Pushing the P300-based brain–computer interface beyond 100 bpm: extending performance guided constraints into the temporal domain

2016 J. Neural Eng. 13 026024

(http://iopscience.iop.org/1741-2552/13/2/026024)

IOP Publishing
Journal of Neural Engineering
J. Neural Eng. **13** (2016) 026024 (15pp)
doi:10.1088/1741-2560/13/2/026024

Pushing the P300-based brain–computer interface beyond 100 bpm: extending performance guided constraints into the temporal domain

G Townsend and V Platsko

Algoma University, Sault Ste. Marie, Ontario, P6A 2G4, Canada

E-mail: townsend@algomau.ca

Received 21 August 2015, revised 4 January 2016
Accepted for publication 20 January 2016
Published 25 February 2016

Abstract

Objective. A new presentation paradigm for the P300-based brain–computer interface (BCI) referred to as the 'asynchronous paradigm' (ASP) is introduced and studied. It is based on the principle of performance guided constraints (Townsend *et al* 2012 *Neurosci. Lett.* **531** 63–8) extended from the spatial domain into the temporal domain. The traditional constraint of flashing targets in predefined constant epochs of time is eliminated and targets flash asynchronously with timing based instead on constraints intended to improve performance. *Approach.* We propose appropriate temporal constraints to derive the ASP and compare its performance to that of the 'checkerboard paradigm' (CBP), which has previously been shown to be superior to the standard 'row/column paradigm' introduced by Farwell and Donchin (1988 *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23). Ten participants were tested in the ASP and CBP conditions both with traditional flashing items and with flashing faces in place of the targets (see Zhang *et al* 2012 *J. Neural Eng.* **9** 026018; Kaufmann and Kübler 2014 *J. Neural Eng.* **11**; Chen *et al* 2015 *J. Neurosci. Methods* **239** 18–27). Eleven minutes of calibration data were used as input to a stepwise linear discriminant analysis to derive classification coefficients used for online classification. *Main results.* Accuracy was consistently high for both paradigms (87% and 93%) while information transfer rate was 45% higher for the ASP than the CBP. In a free spelling task, one subject spelled a 66 character sentence (from a 72 item matrix) with 100% accuracy in 3 min and 24 s demonstrating a practical throughput of 120 bits per minute (bpm) with a theoretical upper bound of 258 bpm. The subject repeated the task three times in a row without error. *Significance.* This work represents an advance in P300 speller technology and raises the ceiling that was being reached on P300-based BCIs. Most importantly, the research presented here is a novel and effective general strategy for organising timing for flashing items. The ASP is only one possible implementation of this work since in general it can be used to describe all previous existing presentation paradigms as well as any possible new ones. This may be especially important for people with neuromuscular disabilities.

Keywords: brain–computer interface, EEG, P300, event-related potential, rehabilitation

1. Introduction

An electroencephalographic (EEG) based brain–computer interface (BCI) can provide a non-muscular method of communication for people whose motor abilities have been impaired by neuromuscular disease [6]. In amyotrophic lateral sclerosis (ALS) motor neuron death eventually paralyzes the patient, but cognitive function is usually unaffected. BCI technology can help restore communication for ALS patients because it does not require voluntary muscle control.

P300-based BCIs use EEG to recognise the P300 response to an attended stimulus [2]. The P300 is an event-related potential (ERP) consisting of a positive deflection in the EEG over the parietal cortex occurring approximately 300 ms after the presentation of a rare or meaningful stimulus. Due to other brain activity and noise in EEG signals, several P300 responses must typically be averaged in order for the response to be recognised.

In the row/column paradigm (RCP) introduced by Farwell and Donchin [2], a 36-character matrix containing 6 rows and 6 columns is displayed on the screen, and the subject attends to the character they wish to select. The rows and columns flash randomly, and flashes that contain the attended item will, in theory, elicit a P300 response. Variations on the RCP have been introduced to improve its accuracy, including flash colour [7], item size and distance [8], and stimulus type [9].

Errors in the RCP have been shown to occur most frequently on stimuli adjacent to target flashes [10]. The checkerboard paradigm (CBP) attempted to eliminate this problem by constraining which targets may flash together [11]. Additionally, the CBP eliminates the double flash problem which occurs in the RCP where a row containing a target flashes immediately following a column containing the same target (or vice versa). Increasing the distance between target flashes increases the amplitude of the P300 response [12]. Extending the original work of the CBP, Townsend *et al* designed a paradigm in which spatial constraints were eliminated altogether and the targets were grouped based on abstract sets. Then constraints were enforced only to improve the interface leading to the development of the concept of 'performance guided constraints' [1].

Many recent papers use BCI2000 [13, 14], a general-purpose platform for BCIs [15]. BCI2000 does not synchronise its stimulus markers to what is physically happening on the display, which results in inherent inaccuracies. This may be exacerbated depending on the stimulus presentation rate. For example, an 8 Hz paradigm implemented on a 60 Hz monitor cannot have each epoch be precisely the same length. (Note that the stimulus markers could, in theory, still be correct if the markers were carefully synchronised with the display by means of a photo-sensor.)

1.1. Evolution of stimulus presentation

Some of the problems with the traditional RCP were addressed with the introduction of the CBP [11]. The CBP addressed the 'double flash' and the 'adjacency distraction' problems by focusing on eliminating the spatial constraints of flashing items in rows and columns. However, the paradigm continued to flash items in rows and columns in 'virtual matrices' hidden from the subject. The next step in the evolution of stimulus paradigms was to eliminate spatial organisations altogether and to organise flash groups in a completely abstract way described in [1]. This paper (designated by the editor as a 'plenary paper') described the elimination of all spatial constraints on the flashing pattern initially, and then only constraints designed to improve the performance of the interface were added. It captured all possible spatial organisations previously described (RCP, CBP, etc) as well as any still to be invented. In the present work, we continued this evolution, now extending it from the spatial domain to the temporal domain. Initially we remove all temporal constraints on how stimuli are presented, and then only introduce constraints designed to improve performance. Therefore, items no longer flash in epochs as in the past, but are initially free to flash at any time; alone, near, or with any other targets. Then constraints intended to address specific performance issues are added to improve the interface. This is the principle of 'performance guided constraints' presented in [1] now extended to the temporal domain.

1.2. The present study—the asynchronous paradigm (ASP)

Since the presentation of stimuli is no longer synchronized to epochs, we call this new paradigm the 'asynchronous paradigm' or 'ASP' for short. NB: in the present work we define a 'flash period' to be the portion of a flash sequence in which all targets flash once. We define a 'flash sequence' to consist of ten such flash periods (during training) or fewer (when reduced to optimise throughput during testing). We now introduce the temporal 'performance guided constraints' employed in the paradigm:

(1) An item that flashes temporally near another item should never flash near to that same item again for the rest of the sequence. This improves performance by not having a target score shared with another item multiple times since when that happens, the other item has a higher probability of being incorrectly selected by the classifier. How we determined what constitutes 'near' in this context will be discussed later.

(2) Similarly, items flashing just a little farther apart than 'near' from one another may only be permitted to do so again at most one additional time. Note that whereas our primary constraint above is strictly honoured, this secondary constraint is not strictly observed. The reason for this is that if we were to strictly enforce this constraint we may quickly run out of eligible neighbours who have not flashed together and items may eventually have to flash isolated from one another as a result. This would extend the time required for an entire typical training sequence.

(3) A given target must not flash any sooner than 500 ms after its previous flash. This avoids the double flash problem and although it is a little quicker than some of the literature recommends, we have found that it produces acceptable cognitive responses.

(4) All items should flash $N-1$ times before any item flashes for the Nth time. In other words, flash periods may not overlap. Since not all subjects will require the same number of flashes to achieve their maximum throughput, it is reasonable to enforce this constraint to avoid wasting time flashing targets more times than required for a particular subject.

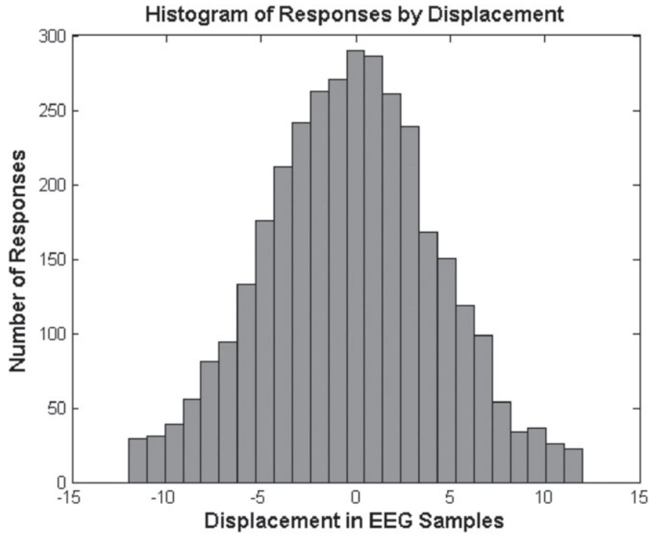

Figure 1. Distribution of at what displacement P300 target response scores peaked compared to the original offset (i.e. zero EEG samples sampled at 256 Hz) at which each response occurred.

