gave more buffer for eye movements during imagination too, and if the eye tracker was 'shaky', a larger target helps by providing a larger activation area.

The time for gaze to steer the cursor into the target is much faster than for the BCI to classify a new EEG signal; however, BCI is the main contributor to AT. The expected effect of target amplitude and width were possibly obscured and rendered insignificant by the slow BCI.

The gaze pointing movement time recorded in our experiments were much longer than reported by, for example, Bækgaard et al. [Bækgaard et al. 2019] and Mateo et al. [Mateo et al. 2008]; they report movement time less than 500 ms. Our participants rated an average 6 out of 10 for gaze control, suggesting the long movement time could be caused by imperfect eye tracker calibration or usage.

Whether and how the mental difficulty of the MI may have impacted movement time needs to be investigated further. Some users commented on not being able to stay relaxed, or needing to stop imagining immediately once the target had been selected. This implies the BCI may have issued a false-positive selection even though the user was not in an MI state. If this happens, for the next trial the user may consciously relax more and not work as hard so as to avoid over-activation. The effort to relax more may have started before moving the eyes, which would then slow down movement time.

Contrary to what we might have expected from the difficulty and load caused by motor task complexity and movement planning, a smaller target width, in this case, decreased pupil size. Picking approximately equiliumant targets and background across all conditions aimed at avoiding the pupillary light reflex [Ellis 1981] as an effect on the pupil size. Hence, our findings may suggest that the increased demand for precision caused by the smaller target width is the stronger effect, in line with findings reported by Fletcher et al. [2017].

The mean value of the pupil diameter by trial within each sequence varies with time. Bækgaard et al. [Bækgaard et al. 2019] reported that the pupil diameter increased approximately 8% over the first trials, stabilizing over the remaining sequence. For all selection methods we similarly observed a comparable start-up effect from the first to the second task of approximately 5% (cf. Figure 9). For the selection method "either" we also observed an approximately stable pupil size hereafter, but for the selection methods "legs" and "left-hand", a decrease in pupil size is seen after the second selection. However, we only had five targets per sequence, which may be too few to see any convincing trend, compared to 21 targets per sequence reported by Bækgaard et al. [Bækgaard et al. 2019]. Some users commented they had difficulties staying relaxed, which caused accidental/premature selections. They may have compensated by lowering their arousal, leading to a decrease in pupil size. Additionally, it may be that forcing users to their non-preferred selection method could also cause a disengagement, leading to a lower invested effort and cognitive load, resulting in a decrease in pupil size.

Finally, two practical issues should be considered, namely the time it takes to train the system and the comfort of the cap. The BCI signal was processed using the FBCSP and NBPW algorithms [Ang et al. 2012], which took about five minutes to train for each participant. A more accurate 2D convolutional neural network (CNN) classifier, similar to Sakhavi et al. [Sakhavi et al. 2018] was implemented, however, the CNN model took about 45 minutes to train per participant, which is a challenge in the EEG lab, where we had limited time to complete the experiments, the electrical conducting gel started to dry and participants started to get tired; this negatively impacted signal quality and performance. Therefore, we chose the faster NBPW algorithm to meet this logistic demand.

The EEG cap took time to setup. Each electrode has to be filled with gel which may dry up in long experiments, requiring the subjects to wash their hair, wait for their hair to dry and reapply the gel. Additionally, the cap may not fit all participants perfectly; some had the edge of the cap over their eyebrows, Although none of the subjects reported obstructed vision, the cap was uncomfortable for some participants.

7 CONCLUSION

Our experiment has shown the GIMIS method to be feasible. The best performance was obtained when participants were free to

choose if they would use either the left-hand or legs for motor imagery selection. However, throughputs were lower and error rates were higher than values reported in other Fitts' law studies. Therefore, we suggest that GIMIS should primarily be considered as a way to introduce BCI for patients who are already familiar with gaze interaction but in danger of loosing their eye control.

ACKNOWLEDGMENTS

REFERENCES

2014. Basics: eyetribe-docs. https://theeyetribe.com/dev.theeyetribe.com/dev.theeyetribe.com/general/index.html

Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience* 6 (2012), 39.

Karissa C Arthur, Andrea Calvo, T Ryan Price, Joshua T Geiger, Adriano Chio, and Bryan J Traynor. 2016. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. *Nature Communications* 7 (2016), 12408.

Per Bækgaard, John Paulin Hansen, Katsumi Minakata, and I Scott MacKenzie. 2019. A Fitts' law study of pupil dilations in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications – ETRA '19*. ACM, New York, 32:1–32:5.

H Bouma and L C J Baghuis. 1971. Hippus of the pupil: Periods of slow oscillations of unknown origin. *Vision Research* 11, 11 (1971), 1345–1351.

