An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli

2012 J. Neural Eng. 9 026011

(http://iopscience.iop.org/1741-2552/9/2/026011)

IOP PUBLISHING JOURNAL OF NEURAL ENGINEERING

J. Neural Eng. **9** (2012) 026011 (13pp) doi:10.1088/1741-2560/9/2/026011

An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli

N J Hill[1] and B Schölkopf[2]

[1] Wadsworth Center, New York State Department of Health, Albany, NY, USA
[2] Max Planck Institute for Intelligent Systems, Tübingen, Germany

E-mail: njh07@wadsworth.org and bs@tuebingen.mpg.de

Received 28 August 2011
Accepted for publication 20 December 2011
Published 15 February 2012
Online at stacks.iop.org/JNE/9/026011

Abstract

We report on the development and online testing of an electroencephalogram-based brain–computer interface (BCI) that aims to be usable by completely paralysed users—for whom visual or motor-system-based BCIs may not be suitable, and among whom reports of successful BCI use have so far been very rare. The current approach exploits covert shifts of attention to auditory stimuli in a dichotic-listening stimulus design. To compare the efficacy of event-related potentials (ERPs) and steady-state auditory evoked potentials (SSAEPs), the stimuli were designed such that they elicited both ERPs and SSAEPs simultaneously. Trial-by-trial feedback was provided online, based on subjects' modulation of N1 and P3 ERP components measured during single 5 s stimulation intervals. All 13 healthy subjects were able to use the BCI, with performance in a binary left/right choice task ranging from 75% to 96% correct across subjects (mean 85%). BCI classification was based on the contrast between stimuli in the attended stream and stimuli in the unattended stream, making use of every stimulus, rather than contrasting frequent standard and rare 'oddball' stimuli. SSAEPs were assessed offline: for all subjects, spectral components at the two exactly known modulation frequencies allowed discrimination of pre-stimulus from stimulus intervals, and of left-only stimuli from right-only stimuli when one side of the dichotic stimulus pair was muted. However, attention modulation of SSAEPs was not sufficient for single-trial BCI communication, even when the subject's attention *was* clearly focused well enough to allow classification of the same trials via ERPs. ERPs clearly provided a superior basis for BCI. The ERP results are a promising step towards the development of a simple-to-use, reliable yes/no communication system for users in the most severely paralysed states, as well as potential attention-monitoring and -training applications outside the context of assistive technology.

(Some figures may appear in colour only in the online journal)

1. Introduction

The aim of research into brain–computer interfaces (BCIs) is to develop systems that allow a person to interact with his or her environment using signals from the brain, without the need for any muscular movement or peripheral nervous system involvement—for example, to allow a completely paralysed person to communicate. Total or near-total paralysis can result in cases of brain-stem stroke, cerebral palsy and amytrophic lateral sclerosis (ALS, also known as Lou Gehrig's disease), among other disorders. It has been shown [1] that some people in a 'locked-in' state (LIS), in which most cognitive functions are intact despite almost-complete paralysis, can learn to communicate via an interface that interprets electrical signals from the brain, measured externally by electroencephalogram (EEG). However, for people in the so-called completely locked-in or totally locked-in state (CLIS or TLIS), in which absolutely no communication is possible

Table 1. Auditory ERP-based BCI studies.

Study	Interface design	Stimulus arrangement	Analysis
Hill *et al* [8]	Binary choice	*Streaming*	Offline
Sellers and Donchin [9]	4-way choice	Sequential	Offline
Furdea *et al* [10, 17]	5+5-choice speller	Sequential	*Online*
Klobassa *et al* [11]	6+6-choice speller	Sequential	*Online*
Kanoh *et al* [12]	Binary choice	*Streaming*	Offline
Halder *et al* [13]	Binary choice	Sequential	Offline
Schreuder *et al* [18, 14]	6+6-choice speller	Sequential	Offline
Schreuder *et al* [19]	6+6-choice speller	Sequential	*Online*
Belitski *et al* [20]	6+6-choice speller	Sequential	*Online*
Höhne *et al* [21, 15]	9-choice speller	Sequential	*Online*

via muscular movement, successful communication even via BCI has proved more elusive [2]—although there have been some encouraging early reports of success [3].

There is, therefore, still considerable room for development of BCI systems targeted at those people in the most-severely paralysed states, who need the technology most. The majority of BCI studies have so far been devoted to one of two approaches: the first approach is the exploitation of event-related potentials (ERPs) in response to visual stimuli [after 4, 5]. However, this is rather unsuitable for people in TLIS, whose eye movements are uncontrollable or entirely absent—among other problems, the inability to direct their gaze, focus to the desired depth or blink their eyes to prevent corneal disease and eye infections, all tend to add up to very poor spatial vision or none at all. The second popular approach is based on signals from the motor and pre-motor cortex in response to imagined muscle movements. However, since TLIS often results from progressive degeneration of the motor system, it is still unclear for how long users in TLIS, having hypothetically made the breakthrough of using such a system successfully in TLIS, might continue to be able to rely on motor-system signals in this way.

Hence there is considerable motivation to continue exploring BCI modalities that have been relatively little explored in the literature: those based on non-motor mental tasks such as the mental calculation and music imagery tasks used by Naito *et al* [3], or those based on attention to tactile stimuli [e.g. 6, 7] or auditory stimuli [8–15]. The current design is a development of the first such auditory approach to be published [8]. It is based on voluntary shifts of attention in a two-stream (and in the current paper, dichotic) listening task. Two sequences or 'streams' of auditory stimuli are played simultaneously, and the user may make a binary decision by focusing on one stream and ignoring the other. This leads to a modulation of ERPs in response to the stimuli of the two streams—an effect reported in 1973 by Hillyard *et al* [16], which Hill *et al* [8] showed could be classified on a single-trial basis for potential use in the BCI.

1.1. BCIs driven by auditory ERPs

Various auditory-ERP-based BCI approaches have been reported since 2005. In table 1, these are categorized according to whether they used as a *streaming* or *sequential* stimulus arrangement.

In streaming approaches, users may or may not be asked to monitor the attended stream for particular (relatively infrequent) target stimuli. However, the BCI is driven not by the contrast of target responses versus non-target responses, but rather by the contrast between responses to stimuli in the attended stream versus responses to stimuli in the unattended stream. The stimuli used for this contrast could be non-targets, targets (if present in the design) or both.

In sequential approaches, by contrast, the user monitors a single stream for a target stimulus, and it is the brain response to the target that carries the crucial information. This has a disadvantage in the two-class case that the system must wait longer between information-bearing stimuli. However, it has the distinct advantage that the stream may consist of a large number of different targets, allowing a multi-way choice. Two or more multi-way choices, made in succession, allow a letter to be selected in a spelling application, provided the user is sufficiently familiar with the layout of letters in, for example, a grid [10, 17, 11, 15] or a nested pattern of hexagons [14]. For a subject able to use both systems, a fully fledged spelling application is clearly superior to a binary chooser considered in isolation. For now, however, our aim is not to design a full speller, but rather to create a reliable binary interface that places little demand on working memory. The long-term aim is that this might be useful for re-establishing simple, initial contact with a person who has entered the totally locked-in state—either as a basis for, or as a stepping-stone towards, more sophisticated communication. Hence, we continue to pursue the streaming method.

Table 1 also categorizes the studies as presenting either 'offline' or 'online' analyses. Both of the streaming studies, and some of the sequential studies, assessed classification performance offline (i.e. by dividing data into training and test sets after all the data were gathered). A vital step in developing such methods for BCI is to ensure that the system works online (i.e. that the system can act on its interpretation of a decision made by the user, and report this to the user, in time for the user to make the next decision). One function of this paper is therefore to provide an in-depth assessment of *online* performance of the streaming approach. The second goal is to use the online attention paradigm as a platform for investigating the usefulness of a second class of brain signal, as described in the following section.