1.3. Determination of timing parameters

If the classifier could consistently score a target response represented in the EEG with its maximum score exactly where it occurred and if all responses appeared with exactly the same latency from the stimulus associated with them every time, then our definition of 'near' in our first constraint above would essentially collapse to zero. However, in practice there is some variability or 'jitter' involved and as a result we do not expect a classifier to be able to reliably distinguish between target and non-target responses if they occur too close together. Determining the amount of jitter in this process will help to determine how narrow of a window around a response we can construct such that we expect to be able to reliably distinguish a target response within this window from a non-target response outside of the window.

To determine the width of this window, we analysed previous CBP data in which a photo-sensor was used to mark the exact location of target stimuli in the EEG. Data consisting of 3600 target responses sampled at 256 Hz from subjects performing a P300 spelling task was analysed by artificially shifting the event marker ± 12 EEG samples (± 46.9 ms) from the original classification point and observing how the classifier score of each target response varied over that range. Figure 1 above shows the distribution of where these 3600 responses peaked relative to the point where the classifier originally scored them.

It is clear that the responses are normally distributed, peaking at various distances from their nominal positions and we deemed the range over which the vast majority (i.e. approximately 90%) of target responses peak as defining the size of our 'near' window implicitly described in our first constraint above. Figure 2(a) through (d) provides a more detailed analysis of the responses analysed. The curves in this sequence of figures are plotted as the raw classifier scores evaluated between ± 16 EEG samples from the original classification points. The numerical values of the classifier scores are of no consequence in this context and have intentionally been omitted from the vertical axes. We are only interested in the shapes of the curves and the locations of their peaks.

Fewer than 10% of the responses peak at their original classification position while the majority peak within ± 4 EEG samples of that position. When the window is widened to ± 8 EEG samples, almost 90% of the responses peak within that window. This width corresponds to ± 31.3 ms and defines 'near' in the context of our primary constraint. The wider range of ± 12 samples captures almost all of the responses (95%) and corresponds to ± 46.9 ms. This defines the alternate window described in our second constraint earlier. For convenience we will round these window widths up to 32 ms and 48 ms.

We now have parameters with which we can describe the ASP. Specifically, our primary constraint requires that once an item flashes within 32 ms of another, it is never allowed to flash that close again. Our secondary constraint requires that most of the time items flashing within 48 ms of one another may ideally do so again no more than one additional time. In figure 3 below, an example section of a flashing sequence is given in the context of the item 'G' implicitly showing examples of 'near,' 'farther,' and 'distant' neighbours.

1.4. Practical implementation issues

We will implement this paradigm using a computer monitor only capable of allowing items to flash with a time resolution equal to the frame rate of the monitor, typically 60 frames per second. The reader should resist the temptation of thinking of this as simply being a variation of the traditional implementation done in epochs of 16.67 ms. Given the variability of brain responses and the expected temporal accuracy of the classifier it would be difficult to argue that we should expect different results if we implemented this on, say a LED-based display where the resolution of the timing based on a microcontroller could be in the microsecond range. In our opinion, the constraints imposed by the monitor do not prevent us from implementing and testing our paradigm. We merely need to shift our theoretical stimuli positions by no more than ± 8.3 ms (i.e. ± 2 EEG samples) to the start of the nearest video frame. In the context of the phenomenon we are attempting to classify (EEG), we do not expect ± 2 EEG samples in one direction or the other to matter. In fact, figures 2(a)–(d) below shows very little variation in the classifier scores over that range.

Since a video frame is 16.67 ms in duration, we will modify our constraints to integral multiples of this time, specifically 33.33 ms and 50 ms instead of the originally discussed 32 ms and 48 ms respectively. These changes are less than one EEG sample difference. We may now think of the constraint windows in terms of video frames as depicted in figure 4 below which is the 'video frame' version of figure 3.

Note that the secondary constraint of ± 3 video frames is not honoured strictly which allows some leeway to pack items

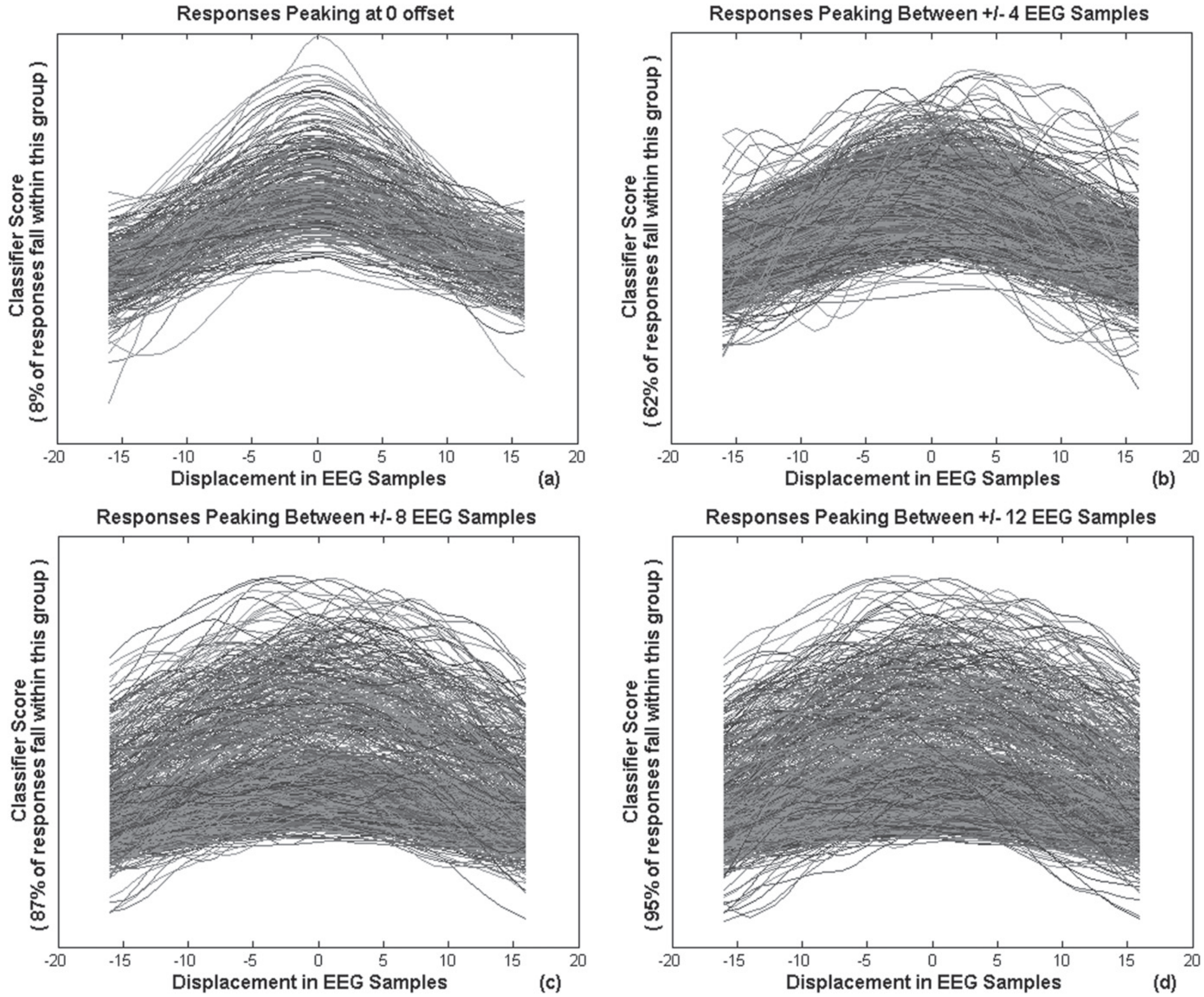

Figure 2. The classifier scores as the responses are shifted earlier or later. In (a), only those responses that peaked right at the original point of classification appear. In (b), all responses peaking between ± 4 EEG samples of that position appear including those in (a). In (c) the responses peaking within ± 8 EEG samples appear, and in (d) those peaking within ± 12 samples are shown.