C Brunner, R Leeb, G Müller-Putz, A Schlögl, and G Pfurtscheller. 2008. BCI Competition 2008–Graz data set A. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* 16 (2008).

Xiaogang Chen, Zhikai Chen, Shangkai Gao, and Xiaorong Gao. 2014. A high-itr ssvep-based bci speller. *Brain-Computer Interfaces* 1, 3-4 (2014), 181–191.

Inchul Choi, Ilsun Rhiu, Yushin Lee, Myung Hwan Yun, and Chang S Nam. 2017. A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. *PLoS one* 12, 4 (2017), e0176674.

Justin Cuaresma and I Scott MacKenzie. 2017. FittsFace: Exploring navigation and selection methods for facial tracking. In *Proceedings of Human-Computer Interaction International - HCII 2017 (LNCS 10278)*. Springer, Berlin, 403–416. https://doi.org/10.1007/978-3-319-58703-5_30

Edwin Dalmaijer. 2014. *Is the low-cost EyeTribe eye tracker any good for research?* Technical Report. PeerJ PrePrints.

C J Ellis. 1981. The pupillary light reflex in normal subjects. *British Journal of Ophthalmology* 65, 11 (1981), 754–759.

T Felzer, I Scott MacKenzie, and John Magee. 2016. Comparison of two methods to control the mouse using a keypad. In *Proceedings of the 15th International Conference on Computers Helping People With Special Needs – ICCHP 2016*. Springer, Berlin, 511–518. https://doi.org/10.1007/978-3-319-41267-2_72

Kingsley Fletcher, Andrew Neal, and Gillian Yeo. 2017. The effect of motor task precision on pupil diameter. *Applied Ergonomics* 65 (2017), 309–315.

Antonio Frisoli, Claudio Loconsole, Daniele Leonardis, Filippo Banno, and Massimo Bergamasco. 2012. A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks. *IEEE Transactions on Systems Man & Cybernetics - Part C* 42, 6 (2012), 1169–1179.

Mehedi Hassan, John Magee, and I Scott MacKenzie. 2019. A Fitts' law evaluation of hands-free and hands-on input on a laptop computer. In *Proceedings of the 21st International Conference on Human-Computer Interaction – HCII 2019 (LNCS 11572)*. Springer, Berlin, 234–249. https://doi.org/10.1007/978-3-030-23560-4

Elisa Mira Holz, Johannes Höhne, Pit Staiger-Sälzer, Michael Tangermann, and Andrea Kübler. 2013. Brain–computer interface controlled gaming: Evaluation of usability by severely motor restricted end-users. *Artificial intelligence in medicine* 59, 2 (2013), 111–120.

Robert J K Jacob. 1991. The use of eye movements in human–computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9 (1991), 152–169. https://doi.org/10.1145/123078.128728

Xianta Jiang, M Stella Atkins, Geoffrey Tien, Bin Zheng, and Roman Bednarik. 2014. Pupil dilations during target-pointing respect Fitts' law. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 175–182.

Xianta Jiang, Bin Zheng, Roman Bednarik, and M Stella Atkins. 2015. Pupil responses to continuous aiming movements. *International Journal of Human-Computer Studies* 83 (2015), 1–11.

Ivo Käthner, Andrea Kübler, and Sebastian Halder. 2015. Comparison of eye tracking, electrooculography and an auditory brain-computer interface for binary communication: A case study with a participant in the locked-in state. *Journal of neuroengineering and rehabilitation* 12, 1 (2015), 76.

Yongwon Kim and Sungho Jo. 2015. Wearable hybrid brain-computer interface for daily life application. In *The 3rd International Winter Conference on Brain-Computer Interface*. IEEE, 1–4.

Eui Chul Lee, Jin Cheol Woo, Jong Hwa Kim, Mincheol Whang, and Kang Ryoung Park. 2010. A brain-computer interface method combined with eye tracking for 3D interaction. *Journal of Neuroscience Methods* 190, 2 (2010), 289–298. https://doi.org/10.1016/j.jneumeth.2010.05.008

Jeong Hwan Lim, Jun Hak Lee, Han Jeong Hwang, Hwan Kim Dong, and Chang Hwan Im. 2015. Development of a hybrid mental spelling system combining SSVEP-based brain–computer interface and webcam-based eye tracking. *Biomedical Signal Processing & Control* 21, 2013 (2015), 99–104.

Robert Gabriel Lupu, Florina Ungureanu, and Valentin Siriteanu. 2013. Eye tracking mouse for human computer interaction. In *2013 E-Health and Bioengineering Conference - EHB '13*. IEEE, New York, 1–4.