1.2. Steady-state evoked potentials

A very different approach to auditory stimulus-driven BCI was attempted by Kallenberg [22] using a streaming design, and by Farquhar *et al* [23] using both streaming and sequential designs. Here, the focus was on a different class of brain responses known as steady-state auditory-evoked potentials (SSAEP) or auditory steady-state responses (ASSR). These are sustained responses to continuous, fluctuating stimuli [24]. They are typically elicited by trains of click stimuli, tone pulses or amplitude-modulated tones, with a repetition or modulation rate between 20 and 100 Hz. The resulting brain response can be localized in the primary (and, for lower frequencies, also secondary) auditory cortex [25] and are frequency-matched and phase-locked to the modulation. Ross *et al* [26] found that the largest signal-to-noise ratio was produced by modulation frequencies around 40 Hz, and indeed this is the frequency typically used in many SSAEP studies.

For BCI use, first the signal needs to be modified by the user's voluntary shifts of attention; second, this modulation needs to be detectable on a single-trial basis, where a 'trial' lasts some reasonably small number of seconds. Steady-state evoked potentials (SSEPs) in other modalities have been shown to meet these criteria. Visual SSEPs are well established as a basis for the BCI, with users able to modulate them by overt [27] or covert [28–30] shifts of attention to spatial [27, 28] or non-spatial [29, 30] aspects of a stimulus array. Promising results have also been shown for somatosensory SSEPs [6].

Auditory SSEPs, however, have had more difficulty in living up to this promise. A 1987 EEG study [31] failed to find any significant attention modulation of SSAEPs at all, and it took another 17 years for a measurable effect to be found: first in MEG using a cross-modal paradigm [32], then using pure-auditory streaming designs in ECoG [33] and MEG [34], and finally in EEG using a sequential design [35].

In their attempts at a single-trial classification, both Kallenberg [22] and Farquhar *et al* [23] report performance that is mostly below 65%. This is well below the level of performance at which one can expect to construct any kind of BCI system for independent use. Lopez *et al* [36] also reported significant attention modulation of SSAEPs in their offline analysis of a BCI-like experiment, but concluded that the effect may still be too weak for practical use since the time required per trial (over 40 s) was excessively long. Most recently, Kim *et al* [37] have shown slightly more encouraging preliminary results from an SSAEP-based BCI system using 20 s stimuli (however, see section 4). Note that offline performance levels in all these studies were significantly above chance: they do at least, therefore, add to the evidence that SSAEPs can be modulated by attention *at all*. However, for BCI, which requires high classification accuracy in a short time, the results seem discouraging.

Nonetheless, it is impossible to draw definitive conclusions from studies that fail to find a large effect, since any such failure can be ascribed to a large number of potential factors. For example, these previous studies did not verify, in any independent and objective way, the extent to which subjects were able to focus their attention on one stimulus and ignore the other. It may be, therefore, that some aspect of the stimulus or task design made it difficult for subjects to shift their attention optimally. The current study aims to control for this possibility: it uses the existing auditory ERP-based BCI design to confirm that subjects can indeed modulate their attention in a single-trial-classifiable way, while at the same time, and in the same stimuli, introducing frequency 'tagging' to examine the effects of attention on SSAEPs.

2. Methods

2.1. Subjects

Subjects were 13 healthy participants (nine males, four females) with an age range of 23–37 years (27.8 ± 4.6). All subjects were right-handed, had corrected-to-normal vision and no history of significant hearing defects. They had answered public advertisements for experimental subjects, had given informed consent and were paid for their participation. Experiments were performed at, and approved by, the Max Planck Institute for Biological Cybernetics. No subjects were excluded from the analysis.

2.2. Stimuli and task design

One *trial* was defined as the attempt to make one binary choice (left or right). One *block* consisted of 20 trials and lasted about 5 min. After each block, the subject could rest for a few minutes if they so desired. In a single sitting lasting 2 h (excluding setup), each subject performed ten 20-trial blocks of the normal *attention* condition for a total of 200 attention trials. For comparison with the studies of Kallenberg [22] and Farquhar *et al* [23], a *perception* condition was included: here, the instructions and tasks are the same, as are the stimuli except in that the unattended stream is silent. Two 20-trial blocks of the perception condition were performed: one right at the beginning of the measurement period (which helped in introducing the task to the subject) and one half-way though the session between attention blocks for a total of 40 perception trials per subject.

At the start of each trial, subjects were given a visual cue (either the word 'LEFT' or the word 'RIGHT' appearing in the centre of the screen for 2 s) instructing them which stream to attend to. Half the trials in a given block were left, and half were right, in random order. The very last block was an exception to this: it was a *free-choice* attention block in which, instead of an explicit instruction, the word 'CHOOSE' appeared for 3 s, during which subjects decided freely whether to choose left or right, and wrote their choice down on paper. The purpose of this was to verify, for the benefit of both subject and experimenter, that the system's classification performance must be based on the EEG input alone.

Subjects were instructed that from the moment the cue appeared, they should keep their gaze fixed on the centre of the screen, and refrain as much as possible from blinking, swallowing or moving. After 2 s (or 3 in the free-choice block), the cue was replaced by a fixation cross and the sound stimulus began.

The stimulus is illustrated in figure 1. It lasted 5 s in total including 250 ms attack and decay periods. The stimulus was dichotic: a different stimulus stream was presented to each ear. Each stream consisted of an anti-aliased sawtooth carrier wave (500 Hz on the left, 769.231 Hz on the right), amplitude-modulated to 100% depth by a sine wave (41.667 Hz on the left, 38.462 Hz on the right). For most of the time, the peak-to-peak amplitude of the stimulus was at 30% of the soundcard's maximum output. However, starting at 504 ms on the left, and 598 ms on the right, the stimulus began to 'pulse': that is, with a raised-cosine attack of 5 ms, a plateau lasting 45 ms and a decay of 50 ms, the amplitude of the stimulus was raised to 100% output. These pulses were repeated with a period of 504 ms (left) and 546 ms (right) for a total of eight on the left and seven on the right. The pulses were designed to elicit ERPs, analogously to the beeps of Hill *et al* [8].

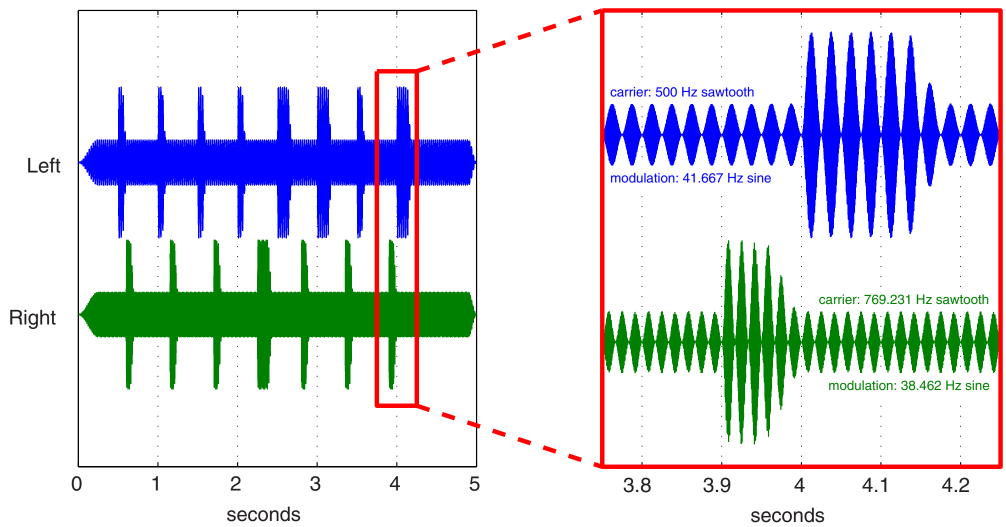

Figure 1. An example of the dichotic stimulus used on each trial. Amplitude-modulation at close to 40 Hz induces ASSRs, whereas the periodic 'pulsing' induces ERPs. Target pulses for the counting task are longer than standard pulses (in this example, there are three targets in the left stream and one in the right).