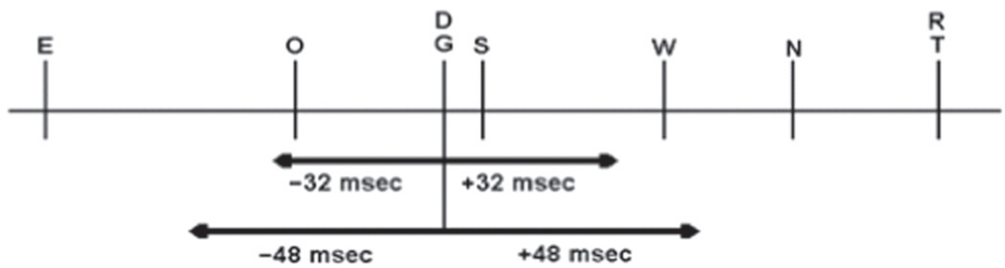

Figure 3. The item G above may never again flash within ± 32 ms of the D, O, or S and should only flash within ± 48 ms of the W at most once more. This is also the case for the D since it occurs at the same point in time as the G. Similar boundaries may be set up for all the other items. For example, the O can never flash within ± 32 ms of the D or G again, however it would be fine if the O and S flashed closer than that to one another in the future.

more tightly together, however the primary constraint of ± 2 video frames is strictly honoured. By making the secondary constraint flexible, we can avoid having the flash periods deteriorate very early into flashing individual targets with no neighbours nearby.

At the start of the flash sequence, it is easy to honour the constraints allowing stimuli to be packed in tightly thereby reducing the time required for the first flash of all items to occur. However if we pack things arbitrarily close together, we quickly use up the available combinations of neighbour

Figure 4. This figure is the same as the previous one but with the timing 'discretized' into integral multiples of 16.67 ms to fall at the start of each video frame. Here, items that flash within ±2 video frames of one another can never flash that close again. Items flashing within ±3 video frames can flash that close no more than twice.

clusters possible and the interface then degrades to the point of having to flash individual stimuli 50 ms apart. The development of an appropriate sequence that minimises the total time taken for ten flashes of all targets can be found by using an algorithm that employs random searching guided by appropriate heuristics. The details are beyond the scope of this paper, however we recommend starting off with a sequence in which groups of two or three items flash together in each frame with a few randomly occurring frames in which no flash onset occurs. In our case we 'seeded' the first flash period randomly to fill most of the first 39 video frames with two or three stimuli each so that all 72 matrix items had flashed once by the end of 650 ms. The length of this first flash period will qualify many targets to begin their second flash immediately after the end of the period while still honouring our third constraint that consecutive flashes of the same target be at least 500 ms apart. It also keeps items sufficiently spread out in the first flash period so that an abundance of combinations of neighbours that are still allowed to flash near to one another remain possible. One of the heuristics we employed in our algorithm was that the flash onset of no more than three items would occur in a single video frame. Although this is not a constraint imposed by the paradigm, it seems to permit the algorithm to develop shorter sequences than it otherwise would. Our algorithm included some randomness in its search, and we automatically ran it thousands of times searching for shorter and shorter sequences that honoured all the constraints. The final sequence was a total of 536 frames in length and will be described in more detail later. The algorithm found a 535 frame sequence after experiments had already begun. Optimising the sequence is a study for a different discipline. Once a successful sequence containing ten flash periods is found, the 72 targets may be randomly mapped into that sequence in multiple ways to randomise the presentation and 'recycle' the successful sequence.

It should be noted that once the onset of a stimulus is scheduled to appear in a particular video frame, it will persist over the following four video frames so that the stimulus duration matches that used in the control paradigm (CBP). Only the positions of the flash onsets of the stimuli are depicted in figure 4 above.

1.5. Comparison of timing to CBP

Table 1 below compares the total time required to flash through each of the ten flash periods from the start of the flash sequence for each paradigm for a 72 item matrix. Note that every flash in the CBP requires an additional 1.5 s, whereas early flash periods in the ASP take less time than later flash periods leading to a nonlinear relationship between the number of flashes and the total time taken to present them. However, even the longest flash period in the ASP is still much shorter than the 1.5 s that every flash period requires in the CBP. The lengthening periods in the ASP are the result of exhausting unused combinations of nearby neighbours as the flashing progresses causing subsequent flash periods to become progressively longer. Not all subjects will require ten flash periods, however all subjects are able to benefit from the faster flashing since it occurs at the start of the sequence.

Table 1. A comparison of total flash times and associated numbers of video frames for the paradigms. The CBP takes a constant number of video frames (90) for each flash period in the flash sequence. By comparison, the amount of time per flash period is variable in the ASP and increases as the flash sequence progresses. The values in the table represent the cumulative sums of the previous flash periods.

	Time in seconds		Video frames	
Flash	CBP	ASP	CBP	ASP
1	1.5	0.65	90	39
2	3.0	1.43	180	86
3	4.5	2.13	270	128
4	6.0	2.92	360	175
5	7.5	3.75	450	225
6	9.0	4.63	540	278
7	10.5	5.60	630	336
8	12.0	6.67	720	400
9	13.5	7.78	810	467
10	15.0	8.93	900	536

In our implementation of this paradigm we included three seconds of 'dead time' between selections to provide the subject ample time to locate the next item to be selected. Table 2 below shows the resulting practical throughput of the system for a given number of flashes required to reach 100% accuracy for the two paradigms as well as the upper bound on the theoretical limit based on excluding the three seconds of dead time between selections. The results are expressed as the 'written symbol rate' (WSR) as well as the bit rate based on selecting items from a 72 item matrix.

1.6. The control paradigm

The nature of presentation paradigms reported up to now have not relied heavily on precise display timing. For example, the results reported in [1, 11–15] all used the BCI2000 platform where despite the consistent spacing between event markers in the recorded data, there are significant inconsistencies between the actual stimuli and their associated markers. We compared these BCI2000 markers to digital markers from the EEG amplifier connected to a photo-sensor placed over

Table 2. A comparison of the theoretical and practical throughputs of the ASP and CBP by number of flashes.

Flash number	Theoretical WSR (symbols/minute)		Practical WSR (symbols/minute)		Theoretical bit rate (bits per minute)		Practical bit rate (bits per minute)	
	CBP	ASP	CBP	ASP	CBP	ASP	CBP	ASP
2	20.00	41.86	10.00	13.53	123.40	258.28	61.70	83.50
3	13.33	28.13	8.00	11.69	82.27	173.53	49.36	72.12
4	10.00	20.57	6.67	10.14	61.70	126.92	41.13	62.57
5	8.00	16.00	5.71	8.89	49.36	98.72	35.26	54.84
6	6.67	12.95	5.00	7.86	41.13	79.90	30.85	48.50
7	5.71	10.71	4.44	6.98	35.26	66.11	27.42	43.05
8	5.00	9.00	4.00	6.21	30.85	55.53	24.68	38.30
9	4.44	7.71	3.64	5.56	27.42	47.56	22.44	34.33
10	4.00	6.72	3.33	5.03	24.68	41.44	20.57	31.02

targets on the monitor to confirm this. Even in other results reported on other platforms the stimulus frequency and the frame rate of the monitor might make it impossible to avoid jitter (see relevant portion of introduction). Despite these problems, researchers have continued to develop and report on systems with imprecise timing.

This new paradigm by contrast requires precise frame-by-frame control of the video since the onset of a series of consecutive stimuli may appear in many consecutive video frames. As a result, unlike in previously reported paradigms, we do not have the option of being sloppy with our timing. This leaves us with a dilemma. On the one hand we want to compare the existing implementation of a well-known paradigm with good performance (i.e. the CBP) to our new paradigm, but if we 'fix' that implementation to have the same precise timing characteristics that are crucial to our new paradigm, we are instead comparing a 'new and improved' version of the CBP to our new paradigm. On the other hand, if we compare the original CBP implementation to the new paradigm, we draw the criticism that the timing is precise in the new paradigm and imprecise in the old paradigm and that the improvements are due to the exactness of the timing rather than the nature of the paradigm. We have given this much consideration and decided that our CBP results would look inflated unless we used an implementation comparable to those reported on in the past and that this criticism would outweigh the other. Therefore, we elected to conduct the study implementing the CBP in the traditional existing manner, but to conduct an offline analysis of the final results to determine the effect that correcting the timing would have had. The digital marking channel of the EEG amplifier was used to record the precise timing of the frame updates on the monitor to facilitate this as described later. When the correct timing is accounted for, the mean accuracy of the CBP increased by less than 3% ($\rho > 0.07$) and the mean throughput increased by less than 3 bpm ($\rho > 0.09$). Not only are these improvements minimal, but they are not statistically significant. Nevertheless, it is important that this has been considered and quantified in the context of the present work. Note that the significance tests reported here and wherever else ρ appears in the present work are all t-tests.