I Scott MacKenzie. 2013. *Human-computer interaction: An empirical research perspective.* Morgan Kaufmann, Waltham, MA.

I S MacKenzie and R. J. Teather. 2012. FittsTilt: The application of Fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction - NordiCHI 2012*. ACM, New York, 568–577. https://doi.org/10.1145/2399016.2399103

Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, London, 39–65.

Julio C Mateo, Javier San Agustin, and John Paulin Hansen. 2008. Gaze beats mouse: Hands-free selection by combining gaze and emg. In *Proceedings of the CHI '08 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, 3039–3044.

Nils Nappenfeld and Gerd-Jürgen Giefing. 2018. Applying Fitts' Law to a Brain-Computer Interface Controlling a 2D Pointing Device. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 90–95.

Bernhard Obermaier, Christa Neuper, Christoph Guger, and Gert Pfurtscheller. 2001. Information transfer rate in a four-classes brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9, 3 (2001), 283–288.

K Ohtsuka, K Asakura, H Kawasaki, and M Sawa. 1988. Respiratory fluctuations of the human pupil. *Experimental Brain Research* 71, 1 (1988), 215–217.

Emanuele Pasqualotto, Tamara Matuz, Stefano Federici, Carolin A Ruf, Mathias Bartl, Marta Olivetti Belardinelli, Niels Birbaumer, and Sebastian Halder. 2015. Usability and workload of access technology for people with severe motor impairment: A comparison of brain-computer interfacing and eye tracking. *Neurorehabilitation and Neural Repair* 29, 10 (2015), 950–957.

Gert Pfurtscheller, Brendan Z Allison, Clemens Brunner, Gunther Bauernfeind, Teodoro Solis-Escalante, Reinhold Scherer, Thorsten O. Zander, Gernot Mueller-Putz, Christa Neuper, and Niels Birbaumer. 2010. The hybrid BCI. *Frontiers in Neuroscience* 4, 30 (2010), 30–30.

Angela Riccio, Elisa Mira Holz, Pietro Aricò, Francesco Leotta, Fabio Aloise, Lorenzo Desideri, Matteo Rimondini, Andrea Kübler, Donatella Mattia, and Febo Cincotti. 2015. Hybrid P300-based brain-computer interface to improve usability for people with severe motor disability: Electromyographic signals for error correction during a spelling task. *Archives of Physical Medicine and Rehabilitation* 96, 3 (2015), S54–S61.

Francois Richer and Jackson Beatty. 1985. Pupillary dilations in movement preparation and execution. *Psychophysiology* 22, 2 (1985), 204–207.

M F Roig-Maimó, I S MacKenzie, C Manresa, and J Varona. 2017. Evaluating Fitts' law performance with a non-ISO task. In *Proceedings of the 18th International Conference of the Spanish Human-Computer Interaction Association*. ACM, New York, 51–58. https://doi.org/10.1016/j.ijhcs.2017.12.003

R Rupp. 2014. Challenges in clinical applications of brain computer interfaces in individuals with spinal cord injury. *Frontiers in Neuroengineering* 7 (2014), 38. https://doi.org/10.3389/fneng.2014.00038

Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. 2018. Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Transactions on Neural Networks & Learning Systems* 29, 11 (2018), 5619–5629.

R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

Rosella Spataro, Maria Ciriacono, Cecilia Manno, and Vincenzo La Bella. 2014. The eye-tracking computer device for communication in amyotrophic lateral sclerosis. *Acta Neurologica Scandinavica* 130, 1 (2014), 40–45.

Lawrence Stark, Fergus W Campbell, and John Atwood. 1958. Pupil unrest: an example of noise in a biological servomechanism. *Nature* 182, 4639 (1958), 857–858.

Zhichuan Tang, Shouqian Sun, Sanyuan Zhang, Yumiao Chen, Chao Li, and Shi Chen. 2016. A brain-machine interface based on ERD/ERS for an upper-limb exoskeleton control. *Sensors* 16, 12 (2016), 2050.

Roman Vilimek and Thorsten O. Zander. 2009. BC(eye): Combining eye-gaze input with brain-computer interaction. In *Proceedings of Universal Access in Human-Computer Interaction - UAHCI '09*. Springer, Berlin, 593–602.

Ivan Volosyak, Diana Valbuena, Tatsiana Malechka, Jan Peuscher, and Axel Gräser. 2010. Brain–computer interface using water-based electrodes. *Journal of Neural Engineering* 7, 6 (2010), 066007.

Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. 2002. Brain–computer interfaces for communication and control. *Clinical Neurophysiology* 113, 6 (2002), 767–791.

Thorsten O Zander, Matti Gaertner, Christian Kothe, and Roman Vilimek. 2010. Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *International Journal of Human–Computer Interaction* 27, 1 (2010), 38–51.