These parameters were chosen by hand over the course of several exploratory parameterizations and pilot experiments to meet multiple criteria: AM frequencies should be as close as possible to 40 Hz (in order to produce measurable SSAEP responses from as many subjects as possible) while still being distinguishable from each other in the EEG; AM cycles should last an integer number of EEG samples (to aid in analysis); carrier frequencies should be integer multiples of the modulation frequency (to aid in stimulus generation); pulse periods should be such that responses to pulses on the left are minimally correlated with responses to pulses on the right when averaged over the whole stimulus (as in Hill *et al* [8]); the pulses should sound, subjectively, like they are 'part of' the 30%-amplitude background trilling of their respective streams; finally and most importantly, it should be as easy as possible to focus one's attention on one stream and ignore the other, there being as little as possible perceptual 'binding' between the two streams. The streams' opposite laterality, their differing carrier, modulation and pulse frequencies, and the temporal offset between the first pulses on each side, all contributed to this.

Before the experiment began, subjects were asked to listen to the stimulus a few times while adjusting the volume of the left and right headphone outputs using two analogue sliders. The criteria were that the volume should be comfortable, and that attending to the left stream and ignoring the right should be, subjectively, equally easy as vice versa.

Since our design included a target-counting task, one final aspect of the stimulus design was that a minority of the pulses were longer in duration. The first two pulses on each side were always *standard* 100 ms pulses. After this, the remaining pulse sequence might contain one, two or three *target* pulses whose duration was 180 ms. The correct number of targets was chosen uniformly, randomly and independently for each stream on each trial. The example stimulus of figure 1 contains three targets on the left and one on the right.

At the end of the stimulus, the fixation cross was replaced by a question mark in the centre of the screen. This signalled to the subject that they were free to blink, swallow and move. At this moment, they received acoustic feedback (a single 'ding!' of a bell) if the system had correctly classified attention-to-the-left versus attention-to-the-right using their EEG. The question mark also signalled that the subject had up to 5 s in which to press a key on their numeric keypad, to report how many target stimuli had been in the attended stream. As soon as they pressed the key (or after 5 s had elapsed), the screen displayed, for 2 s, the correct number of targets in each stream: the numeral on the attended side was green if the subject had responded correctly, red if not. After a 1–2 s pause, the next trial began.

In the final free-choice block, the classification result could not, of course, be judged as 'correct' or 'incorrect' until after the experiment, so there was no bell sound. To ease the increased complexity of the task (making a free choice and writing it down), we also removed the obligation to press a key and the feedback about the number of targets. Instead, the screen simply reported 'interpreted as LEFT' or 'interpreted as RIGHT' according to the classifier's output.

2.3. Hardware and software

A BrainProducts 136-channel QuickAmp was used in combination with the BCI2000 software platform [38] to acquire signals at 500 Hz from 67 EEG positions, roughly evenly distributed throughout the 10/20 system and mounted on an ElectroCap EEG cap, as well as three EOG electrode positions around the left eye: above (EU1), below (EL1) and lateral to the outer canthus (EO1). Impedances were lowered below 5 kΩ and unused channels were grounded. The cap used a ground electrode at AFz, and the amplifier applied a built-in common average reference across the 70 biosignal inputs. For all online and offline analyses, the electrodes were re-referenced in software to remove this, using the sparse spatial filter option of BCI2000. The result was 66 EEG signals referenced to linked mastoids ($[\mathrm{TP9} + \mathrm{TP10}]/2$), as well as horizontal and vertical EOG ($\mathrm{EO1} - [\mathrm{EU1} + \mathrm{EL1}]/2$ and $\mathrm{EU1} - \mathrm{EL1}$, respectively).

Auditory stimuli were delivered using four independent channels of a multi-channel soundcard: left and right channels to the headphones, plus two channels which were fed via an optical isolator into two auxiliary inputs of the EEG amplifier. These auxiliary channels served to provide a synchronization signal for the timing of the left and right stimuli, in the EEG data stream. Visual stimuli were presented on an LCD monitor at a comfortable distance.

Data were recorded using BCI2000 [38]. Online signal processing, stimulus generation and stimulus presentation were implemented in Python using the 'BCPy2000' add-on to BCI2000 [39]. Offline analysis, and training of classifiers between blocks, was performed using Matlab.

2.4. Analysis

2.4.1. Online (ERPs). Online classification was based entirely on ERPs, using methods and parameters optimized via the experience gained from analysing the results of Hill *et al* [8] offline. First, the re-referenced EEG signals were band-pass filtered using an order-6 Butterworth filter designed to pass frequencies between 0.1 and 8 Hz. Second, following the onset of every pulse in either the left or the right stream, a 600 ms segment or *epoch* of the filtered EEG signal was cut out and stored in memory. Within each trial, the first two left-pulse responses and the first two right-pulse responses were discarded (on the assumption that the subject might require some time to 'lock on' their attention to the correct stream), after which the system maintained a running average X_L of the epochs following left-stream pulses, and an average X_R of the epochs following right-stream pulses. The difference $X = X_R - X_L$, a 68-channel-by-300-sample matrix, was used as a feature set. Every time a new pulse occurred and X changed, a new classifier output was computed by multiplying the elements of X by a set of linear weights. At the end of each trial i, the final $X^{(i)}$ for that trial was written to a disc to provide one training exemplar for future classifiers. In the first attention block, there was no feedback from the classifier, but after each attention block, a classifier was trained on all the attention data gathered so far, and the resulting weights loaded into the system. Thus, from the second attention block onwards, subjects received feedback about how well their signals could be classified.

Spatial *whitening* of the data has previously been shown to produce a benefit in linear ERP classification[3]. Therefore, our first step in classification is to estimate a 68-by-68-channel spatial covariance matrix Σ_s from the training data, and whiten the feature representation for each trial i using $X_P{}^{(i)} = \Sigma_s^{-\frac{1}{2}} X^{(i)}$. The $X_P{}^{(i)}$ were then classified using an L2-regularized linear classifier (for online purposes we used the logistic-regression method), with the regularization parameter being found by tenfold cross-validation within the training set. The weights found by the classifier in the whitened space (call them M_P) were then transformed back to yield weights $M = \Sigma_s^{-\frac{1}{2}} M_P$ that can be applied directly to the unwhitened data.

[3] Comparing the fourth and sixth columns of table 1 of Hill *et al* [8], we can compare offline classification accuracies immediately before and immediately after applying the FastICA algorithm to the data. The percentage-point improvement from applying FastICA was 4.0 on average, with standard error 1.0 across 15 subjects. Since FastICA consists of a PCA whitening step followed by a rotation in the electrode space, and since L2-regularized classifiers are invariant to such rotations, this performance improvement reflects what can be achieved by spatial whitening alone—a fact which our experiments in offline data analysis (results not shown here) have confirmed for both the 2005 data and the current data.

2.4.2. Offline (ERPs and SSAEP). Further offline analysis of the ERPs used very similar methods to those described above, the only differences being that the preprocessing chain was re-created offline in Matlab, and the performance was assessed by tenfold cross-validation (and since the training procedure itself employed cross-validation for model selection, this resulted in double-nested cross-validation). To examine the potential effect of manipulating the length of a trial, the analysis was repeated using only the first beat of each stream in each trial, only the first two beats, only the first three beats, and so on until all beats were used (unlike the online classification procedure, no beats were discarded).

For offline analysis of the SSAEPs, we cut two segments of the re-referenced multi-channel signal of each trial: a baseline segment measured during visual cue presentation and a stimulus segment starting 1 s after the stimulus onset. The segment length of 1872 ms (i.e. 936 samples) was chosen since it was close to the classification interval length of 2 s used by Farquhar *et al* [23], while being an exact integer multiple of both the amplitude-modulation periods. The SSAEPs induced by sinusoidal amplitude modulation are, themselves, almost pure sinusoids, and stand out very clearly in single components of the Fourier transform when this condition is fulfilled. Since, furthermore, the signal is phase-locked to the stimulation, it is appropriate to use linear preprocessing and classification methods. We follow Farquhar *et al* [23] in using a correlation method: for each EEG channel and for each of the two AM frequencies in question, we take the real and imaginary coefficients of the discrete Fourier transform as features (i.e. the correlation coefficients of the segment with a cosine-wave and with a sine-wave, a linear basis for fitting a sinusoid at any phase). This ensures that, as far as possible, only the information from the SSAEPs is being used, to allow their assessment in isolation from that of the ERPs. Once again we classify the features with a linear L2-regularized classifier (logistic regression).