To be clear, we are comparing the existing implementation of the CBP running on the BCI2000 platform against our new paradigm. In order to randomly interleave the two paradigms on one display for the subject to view, we forwarded display information from the native BCI2000 system to our new system and mimicked the timing of the native monitor. Multiple tests were run with photo-sensors placed over the same targets on both the native and remote screens to confirm that the jitter between the two monitors was acceptable (under ±6 milliseconds) and to determine the (constant) latency so that appropriate adjustments could be made to the recorded data to obtain correct, accurate waveforms.

1.7. Faces versus faceless

The main purpose of this paper is to introduce a new paradigm, however given the success of the 'faces paradigm' and the relative ease involved in adding it to our system, it was reasonable to include that component in our experiments to see how much it could improve what we had already accomplished. In this paradigm, a picture of a face is briefly presented in place of an item rather than having the item flash. The brain's automatic visual processing related to the appearance of a face produces a unique visual response which modulates the time-course of the P300 responses. Further details are given in [16–18].

2. Methods

2.1. Participants

The subjects were 10 able-bodied adults (3 men, 7 women), recruited from the Algoma University community of students, staff, and faculty. Four were naïve to BCI use and none had uncorrected visual impairments. The study was approved by the Algoma University board of ethics and subjects gave informed consent.

2.2. Data acquisition, processing

Participants sat in a comfortable chair approximately 1 m from a computer monitor that displayed an 8 × 9 matrix (72 items) of selections. EEG was recorded with an 8-channel electrode cap with tin electrodes (Electro-Cap International). All channels were referenced to the right mastoid and grounded to the left mastoid. An 8-channel monopolar g.tec (Guger Technologies) amplifier was used to acquire the EEG (amplification to ± 2 V before ADC; high-pass 0.1 Hz; low-pass 30 Hz; digitisation rate 256 Hz). The eight electrode sites (based on 10/20) were Fz, Cz, P3, Pz, P4, P07, P08 and Oz [19]. The BCI2000 software platform [15] was used for data collection, word selection and prompting while an Ubuntu platform (discussed later) running custom software controlled stimulus presentation and online processing for the ASP.

2.3. Experimental paradigm

Subjects completed two sessions on different days within a one-week period. Each session consisted of a calibration (training) phase and an online testing phase. The experiment was fully automated by a script that interleaved the paradigms in pseudo-random order during both the testing and training phases. The words chosen to be spelled in each run were also randomized. However to ensure fair testing, no more than two runs of the same paradigm were tested in succession, and the same words were chosen for the training phase for each paradigm. This was the case both during training and testing.

In the calibration phase, subjects were shown a word to spell, displayed at the top left corner of the screen with the current target in parenthesis. Subjects were instructed to attend to each flash of the target item and to repeat this procedure for each letter of the word, as prompted.

Subjects were given six words totalling 36 letters during the calibration phase in each paradigm. Items in the matrix all flashed a total of ten times during this phase providing 360 target responses for each subject.

Subjects were presented with six different words during the online testing phase. Again the order of the paradigms and words during the testing phase was randomized and fully automated. The number of flashes was set to optimise the throughput for each subject based on the results of the training phase.

The subjects participated in two sessions one week apart. The primary study took place on the subjects' first visits and involved traditional flashing of the targets. The secondary study took place on the second visits and flashing targets were replaced with the appearance of a picture of a face. Both the primary and secondary studies were intended to compare the CBP to the ASP.

2.4. Classification

As described in Krusienski *et al* [20], independent SWLDA classifiers were derived for the CBP and ASP [21]. In the both the CBP and ASP calibration phases, each item selection included the same number of target flashes, consistent with the approach used in [11]. That is, 36 selections were made in each paradigm with flash sequences each containing ten target flashes for the CBP and in the ASP. Thus, both paradigms used 360 target flashes to acquire training data. We used the SWLDA algorithm to determine the signal features that best discriminated between target and non-target flashes (MATLAB version 7.6 R2007a, stepwise fit function).

For online classification, epochs from each stimulus item were averaged before applying the SWLDA classification coefficients. In both paradigms, the coefficients were applied to the specific spatiotemporal features of each of the 72 items and summed. The item with the highest score was selected and presented to the subject as feedback.

2.5. Determining the optimal number of flash periods

Due to the P300 response's relatively low signal-to-noise ratio, each item must be flashed multiple times and the results averaged [22]. During calibration, the number of target item flashes was constant across participants and presentation methods. During the online testing phase, we optimised the number of flash periods from each participant's maximum written symbol rate (WSR, or symbols/min; [23]) based on the results of their training sessions. Each of the six runs from the training data was evaluated using a classifier trained on the other five runs (i.e. leave one out) and the cumulative results were then evaluated using the resulting 'penalty adjusted accuracy' (described below in section 3.1) to determine the optimal number of flashes to maximise the WSR for each subject.

2.6. The software and hardware

The BCI2000 platform [15] has many advantages to use as a platform for performing BCI experiments. Although the precise timing required to support our paradigm required that we develop our own system, we continued to utilise BCI2000 to handle the recording of the data, the tracking of experimental parameters, and the display of the words to be spelled to the subject.

The precise frame-by-frame control required for the experiment was provided by a Linux system running Ubuntu 10.4 and OpenGL. Since the P300 spelling module on BCI2000 flashes stimuli in epochs and classifies epochs as opposed to asynchronously flashing individual targets, it was necessary to handle P300 response collection and classification on the Ubuntu system. In order to do this, BCI2000 was modified to send the classifier weights from the BCI2000 parameter files to the Ubuntu system prior to the beginning of each run. Further modifications were made to stream the incoming EEG including the digital marking channel from BCI2000 to Ubuntu. Additional information marking the start and end of each run and the beginning of each trial was also sent from BCI2000 to Ubuntu. All of this communication was handled by dedicated Gigabit Ethernet cards installed in each system with a dedicated crossover cable between the two to minimise latency and eliminate local LAN traffic. The UDP protocol was used to transfer information between the systems.

In the case of the CBP, classification and presentation were handled by the BCI2000 system. However the stimuli to be flashed were sent as a list of targets via UDP and whenever the BCI2000 system executed the display functions called ‘reveal’ or ‘conceal,’ an appropriate UDP packet was sent to mirror the event on the remote (Ubuntu) system. Prior to undertaking the study, careful experiments were conducted as described earlier to ensure that the jitter between the native BCI2000 display and the Ubuntu display was acceptably low and to determine the latency.

To provide the accurate frame-by-frame presentation required by the new paradigm, low-level graphics routines were provided to efficiently generate the frames. OpenGL was used to render these frames using the blocking invocation of the ‘swap buffers’ call. This was done to ensure that every frame would be synchronized to the monitor’s refresh rate and was necessary to ensure that frames were all presented on time in the correct order with no missing or duplicated frames. An interesting side-effect of this approach is that the refresh rate of the monitor directly controls the timing of the stimulus presentation, and therefore faster presentation of the stimuli may be accomplished by merely changing the refresh rate of the monitor from 60 Hz to 75 Hz. In our case, the monitor was set to 60 frames per second. The display used was an Acer x223w LCD monitor connected via a DVI cable to an AMD FirePro V5900 video card.

For the reader familiar with graphics programming, the presence of a FIFO in the video card isolates the timing of the frame transmission from the application. Although using the blocking swap buffers call will ensure that a new frame is presented after every vertical retrace, the call to the blocking swap buffers may block before the FIFO is completely full and might not unblock immediately just because there is one empty spot in the FIFO. As a result, the application has no way of knowing when each frame is actually displayed. This necessitated the provision of a small ‘dummy target’ on the Ubuntu monitor over which a photo-sensor was placed to record when each frame appeared. This ‘marker target,’ initially black, was set to begin toggling between black and white on every frame change starting at the beginning of each flash sequence and its output was digitised and sent to the digital marking channel input of the EEG amplifier. Although this information was not required for the CBP, we recorded it for both paradigms so that we would be able to reanalyse the results of the CBP offline to see how it would have behaved if the precise timing of the stimuli were known. Those results were described earlier.

3. Results

3.1. Accuracy and throughput

Tables 3 and 4 below show the number of flashes required by each subject in the two paradigms to reach their maximum throughput along with the accuracy and throughput achieved in an online test with feedback presented to the subjects. Table 3 shows the results of the ‘faceless’ versions of the paradigms while table 4 shows the ‘face-full’ version.

It is important for the reader to appreciate that the bit rates above would translate in many researchers’ work to values up to 10% higher if the raw accuracy was used to determine the throughput. In order to determine a more conservative estimate for the throughput for a subject in bits per minute, the practical bit-rate from tables 3 and 4 above was decreased according to the subjects’ raw accuracies adjusted as now described. In the event that a mistake is made in the context of a practical interface, the subject would have to select the backspace key and then make the correction. There is a probability that a further mistake could be made at either of these steps, resulting in additional corrections and so on and so forth ad infinitum. This leads to the infinite mathematical series of infinite products presented and solved in [11]. We refer to this as the ‘penalty adjusted accuracy,’ which despite its complicated origins works out to simply $2 \times p - 1$, where p is the raw accuracy of the subject. While using this metric will decrease the values of our throughput calculations, it represents a more realistic computation of the true throughput we can expect a subject to achieve when using the speller in a practical way. The throughput of a subject is then calculated as the 100% throughput (taken from table 2) multiplied by that subject’s ‘penalty adjusted accuracy’. Of course the subject’s reported accuracy is still appropriately reported as the raw accuracy since that is indeed a measure which reflects how many errors we expect a subject to make for any given string of characters to be spelled. Only the throughput should be adjusted according to the ‘penalty adjusted accuracy.’ This metric may be used to determine how many selections, M, a subject is expected to make in order to successfully communicate N selections. Specifically, $M = N/(2 \times p - 1)$.

3.2. Waveform morphology

NB: we have chosen to show our waveforms mathematically ‘right side up’. We have noted a trend of researchers beginning to adopt this approach and in support of that we would like to also use this convention.

The present work is intended on focusing on the introduction of the ASP. Therefore we will not undertake a detailed analysis of the faces paradigm here. Although we will present the waveforms for the ‘face-full’ versions of our experiments, we will only present a brief analysis of those waveforms later and will make a few relevant comments concerning its effect. Our discussion will focus on comparing the CBP and ASP. Figures 7 and 8 below depict the associated waveforms for the ‘faceless’ and ‘face-full’ paradigms respectively.

In all cases the photo-sensor marking channel was used to provide accurate timing information to produce response waveforms. The CBP and the ASP produced very similar waveforms. We focused our analyses on the four electrodes, (Cz, Pz, Po7, and Po8), that reliably capture most of the P300 energy in BCI applications [20, 24]. Figures 7 and 8 depict the averaged target waveforms for each of the 10 participants. We averaged these data across the 36 item selections of the calibration phase in order to keep the amount of data contributing to each average constant across all participants.

Table 3. Number of flashes and time taken to reach peak throughput in each paradigm without using faces. NB: the time includes the three seconds of dead-time between selections which must be incorporated into the calculation of practical bit rate.

	CBP			ASP		
Subject	Flashes	Accuracy	Throughput	Flashes	Accuracy	Throughput
S1	8	94.4%	21.9 bpm	5	94.4%	48.7 bpm
S2	2	94.4%	54.8 bpm	2	94.4%	74.2 bpm
S3	2	77.8%	34.3 bpm	3	94.4%	64.1 bpm
S4	2	88.9%	48.0 bpm	3	94.4%	64.1 bpm
S5	3	80.6%	30.2 bpm	4	86.1%	45.2 bpm
S6	3	77.8%	27.4 bpm	7	97.2%	40.7 bpm
S7	3	91.7%	41.1 bpm	4	91.7%	52.1 bpm
S8	4	91.7%	34.3 bpm	5	94.4%	48.7 bpm
S9	3	91.7%	41.1 bpm	5	97.2%	51.8 bpm
S10	2	86.1%	44.6 bpm	4	88.9%	48.7 bpm
Mean	3.2	87.5%	37.8 bpm	4.3	93.3%	53.8 bpm

Table 4. Number of flashes and time taken to reach peak throughput in each paradigm using faces instead of traditional flashing. NB: the time includes the three seconds of dead-time between selections which must be incorporated into the calculation of practical bit rate.

	CBP with faces			ASP with faces		
Subject	Flashes	Accuracy	Throughput	Flashes	Accuracy	Throughput
S1	2	88.9%	48.0 bpm	2	94.4%	74.2 bpm
S2	2	100.0%	78.9 bpm	2	100.0%	83.5 bpm
S3	2	97.2%	58.3 bpm	2	94.4%	74.2 bpm
S4	2	97.2%	58.3 bpm	2	100.0%	83.5 bpm
S5	4	100.0%	41.1 bpm	3	91.7%	60.1 bpm
S6	2	97.2%	58.2 bpm	3	100.0%	72.1 bpm
S7	2	91.7%	51.4 bpm	2	97.2%	78.9 bpm
S8	4	97.2%	38.8 bpm	3	94.4%	64.1 bpm
S9	3	97.2%	46.6 bpm	3	100.0%	72.1 bpm
S10	3	100.0%	49.4 bpm	2	88.9%	64.9 bpm
Mean	2.6	96.7%	51.2 bpm	2.4	95.8%	72.3 bpm

Figures 7 and 8 include, for each of the four electrodes, the grand means.

3.3. Error analysis

The average number of errors made by subjects in which the incorrect selection flashed once or never with the target was similar in both paradigms, 2.9 in ASP and 3.2 in CBP, $\rho = 0.18$, however, there were an additional 1.5 errors per subject on average in the CBP where the incorrectly selected target had flashed more than once with the target.

The peak amplitude of the P300 response in the ASP was 18% lower ($\rho = 0.008$) than in the CBP and this was likely due to the decreased average target-to-target spacing in the ASP (890 ms) compared with the CBP (1500 ms). This is consistent with the findings reported in [12] where faster flash rates result in lower amplitude responses.

3.4. Final results

As a final test of the capabilities of the faces version of the ASP, a few subjects agreed to return to the laboratory to do some 'free spelling.' The following protocol was used:

(i) The subjects practiced copy spelling a complete sentence with their optimal sequence length using the faces version of the ASP twice to generate weights. Given that each run only contained two flashes per selection for the subjects tested, the task was repeated twice to collect sufficient training data to train the classifier. No feedback was provided to the subjects at this point.
(ii) A classifier trained on the training data was then used for one free spelling trial test conducted to allow the subjects an opportunity to practice the task with feedback once before being 'formally tested'.
(iii) The subjects were then told that the remaining runs would constitute a final test of their performance and that the dead time between selections would be

Table 5. Timing parameters and practical throughput for final free spell test. Results were based on the subjects' penalty adjusted accuracies. The characters were selected from an 8 by 9 matrix containing 72 items.

Subject	Time between selections	Flash time	Total time (mm:ss)	Number of selections	Selections per minute	Accuracy	Throughput (bpm)
S2	2.16 s	1.43 s	3:43	62	16.71	100%	103
S2	1.88 s	1.43 s	3:25	62	18.13	97%	105
S3	2.16 s	1.43 s	3:48	64	16.84	95%	94
S4	2.16 s	1.43 s	3:57	66	16.71	100%	103
S4	1.88 s	1.43 s	3:38	66	18.13	100%	112
S4	1.66 s	1.43 s	3:24	66	19.42	100%	120
S7	2.16 s	1.43 s	3:28	58	16.71	97%	96

incrementally decreased after each attempt until the subject was no longer able to achieve 100% accuracy.

With the subjects' permissions, we are including the string that each attempted to free spell.

Subject S2:
I AM DANNY REID, DIRECTOR OF INFORMATION TECHNOLOY AT ALGOMA.
Subject S3:
I AM JESSICA SHANAHAN, BCI LAB COORDINATOR AT ALGOMA UNIVERSITY.
Subject S4:
I AM SARAH CROWELL, STUDENT SERVICES ADVISOR AT ALGOMA UNIVERSITY.
Subject S7:
I AM JEREMY WILHELM, DIRECTOR OF PHYSICAL PLANT AT ALGOMA.

The practical bit rates attained for the subjects tested appear below in table 5. One subject was tested three times with progressively decreasing dead time between selections and made no errors even at the shortest possible setting. Ironically, that subject reported being more comfortable as the dead time was decreased claiming that the task 'flowed more naturally' as the speed was increased. NB: the results reported in table 5 are not theoretical, but rather they are actual practical bit rates attained by the subjects. Furthermore, the throughput reported has been reduced according to each subject's 'penalty adjusted accuracy'. The number of selections includes blanks and punctuation.

4. Discussion

Although we ran the 'faceless' version of the paradigms prior to the 'face-full' version for all subjects, the effectiveness of introducing faces has already been confirmed and studied elsewhere in depth [3–5, 25, 26]. Therefore the success of applying faces to both the CBP and ASP is unrelated to the order the experiments were performed and is consistent with these previous studies. What we can claim is that the throughput of the ASP remains consistently higher than that of CBP when faces are added, and that both paradigms improve significantly with the introduction of faces.