3. Results

3.1. Online performance

Figure 2 shows the BCI classification performance attained and experienced by the subjects during the experimental session. Since the classifier was retrained after each block of 20 trials, the horizontal axis serves to indicate both the number of training trials and, roughly, the time elapsed during the experimental session. We might expect performance to improve over time as a function of subject learning and of the number of training trials with which the classifier has to work, although we might also anticipate some deterioration of performance over time if the subjects become fatigued. The figure allows us to assess the combined effect of these factors.

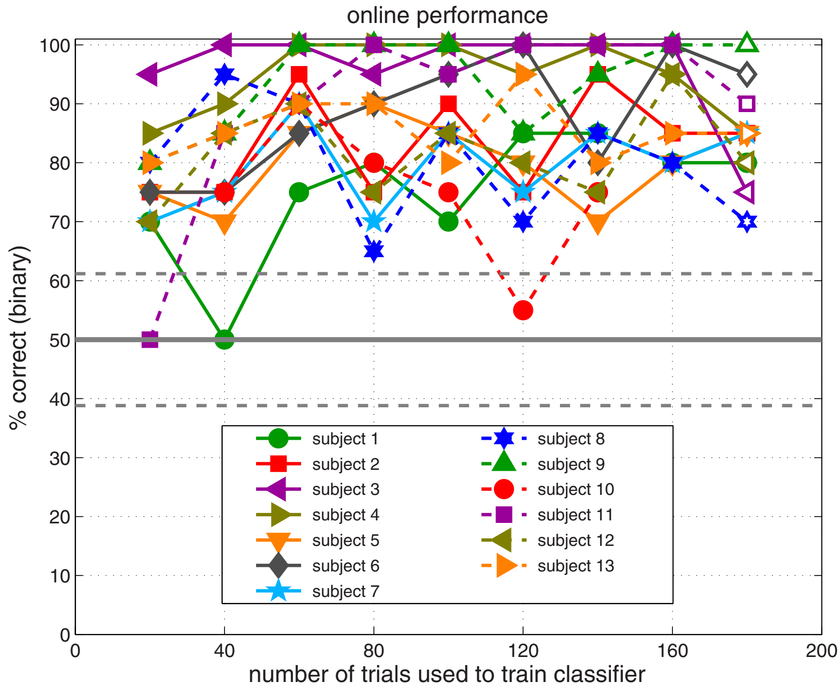

Figure 2. BCI performance expressed as the percentage of trials classified correctly online. Each point denotes one block of 20 trials. After each block, the classifier was re-trained on all the data gathered so far, so the 'number of trials' axis is also effectively a time axis. The different symbol shapes/colours correspond to different subjects. Filled symbols denote blocks of cued trials with trial-by-trial feedback, whereas open symbols denote a final 'free-choice' block in which subjects wrote down their decisions on paper as they went along. The horizontal tramlines indicate chance classification performance (50%) ± 1 standard error for assessing the significance of a single 20-trial block. For assessing the significance of one subject's results over all nine data points (180 trials), the tramlines would indicate ± 3 standard errors.

For many subjects (e.g. subjects 3, 4, 9 and 13), performance appears to reach its peak very quickly, after only 40–60 training trials (i.e. about 10–15 min of calibration). For others, such as subject 1 or even the very high-performing subject 6, it may take longer (120 trials or 30 min). Only two subjects (8 and 10) appear to show a net decline after reaching an early peak—possibly due to fatigue.

Online performance was then averaged across all nine blocks for each subject. This performance statistic has a mean of 84.8% and a standard deviation of 7.2% across all 13 subjects. In figure 3, these values are plotted as a function of each subject's counting accuracy. The overall mean counting accuracy was 82.9% with a standard deviation of 14.6% across subjects. The relationship between the BCI and behavioural measures of attention, though nonlinear, is clearly quite monotonic, with a Spearman rank correlation coefficient of 0.79 ($n = 13$, $p = 6 \times 10^{-4}$). To a large extent, inter-subject variation in BCI performance can therefore be explained by the variation in counting performance (see the discussion in section 5).

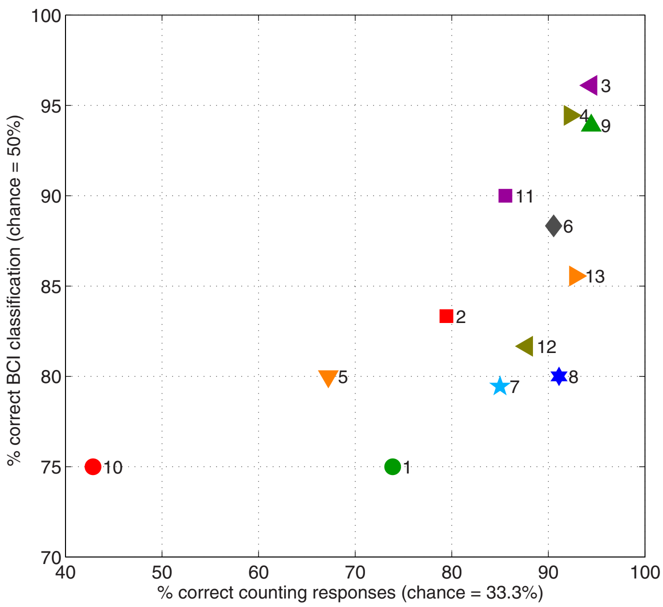

Figure 3. Online BCI classification accuracy for each subject (averaged across all the blocks of figure 2) is plotted against the subjects' accuracy in reporting the number of target stimuli. Numerals next to the symbols denote (chronological) subject ID numbers. Each subject also has a characteristic symbol shape/colour matching those in figures 2, 4 and 7. Our behavioural measure of attentional performance explains a great deal of the between-subject variation in BCI performance (Spearman's rank correlation: $r = 0.79$, $n = 13$, $p = 6 \times 10^{-4}$).

3.2. Offline re-analysis (ERPs)

Figure 4 shows hypothetical performance, as estimated by offline cross-validation, as a function of the length of the trial. Disregarding the ceiling effect for the best subjects, % accuracy increases steadily as more stimuli are averaged (upper panel). Accuracy may be traded off for time taken to measure a single trial: this tradeoff is seen in the lower