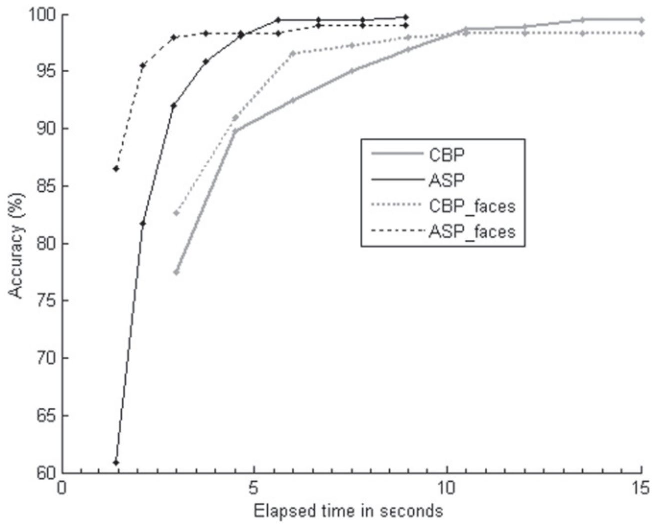

Figure 5. Mean accuracy (%) of the subjects versus elapsed time. The dots on the lines from left to right represent flashes 2 through 10 respectively.

4.1. Accuracy and throughput

Typically, it is of interest to show how the subjects' accuracies increased from flash to flash graphically as accuracy versus flash number. In our case, however, the time taken for each flash period is not only different between the CBP and ASP but is constant in the former and variable in the latter. Because of this, we have elected to present these results in terms of the amount of time which has elapsed since the start of the flash sequence rather than by flash number. The results appear in figure 5 above. Note that in order to generate this data it was necessary to use the training data so that all subjects would have ten flashes worth of data. The results were therefore obtained by using an offline 'leave one out' analysis where the classifier was trained on five of the six runs and tested on the remaining run. This was repeated for all six training runs and then the results were averaged together.

Since we expect subjects to eventually reach 100% accuracy with sufficient flashes, we are more concerned with how high their throughput is and where it peaks. This has been graphed in figure 6 below. For emphasis, please note that it is based on the stricter 'penalty adjusted accuracy' discussed above and not on the raw accuracy. This considerably

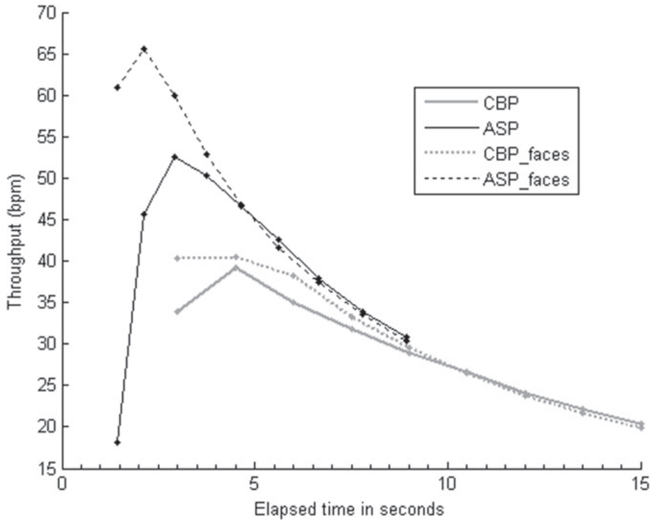

Figure 6. Mean throughput (bpm) of the subjects versus elapsed time and flash number. The dots on the lines from left to right represent flashes 2 through 10 respectively.

decreases the throughput, but as pointed out produces more realistic results. Also, this is the practical bit rate and therefore the results include the 3 s of dead time between selections. Once again, this is based on an offline analysis of the training data using 'leave one out' where the classifier was repeatedly trained on five of the six runs and tested on the remaining run.

The gold standard in measuring the performance of a BCI comes from online tests with feedback. Although it is useful to see the performance over all subjects in a consistent way as above in figures 5 and 6, it is more important to consider how each subject was able to perform online. Since different subjects require different numbers of flashes, it is difficult to produce a meaningful graph consolidating the results as above, therefore we will report on the differences in the average performance of the subjects in each of the test paradigms. The subjects overall performances are given in tables 3 and 4 above. We examined the individual testing runs to provide more data to analyse by paired t-test to determine the statistical significance of these averages.

The ASP accuracy of 93% represented a 6% improvement over the CBP accuracy of 88% ($\rho = 0.005$) when traditional flashing was used. The addition of faces improved the CBP accuracy by 10% increasing it from 88% to 97% ($\rho = 0.002$). However adding faces made no significant difference in the accuracy of the ASP (93% without faces and 96% with faces, $\rho = 0.11$). In the faces versions of ASP and CBP there was no significant differences in accuracy observed (96% and 97% respectively, $\rho = 0.84$).

With traditional flashing, the average throughput of the ASP (54 bpm) represented an increase of 42% ($\rho = 0.000\,000\,005$) over that of the CBP (38 bpm), based on the more strict 'penalty adjusted accuracy' of the subjects. When faces were added, the throughput of the ASP (73 bpm) remained higher (by 46%, $\rho = 6 \times 10^{-13}$) than in the CBP (50 bpm). To put these results in perspective, the ASP throughput here calculated the traditional way would be 79 bpm. Furthermore, the 'penalty adjusted accuracy' becomes more and more severe as the accuracy falls so the throughput of Subject 6 in the regular CBP of 27.4 bpm (who had the lowest accuracy of 77.8%) would be 38.4 bpm if calculated the traditional way.

4.2. Waveform morphology

We examined amplitude and latency differences between targets in each paradigm at each electrode location by paired t-test.

For the averaged target responses located at electrode Cz, we found no significant differences in the latency of the positive peak (approximately 200 ms) between the CBP and the ASP, $\rho = 0.87$. However, we found that the average amplitude of the peak was 17% lower for the ASP, 3.46 μV, than for the CBP, 4.09 μV, $\rho = 0.03$. For the target responses located at electrode Pz, we found that the amplitude of the positive peak (also at approximately 200 ms) was 17% lower for the ASP, 3.06 μV, than for the CBP, 3.69 μV, $\rho = 0.02$. We did not observe any significant differences between the two paradigms at electrode location PO7 ($\rho = 0.34$) nor at location PO8 ($\rho = 0.25$). Similar differences in amplitudes and similarities in latencies over these four electrodes also existed between the 'face-full' versions of the two paradigms.

In comparing the 'faceless' versions of the two paradigms to their 'face-full' counterparts, there was a significant difference in electrodes Cz and Pz. The late negative peak occurring at approximately 500 ms in the faceless paradigms shifted to a much sharper peak with similar amplitude to approximately 300 ms, consistent with the results reported in [16]. In the present work, it is of particular interest that the amplitude of this 300 ms negative peak is significantly more pronounced ($\rho = 0.01$) for the ASP, -1.4 μV than for the CBP, 0.2 μV. This is most likely the explanation for why the required number of flashes is similar in the 'faces' versions of the CBP and ASP, i.e. the lower amplitude of the positive peak at 200 ms in the ASP is compensated for by a higher negative peak at 300 ms. The presence of this sharp peak in both paradigms reduced the number of flashes required compared to their 'faceless' counterparts.

4.3. Error analysis

There was no statistically significant difference in the numbers of errors that occurred in both paradigms where the incorrectly selected target flashed less than twice with the correct target, however there were an additional 47% errors present in the CBP all in which the incorrectly selected item flashed two or more times with the correct target. This provides evidence that constraining the number of times the same items are grouped together has a positive effect on performance.