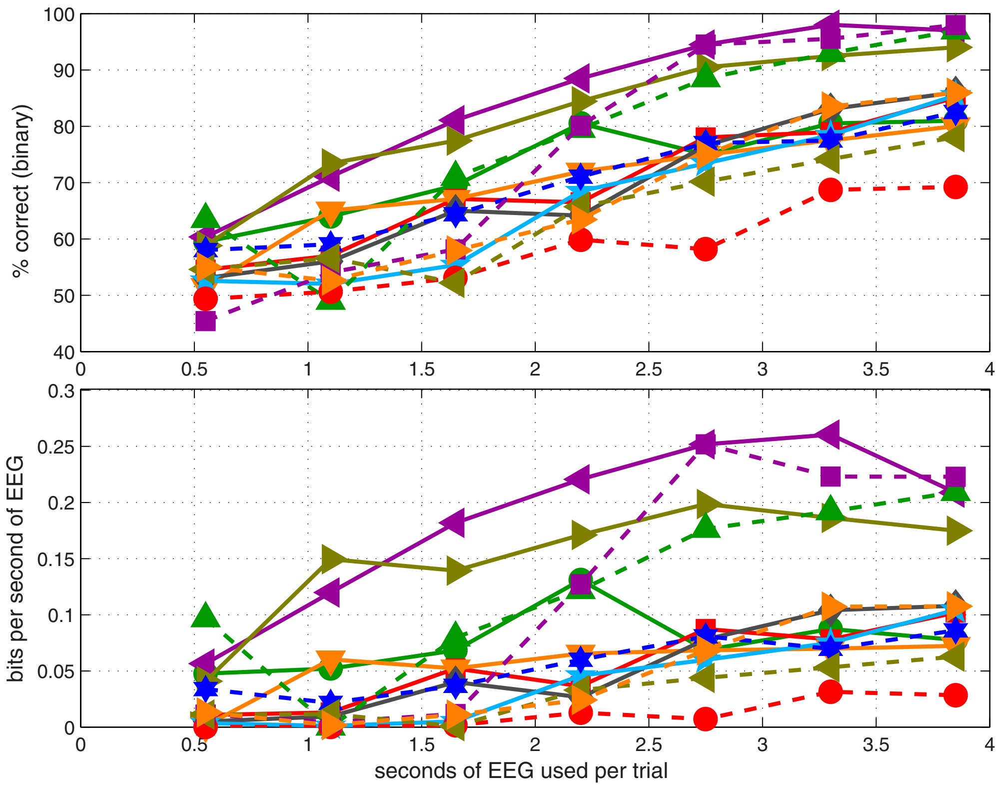

Figure 4. Hypothetical BCI performance estimated by offline cross-validation. In the upper panel, results are expressed as the percentage of trials classified correctly in tenfold cross-validation, each point being based on the full dataset for one subject (200 trials for most; 160 for subject AS). In the lower panel, the same results are re-represented as ITRs in bits per second. The different symbol shapes/colours correspond to different subjects. The analysis was performed repeatedly using only the first n beats from each stream in each trial, with n varying from 1 to 7 along the abscissa.

panel, where the same results are expressed as an information transfer rate (ITR), computed in bits per symbol according to the definition presented in Wolpaw *et al* [40], and then divided by the number of seconds of EEG used per symbol. (Note that we avoid expressing this in bits/min to draw attention to the fact that it is not directly comparable to the ITR frequently reported in bits/min in BCI studies. The latter statistic usually takes into account the 'overhead' of the inter-trial gaps, which are not meaningful in the context of this analysis.)

In this view, there seems to be relatively little value added by playing more than five pulses on each side, or thereby acquiring more than about 3 s of EEG data per trial. (Naturally, this is an upper bound on the amount of useful information that can be transmitted, and may be unrealistic: depending on the level of error correction that can be built in while retaining usability, any real instantiation of a full BCI+HCI system might achieve less than this maximum, and may benefit from trading off more time for accuracy than this apparent optimum would suggest.)

To gain some insight into where and when the useful discriminative information arises in the EEG features, signed coefficients-of-determination (SCD) were computed to measure the extent to which each individual feature separates the two different types of trial (attend-left or attend-right). SCD is also known as 'signed r^2', where r is the correlation coefficient, computed across trials, between the feature value in question and the label value (label -1 denoting attend-left or $+1$ denoting attend-right). r^2 can be interpreted as the proportion of variance in label values that the feature in question, considered alone, can account for, and multiplying it by the sign of the original r preserves the direction of the correlation. Figure 5 shows the results. Note that attend-right trials are assigned the larger label value (+1) and the features were computed as $X_R - X_L$ (response to right-stream stimuli *minus* response to left-stream stimuli). This means that, making the simplifying assumption of equal variances of X_R and X_L, we can show that we would obtain identical SCD values for a single subject, up to a scaling factor of $1/\sqrt{2}$, if we were instead to take each X_R and X_L as a separate data exemplar and redefine the labels such that we were contrasting responses to attended (+1) versus unattended (−1) stimuli. If we do this (results not shown) we do in fact obtain results which are qualitatively almost identical to those of figure 5: the figure may therefore also be understood, perhaps more intuitively, as indicating the contrast between attended and unattended stimuli. This also means that comparisons between the upper-left panel (where SCDs are shown) and the lower-left panel (averaged EEG for attended and unattended stimuli) are meaningful.

The figure shows four temporal and two spatial views of the discriminative information, from SCD values that have been averaged across all subjects. One must be cautious in interpreting the patterns, because of overlap effects: every stimulus-locked average is polluted by responses to previous

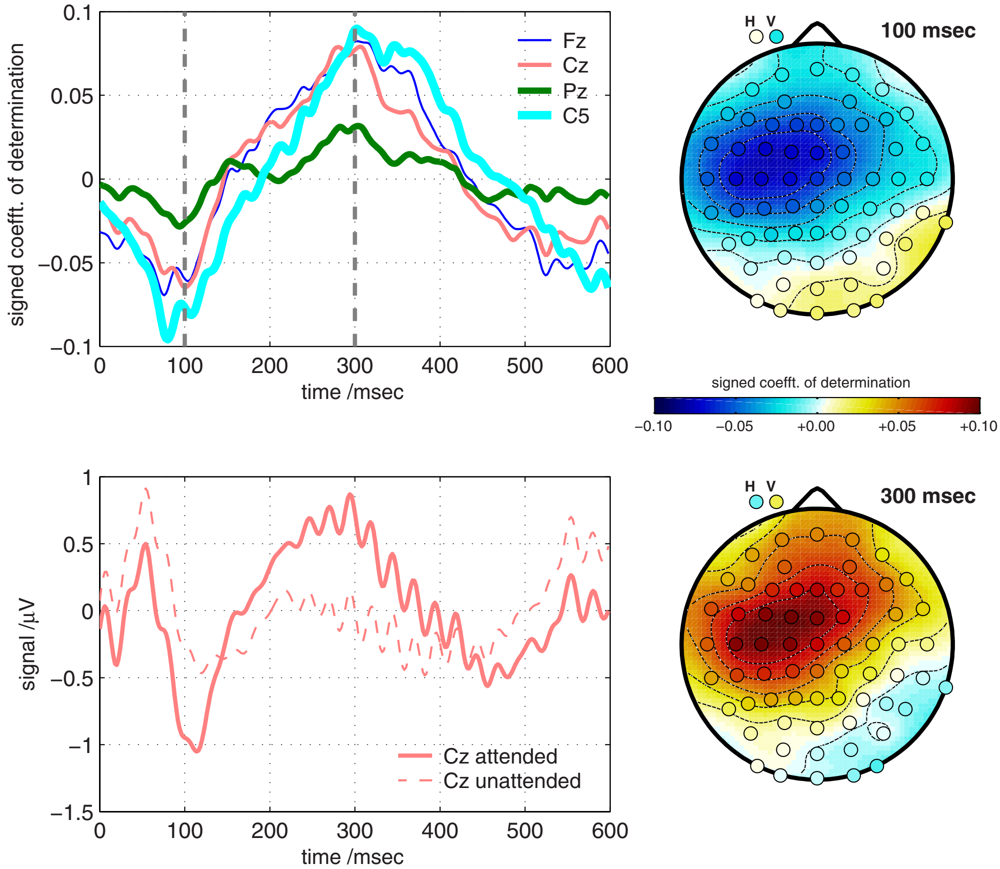

Figure 5. Signed coefficients of determination (SCDs, also referred to as signed r^2 values), averaged across all 13 subjects. These contrast values illustrate how each feature is correlated with the distinction between attend-right (positive class) and attend-left (negative class) trials. The SCD values were computed from a space consisting of 68 channels × 300 time samples: each feature value denotes the response averaged for all right-stream stimuli in a given trial (targets and non-targets) minus the response averaged for all left-stream stimuli (targets and non-targets) in the same trial. The upper-left panel shows the time-course of the SCD values at four different EEG channels from time $t = 0$ (the moment of stimulus presentation) to $t = +600$ ms. The scalp maps show the spatial distributions of SCD values across the EEG montage at the two instants $t = +100$ ms and $t = +300$ ms. The horizontal and vertical EOG channels are marked H and V, respectively. The lower-left panel shows the EEG signal at Cz bandpass-filtered between 0.1 and 45 Hz, time-locked to attended and unattended stimuli (averaged across all stimuli, both streams and all subjects).

and subsequent stimuli from the opposite stream (however, because of the phase drift between attended and unattended sides, the pollution is not time-locked to $t = 0$ in the average and is therefore somewhat attenuated relative to responses to the stimuli whose onset is at $t = 0$). There is a negative peak which appears at Cz/Fz at around 100 ms after stimulus presentation (this actually tended to be lateralized to the left, and tended to originate a little earlier and more lateralized, as reflected by the C5 trace). There is also a positive component originating frontocentrally around 200 ms and evolving into an (also slightly left-lateralized) 250–400 ms peak.

3.3. Offline analysis (SSAEPs)

Figure 6 shows the SSAEP signals from one example subject (subject 3, the best performer in ERP-based BCI). The spectra are computed at Cz, referenced to linked mastoids. With a window length carefully chosen to contain an integer number of cycles of both amplitude-modulation signals, the two SSAEPs components stand out very sharply in the power spectrum as estimated by an FFT on a rectangular-windowed segment. For the purposes of this illustration, the log power was computed at the two precise SSAEP frequencies across the whole mastoid-referenced scalp montage. Then, SCD scores (see section 3.2) were computed for these features to show the degree to which each feature can be used to separate baseline from stimulus segments (second and fourth row of scalp plots) or attend-left versus attend-right segments (first and third row of scalp plots). The scalp distribution of the SSAEPs is somewhat lateralized, centred roughly on C3 and C4, contralateral to the ear to which the corresponding stimulus component was presented. There was no such pattern of features, and very poor overall separation, when attempting to separate attend-left versus attend-right conditions (top row).

Figure 7 presents a summary of the results for all the subjects, showing that the pattern illustrated with subject 3 is consistently repeated. The FFT coefficients at the precise amplitude-modulation frequencies could be used to classify baseline versus stimulus segments at about 89.5% correct (±4.9% across subjects) in offline cross-validation (group C in the figure). They could also be used to classify listen-right trials from listen-left trials in the perception-only condition,

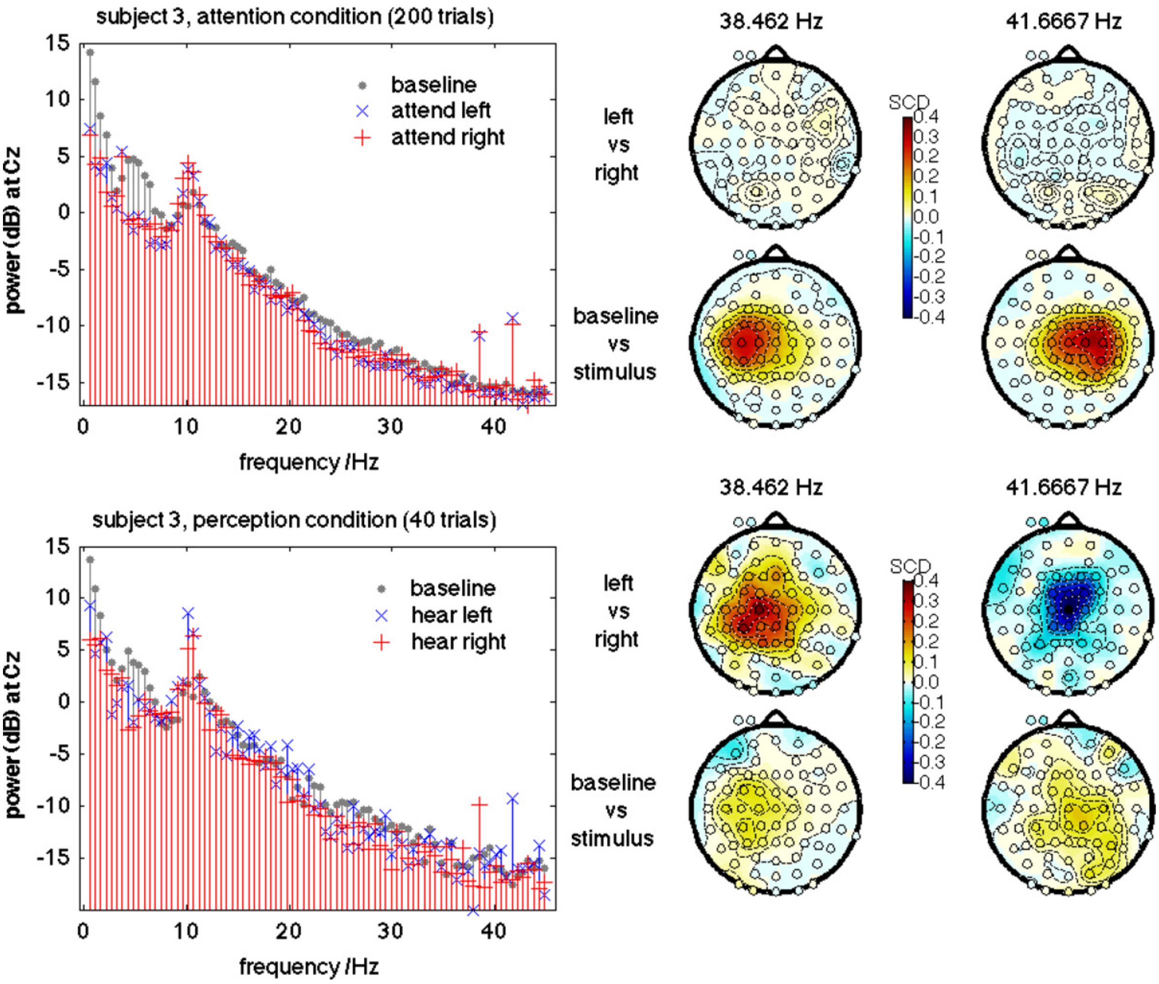

Figure 6. SSAEPs for one example subject—the best-performing subject in the ERP-based BCI. The upper half of the figure shows the attention condition (200 trials), and the lower half shows the perception condition (40 trials). In the leftmost plots, averaged power spectra at Cz are shown for the silent baseline (grey dots), attend-left (blue $\times$) and attend-right (red +) time intervals. The scalp maps all show the signed coefficient of determination (SCD, also known as signed r^2) for individual spectral power features in separating left trials from right trials, or baseline periods from stimulus periods. All scalp maps share the colour scale of 0.0–1.0, the extreme values of which would indicate perfect separation, whereas 0.5 indicates that the feature does not separate the classes at all. For the purposes of illustration, power values are computed without spatial filtering, from signals referenced to the mastoids.

where the unattended stream was silent, at 88.8% $\pm$ 8.6 (group B; the increased variability can be attributed to the smaller number of trials collected in this condition). However, the same feature extraction and classification procedures could not be used to solve the BCI problem of discriminating the attended side when both streams were audible (50.2% $\pm$ 4.2, group D), despite the proven ability of the subjects to modulate their ERPs by shifting their attention to the same stimuli (84.8% $\pm$ 7.2, group E).

3.4. Subjective user reports

The following subjective phenomena were each reported, unprompted, by three or more of the subjects.

(i) Subjects reported that the counting task seemed helpful as a tool for focusing attention at the beginning of the session, but that gradually (as the trial-by-trial feedback became more accurate and the subject grew more accustomed to it) it became redundant. The latter observation is partially supported by the results of the final free-choice block in which counting was made optional: some subjects reported that they had not bothered to count, yet overall BCI accuracy was not significantly lower than in previous blocks.

(ii) Subjects made statements like ‘I knew I had done well/not so well on many of the trials, even before the bell’ suggesting that after a few blocks of trial-by-trial BCI feedback, there was strong sense of being able to evaluate one’s own attention-shifting performance to a degree that seemed correlated with the BCI’s evaluation.

More-rigorous support for these unlooked-for subjective observations may be a valuable aspect of future studies.

4. Discussion

The results demonstrate that attentional shifts to dichotically presented auditory streams are a feasible basis for an effective online binary-choice BCI, in which high single-trial accuracy can be achieved within a few seconds per trial.