The lower amplitude of the P300 responses and the higher number of flashes required in the ASP were both more than compensated for by eliminating multiple groupings of

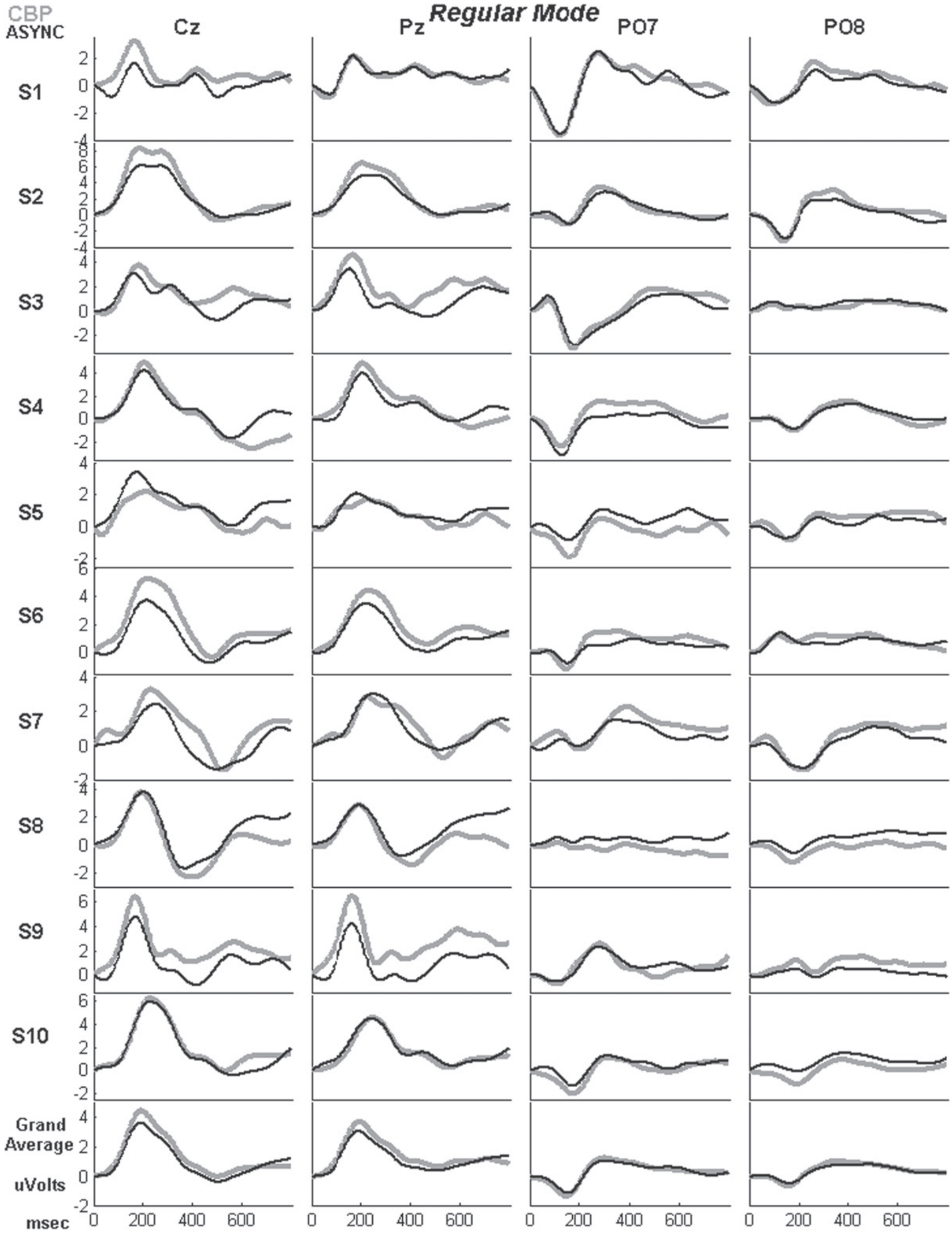

Figure 7. Waveforms for the 'faceless' versions (regular mode) of the CBP (grey lines) and the ASP (black lines). NB: this EEG is shown with positive up, negative down according to the traditional mathematical convention rather than the historically inspired EEG convention.

the same targets together and attaining higher accuracy in the ASP. On average, subjects required more flashes to reach their maximum throughput in the ASP compared with the CBP (4.9 and 3.2 respectively, $\rho = 0.006$) when faces were not used. In the faces variations however, there was no statistically significant difference in the number of flashes required (3.6 in CBP and 3.5 in ASP, $\rho = 0.60$). Although the average time taken to reach the peak throughput was similar for the subjects (7.8 s in CBP and 6.8 s in ASP, $\rho = 0.08$), the improved accuracy in ASP resulted in higher throughput values. Interestingly, when faces were added, the mean difference in the time for maximum throughput between CBP and ASP became significant (8.4 s and 5.7 s respectively, $\rho = 0.003$).

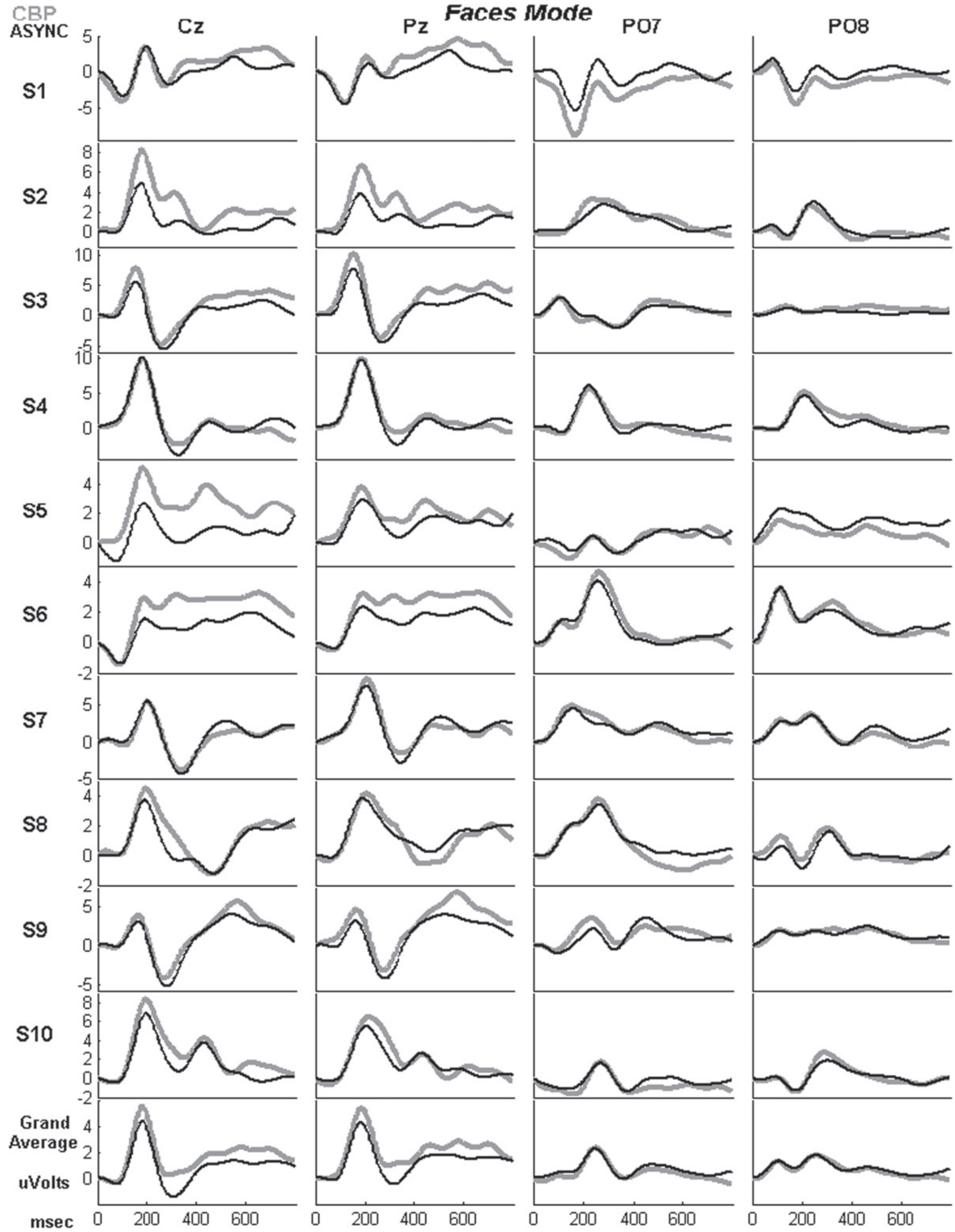

Figure 8. Waveforms for the 'faces' versions of the CBP (grey lines) and the ASP (black lines). NB: this EEG is shown with positive up, negative down according to the traditional mathematical convention rather than the historically inspired EEG convention.

4.4. Final results

Unfortunately, we did not anticipate that the best subject returning to free-spell would be able to continue to reach 100% accuracy even at the fastest speed our implementation provided, so we were unable to determine the extent of their practical upper limit. However the fastest speed we observed was 120 bpm. As far as we are aware, no other P300 speller has ever been able to achieve this level of practical throughput to date. The average time taken over the three tests for the best subject to spell all 66 characters of the sentence was three minutes and forty seconds. The subject's theoretical upper bound on throughput (i.e. with zero dead time between

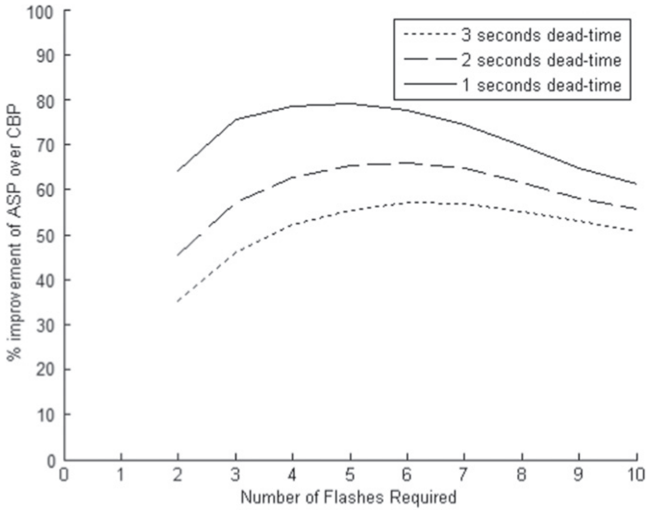

Figure 9. Improvement of ASP over CBP as a function of the number of flashes required to achieve maximum throughput.

selections based on the 'flash time' only) was 258 bits per minute. The mean 'penalty adjusted' practical bit rate for all four subjects tested on free spelling was 104 bpm. The mean accuracy was 97%.

5. Conclusion

Note that the ASP is only one possible application of extending performance guided constraints to the temporal domain. The present work not only captures all previous temporal organisations (including RCP and CBP) but all possible ways of organising the flashing of items that might be invented in the future. To be clear, the present study is more general than the particular implementation of the ASP that we have specifically presented and explored in this paper. Performance guided constraints now applied to both the spatial and temporal domains captures all possible ways of presenting items to be grouped together and flashed.

Prior to this research, it seemed as though a ceiling had been reached with the P300 speller. Now, with the introduction of the ASP the ceiling has been raised, and we can begin pushing this ceiling further, integrating other research such as employing task-oriented optimal designs described by Zhou *et al* in [27].

Given the difference in time taken to present flash sequences between the CBP and ASP, there suddenly becomes a great motivation to reduce the dead-time between sequences. In figure 9 above, we show the theoretical improvement in performance possible as a percentage increase of ASP over CBP for 2 through 10 flashes with dead-time between selections of 3 s, 2 s and 1 s.

If a subject is able to navigate between selections in one second in both paradigms, then the ASP has the potential to be approximately 80% faster than the CBP.

5.1. Suggestions for implementation

In order to replicate the results described here, it is important to ensure that the display is doing what was intended. Using a low latency DVI-connected monitor is a prerequisite of a successful implementation. Many monitors update pixels while the serialized video signal is arriving and this is unacceptable since items in the top left corner of the screen are refreshed before those in the bottom right corner. It is also imperative that the software preparing the images be as efficient as possible and must use the equivalent of the blocking 'swap buffers' call in OpenGL to ensure that there are no missing and no duplicated frames. We highly recommend that during development multiple photo-sensors are employed and that tests be conducted repeatedly using them to ensure that the monitor is displaying things on time, in the correct order, and consistently across the entire physical screen. To reiterate, we used an Acer x223w LCD monitor connected via a DVI cable to an AMD FirePro V5900 video card. We confirmed that the screen refreshes in such a way that the disparity between the appearances of items on different parts of the screen never exceeds more than 2 ms when a DVI cable is used. (The same monitor and video card refreshed the screen in pixel order over the time-course of a frame when a VGA cable was used leading to a disparity approaching 17 ms between different parts of the screen.)

If done correctly, the vertical retrace of the monitor will drive the generation of the images from the software thereby ensuring that every frame is displayed at the correct time. It is also necessary to use a photo-sensor to record every frame update directly in the EEG to ensure synchronization between EEG acquisition and the display.

Note that we are using the more restrictive 'penalty adjusted accuracy' when calculating throughput throughout this paper. The rationale and mathematical development of this metric is explained in [11]. We have seen this metric being used recently in other researchers' work, and we encourage others to begin using it. Although it diminishes the results reported, it is a realistic measure of what your system is actually capable of delivering to a user. We have also seen a trend towards presenting EEG 'right side up' consistent with mathematical graphing methods rather than in the historically inspired 'upside-down' orientation. We have elected to break with the old tradition in this paper and encourage others to follow.

Acknowledgments

The authors would like to thank the Wadsworth BCI Lab for their ongoing financial support, and for providing feedback during the development of this project. The first author would like to sincerely thank his coauthor for her loyalty to the Algoma University BCI lab over the past several years and the fantastic job she did at implementing and debugging the ASP system. And, of course, both authors would like to thank all the volunteers who participated in our study.

References

[1] Townsend G *et al* 2012 A general P300 brain–computer interface presentation paradigm based on performance guided constraints *Neurosci. Lett.* **531** 63–8
[2] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
[3] Zhang Y *et al* 2012 A novel BCI based on ERP components sensitive to configural processing of human faces *J. Neural Eng.* **9** 026018
[4] Kaufmann T and Kübler A 2014 Beyond maximum speed—a novel two-stimulus paradigm for brain–computer interfaces based on event-related potentials (P300-BCI) *J. Neural Eng.* **11** 056004
[5] Chen L *et al* 2015 A survey of the dummy face and human face stimuli used in BCI paradigm *J. Neurosci. Methods* **239** 18–27
[6] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[7] Takano K, Komatsu T, Hata N, Nakajima Y and Kansaku K 2009 Visual stimuli for the P300 brain–computer interface: a comparison of white/gray and green/blue flicker matrices *Clin. Neurophysiol.* **120** 1562–6
[8] Salvaris M and Sepulveda F 2009 Visual modifications on the P300 speller BCI paradigm *J. Neural. Eng.* **6** 046010
[9] Hong B, Guo F, Liu T, Gao X and Gao S 2009 N200-speller using motion-onset visual response *Clin. Neurophysiol.* **120** 1658–66
[10] Fazel-Rezai R 2007 Human error in P300 speller paradigm for brain–computer interface *Conf. Proc.: IEEE Engineering in Medicine and Biology Society 2007* pp 2516–9
[11] Townsend G, LaPallo B K, Boulay C B, Krusienski D J, Frye G E, Hauser C K, Schwartz N E, Vaughan T M, Wolpaw J R and Sellers E W 2010 A novel P300-based brain–computer interface stimulus presentation paradigm: moving beyond rows and columns *Clin. Neurophysiol.* **121** 1109–20
[12] McFarland D J, Sarnacki W A, Townsend G, Vaughan T and Wolpaw J R 2011 The P300-based brain–computer interface (BCI): effects of stimulus rate *Clin. Neurophysiol.* **122** 731–7
[13] De Vos M, Kroesen M, Emkes R and Debener S 2014 P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier *J. Neural Eng.* **11** 036008
[14] McCane L M *et al* 2015 P300-based brain–computer interface (BCI) event-related potentials (ERPs): people with amyotrophic lateral sclerosis (ALS) versus age-matched controls *Clin. Neurophysiol.* S1388–2457 **126** 2124–31
[15] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain–computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43
[16] Kaufmann T *et al* 2011 Flashing characters with famous faces improves ERP-based brain–computer interface performance *J. Neural Eng.* **8** 056016
[17] Jin J *et al* 2012 The changing face of P300 BCIs: a comparison of stimulus changes in a P300 BCI involving faces, emotion, and movement *PLoS ONE* **7** e49688
[18] Jin J *et al* 2014 An ERP-based BCI using an oddball paradigm with different faces and reduced errors in critical functions *Int. J. Neural Syst.* **24** 1450027
[19] Sharbrough F C G, Lesser R P, Lüders H, Nuwer M and Picton W 1991 AEEGS guidelines for standard electrode position nomenclature *Clin. Neurophysiol.* **8** 202–4
[20] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21
[21] Draper N R and Smith H 1981 *Applied Regression Analysis* 2nd edn (New York: Wiley)
[22] Cohen J and Polich J 1997 On the number of trials needed for P300 *Int. J. Psychophys.* **25** 249–55
[23] Furdea A *et al* 2009 An auditory oddball (P300) spelling system for brain–computer interfaces *Psychophysiology* **46** 617–25
[24] Kaper M, Meinicke P, Grossekathoefer U, Lingner T and Ritter H 2004 BCI competition 2003—data set IIb: support vector machines for the P300 speller paradigm *IEEE Trans. Biomed. Eng.* **51** 1073–6
[25] Kaufmann T *et al* 2011 Flashing characters with famous faces improves ERP-based brain–computer interface performance *J. Neural Eng.* **8**
[26] Jin J *et al* 2014 An ERP-based BCI using an oddball paradigm with different faces and reduced errors in critical functions *Int. J. Neural Syst.* **24** 1450027
[27] Zhou Y *et al* 2014 A novel task-oriented optimal design for P300-based brain–computer interfaces *J. Neural Eng.* **11** 056003