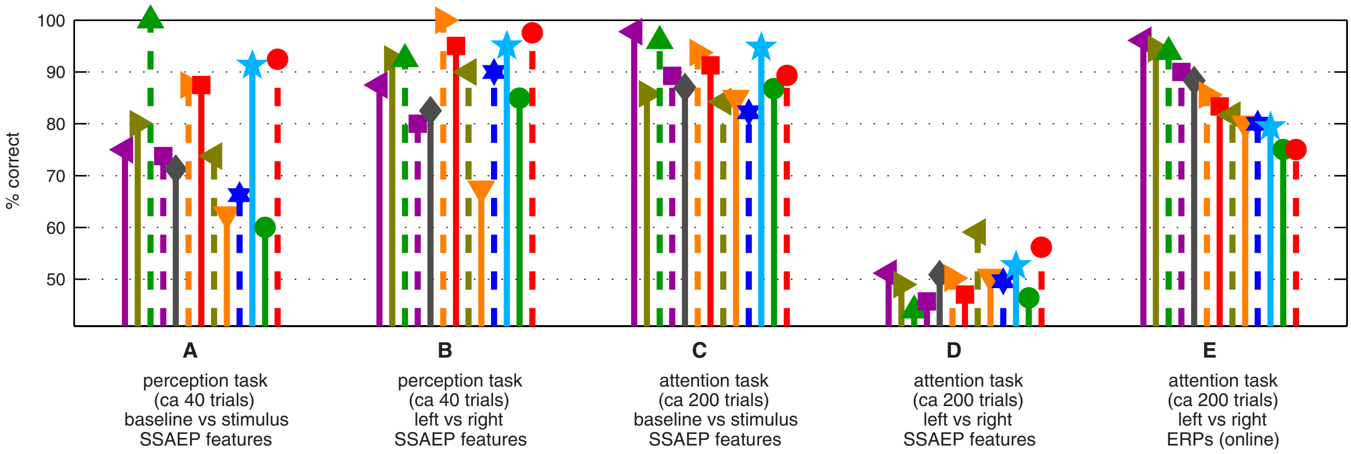

Figure 7. Classification results. Groups A through D show classification using SSAEP features only (FFT coefficients at the precisely known SSAEP frequencies). A and B show results from the 'perception' condition where the unattended stimulus stream was silent, whereas C and D reflect the normal attention condition. A and C show performance in distinguishing stimulus-presentation periods from baseline (silence) periods, whereas A and D show performance in identifying which of the two stimulus streams, left or right, was attended. Each subject is shown individually with his or her unique symbol shape and colour, and within each group the subjects are ordered from left to right in decreasing order of the online (ERP-based) BCI performance they achieved using the very same stimuli. For comparison, these ERP-based BCI performance values are shown in group E.

The ERP-based BCI appears to rely on early (N1) as well as later (P3) components. The usefulness of the N1 component is consistent with the observation of attention-modulation of the N1 component by Hillyard *et al* [16] and is an encouraging sign that BCIs might be constructed using every stimulus in a periodic sequence, rather than necessarily having to rely on (and wait for) a smaller number of less-frequent 'oddball' stimuli, on which purely P3-driven stimulus designs are traditionally assumed to rely.

4.1. Performance comparison with other auditory-ERP studies

Our subjects' average online accuracy of 84.8% is very slightly higher than the 82.4% that was predicted by the best analysis method in our previous offline study [8].

Unfortunately we cannot compare against the results of Kanoh *et al* [12], since the latter authors only reported offline accuracies in which *all the measured responses were used as both sample and test data* (ibid p 38). We interpret this to mean that performance was measured on the classifier's own training set and was therefore a drastically inflated performance measure.

Our results can be compared with those of Halder *et al* [13] if we convert to ITRs. Computing ITR for each subject separately before averaging, we obtain a mean of 0.415 bits/trial with a standard deviation of 0.195 across subjects. Following Halder *et al* 's convention of taking into account only the time used to play the stimulus and discarding inter-trial gaps, we obtain 4.98 bits $\min^{-1} \pm 2.3$. This is twice the mean ITR of 2.46 bits $\min^{-1}$ reported in the best condition in Halder *et al*'s table 3, which in turn exceeds all the other auditory studies reviewed in that table. Our interpretation of this performance difference highlights an important point about streaming designs. In a streaming design, subjects may be asked to discriminate between 'targets' and 'non-targets', or they may not. *If* they do so, as in the current study, then this has the advantage of providing a concrete strategy and incentive for shifting attention, and an opportunity for the experimenter to verify the level of attention using a behavioural response. However, the target–non-target contrast is not a necessary part of the BCI design, and it is not solely the brain response to the target stimulus that is important in classification. Rather, as Hilyard's original paper also reported, attention modulates the response to *every* stimulus that the user is monitoring (assessing it, if required to do so, as a target or non-target) relative to the responses to the stimuli in the stream that the user is ignoring. This has the advantage that the system does not need to wait for an infrequent 'oddball' target before updating its interpretation of the user's attentional state: a meaningful update to the BCI system's output could be made, and assessed to see whether enough information has been gathered to make a decision, every time a stimulus is presented (roughly twice per second in the current stimulus design). Seen in this light, the Halder *et al* study had the worst of both worlds: it was a sequential rather than a streaming design (so the system has to wait for the target) but had only two classes of target, plus a majority of unused non-targets.

It is possible, however, to take advantage of the non-targets by diversifying them into multiple classes. To compare our study with multi-class sequential designs, however, one must be aware that the performance metric is usually different from that used above, because such systems are usually assessed in the context of a more complex real-world task. Before pruning their subject group to select only the better subjects, Schreuder *et al* [19] report a mean ITR of 2.84 bit $\min^{-1}$ across all their 21 subjects. Note that the metric now *includes* the time taken outside of stimulus presentation, i.e. the 'overhead' of following cues and choosing a letter in a real, practical task—something for which neither our minimal two-class experiment nor that of Halder *et al* has any analogue. Our subjects performed an average of four trials per minute: one trial every 15 s on average of which the stimulus was played for 5 s. To match the mean ITR of Schreuder *et al*, our subjects would have had to perform 6.84 trials per minute (5 s stimuli out of every 8.8 s). While this might easily have been possible with our task (healthy subjects mindlessly repeating left–right choices that are prescribed for them), it would be a considerable challenge to design a communication interface based on free binary choices, including all the necessary auditory prompts and feedback, that could achieve this. For communication, binary BCIs will probably only be the preferred choice for users who are unable to use a more complicated speller.

4.2. Implications for SSAEP BCIs

Our results also provide evidence that SSAEPs are a much poorer neural basis for the attention-based BCI than auditory ERPs. Furthermore, it seems unlikely that the negative finding in this study (and hence probably also the previous negative finding of Farquhar *et al* [23]) can be attributed to an inadequate attention-task design, since exactly the same stimuli that elicited the SSAEPs also, simultaneously, elicited ERPs that *were* very successfully modulated by attention. Although SSAEP features could clearly be seen, and could easily be used to detect which of the two stimuli was *heard*, there was absolutely no evidence that they could be used to detect attention: SSAEP classifier performance did not tend to predict even the rank order of the subjects' attention-modulation ability, as measured either behaviourally by the counting task or electrophysiologically by their ERP-based BCI performance.

From the current results, one might even conclude that absolutely *no* attention-modulation of SSAEP is detectable on a practicable single-trial basis. Recent results, however, taken together with the literature reviewed in the introduction, suggest that this cannot be entirely true: Kim *et al* [37] played 20 s long trials consisting of two-stream amplitude-modulated stereo sound, used FFT features from EEG measured from only four electrodes, and showed that one pilot subject may have been able to use attention modulation for online BCI selection, obtaining 10 correct trials out of 14. Note that this is only tentative positive evidence: this online performance of 71.4% is only 1.6 standard errors above chance (under the null hypothesis of probability correct = 0.5, the standard error would be $\sqrt{(0.5 \times 0.5)/14} = 0.1336$). Offline results from the same study are also moderately encouraging: averaged over six subjects each performing 50 trials, percentage accuracy in cross-validated offline tests ranged from the low 70's to the mid 80's depending on the feature representation in use, provided the stimulus length was 10 s or more. (Note, however, that since their offline analysis methods did not appear to use double-nested cross-validation, it is not appropriate to take the averaged-across-subjects *maximum*-across-feature-set performance of 86.3% as a fair indication of expected generalization performance.) Despite their relative statistical weakness, these results are suggestive of significant attention modulation of SSAEP, of the kind that might be harnessed for BCI, even if the strength of the effect is inferior to the attention modulation of ERPs. The minor differences between Kim *et al* [37] and this study are too many to be able to pin down the reason for this difference in findings exactly: the different numbers of electrodes, the use of speakers in free field instead of dichotic listening through headphones, the constant rather than pulsed envelope of the stimuli and (the most likely influential factor) the much longer stimulus durations may all play a role which bears closer investigation.

5. Outlook

As in most BCI systems, accuracy varies widely from subject to subject—however, we have evidence that this variation can in large part be explained by the subject's ability (or perhaps motivation) to focus attention on the stimuli, as indicated by our behavioural measure of counting performance. This is somewhat unusual in the BCI literature: although most BCI studies report large inter-subject variation, it is rare for the design to include an independent behavioural test of attention. The finding raises the hope that BCI performance for many poor-performing subjects could be improved considerably by training, since clearly there is still a behavioural learning curve that the poorer BCI performers can attempt to climb. Furthermore, the design of the system allows for a quantitative measure of auditory attention to be computed online and updated almost in real time—roughly three to four times per second in the current design, although this output signal would need to be smoothed. Such a signal could be fed back to the subject, perhaps as a visual or tactile stimulus: this would potentially breathe life into neurofeedback methodologies for improving attention, now based on *direct*, immediate correlates of performance in an attention-driven task, i.e. the attention modulation of brain responses to specific stimuli, in contrast to more traditional approaches that train the amplitude of oscillatory components known to be correlated indirectly with attention-task performance (see, for example, [41, 42]). The effectiveness of this kind of neurofeedback (as compared to ordinary behavioural training on the same counting task) remains to be tested. Favourable results in such a comparison might establish auditory-streaming BCI as a valuable tool *outside* the sphere of BCI-for-communication. One example of its application might be as a treatment for people who have specific difficulty with dichotic listening and auditory attention tasks, as in some cases of central auditory processing disorder (CAPD) [43]. Another might be as a training tool—for example, for people in occupations such as simultaneous interpreting [44] that require dichotic listening skills or other forms of selective attention in 'cocktail-party'-style acoustic environments.

Future extensions of the current paradigm might investigate

- whether stimuli can be speeded up without sacrificing performance, to yield a larger ITR;
- whether the stimuli are suitable for an older population, more matched to the demographics of the target population of people in the totally locked-in state;
- whether other more-pleasant or more-intuitive stimuli (e.g. voices repeating task-relevant pairs of words such as 'yes' and 'no') might be used without sacrificing performance;

- to what extent usability and performance are served by an adaptive stopping mechanism, whereby stimuli continue indefinitely until the classifier has enough confidence in its output on each trial (this is related to the question of how well the *no-intentional-control* brain state can be distinguished from the brain states corresponding to left-selection and right-selection).

Further development of the system will also require integration of the BCI-based left–right selection into a comprehensive human–computer interface based on binary decision trees, through which a user might spell, or control other systems such as domotic control interfaces that are useful for the target population. Integration of the BCI described here into such a wider system would provide a valuable expansion of communication and control possibilities for people who are paralysed and (hence, or otherwise) have limited vision.

Acknowledgments

This research was supported by the Max-Planck-Gesellschaft, and by the NIH (NIBIB; EB00856). The authors would like to thank Bernd Battes for his excellent work as a lab technician and all-round EEG troubleshooter, and for coordinating with subjects and gathering the data. We would also like to thank Peter Desain and Jason Farquhar for helpful discussions during the development of the paradigm.

The authors report no conflicts of interest. None of the material presented in this manuscript has been published or is under consideration elsewhere, including the Internet.

References

[1] Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E and Flor H 1999 A spelling device for the paralysed *Nature* **398** 297–8

[2] Birbaumer N, Murguialday A R and Cohen L 2008 Brain–computer interface in paralysis *Curr. Opin. Neurol.* **21** 634–8

[3] Naito M, Michioka Y, Ozawa K, Ito Y, Kiguchi M and Kanazawa T 2007 A communication means for totally locked-in ALS patients based on changes in cerebral blood volume measured with near-infrared light *IEICE Trans. Inf. Syst.* **E90-D** 1028–37

[4] 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

[5] Donchin E, Spencer K M and Wijesinghe R 2000 The mental prosthesis: assessing the speed of a P300-based brain–computer interface *IEEE Trans. Rehabil. Eng.* **8** 174–9

[6] Müller-Putz G R, Scherer R, Neuper C and Pfurtscheller G 2006 Steady-state somatosensory evoked potentials: suitable brain signals for brain–computer interfaces? *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 30–7

[7] Brouwer A-M and van Erp F J B 2010 A tactile P300 brain–computer interface *Front. Neuroprosthetics* **4** 19

[8] Hill N J, Lal T N, Bierig K, Birbaumer N and Schölkopf B 2005 *An Auditory Paradigm for Brain–Computer Interfaces (Advances in Neural Information Processing Systems* vol 17*)* ed L K Saul, Y Weiss and L Bottou pp 569–76

[9] Sellers E W and Donchin E 2006 A P300-based brain–computer interface: initial tests by ALS patients *Clin. Neurophysiol.* **117** 538–48

[10] Furdea A, Halder S, Krusienski D, Bross D, Nijboer F, Birbaumer N and Kübler A 2009 An auditory oddball (P300) spelling system for brain–computer interfaces *Psychophysiology* **46** 617–25

[11] Klobassa D S, Vaughan T M, Brunner P, Schwartz N E, Wolpaw J R, Neuper C and Sellers E W 2009 Toward a high-throughput auditory P300-based brain–computer interface *Clin. Neurophysiol.* **120** 1252–61

[12] Kanoh S, Miyamoto K I and Yoshinobu T 2010 A brain–computer interface (BCI) system based on auditory stream segregation *J. Biomech. Sci. Eng.* **5** 32–40

[13] Halder S, Rea M, Andreoni R, Nijboer F, Hammer E, Kleih S, Birbaumer N and Kübler A 2010 An auditory oddball brain–computer interface for binary choices *Clin. Neurophysiol.* **121** 516–23

[14] Schreuder M, Blankertz B and Tangermann M 2010 A new auditory multi-class brain–computer interface paradigm: spatial hearing as an informative cue *PloS ONE* **5** e9813

[15] Höhne J, Schreuder M, Blankertz B and Tangermann M 2011 A novel 9-class auditory ERP paradigm driving a predictive text entry system *Front. Neuroprosthetics* **5** 99

[16] Hillyard S A, Hink R F, Schwent V L and Picton T W 1973 Electrical signs of selective attention in the human brain *Science* **182** 177–80

[17] Kübler A, Furdea A, Halder S, Hammer E M, Nijboer F and Kotchoubey B 2009 A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients *Ann. NY Acad. Sci.* **1157** 90–100

[18] Schreuder E M, Tangermann M and Blankertz B 2009 Initial results of a high-speed spatial auditory BCI *Int. J. Bioelectromagn.* **11** 105–9

[19] Schreuder M, Rost T and Tangermann M 2011 Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI *Front. Neuroprosthetics* **5** 112

[20] Belitski A, Farquhar J and Desain P 2011 P300 audio-visual speller *J. Neural Eng.* **8** 025022

[21] Höhne J, Schreuder M, Blankertz B and Tangermann M 2010 Two-dimensional auditory p300 speller with predictive text system *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* pp 4185–8

[22] Kallenberg M 2006 Auditory selective attention as a method for a brain–computer interface *Masters Thesis* Radboud University Nijmegen

[23] Farquhar J, Blankespoor J, Vlek R and Desain P 2008 Towards a noise-tagging auditory BCI-paradigm *Proc. 4th Int. Brain–Computer Interface Workshop and Training Course* pp 50–5

[24] Galambos R, Makeig S and Talmachoff P J 1981 A 40 Hz auditory potential recorded from the human scalp *Proc. Natl Acad. Sci.* **78** 2643–7

[25] Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J and Chauvel P 2004 Temporal envelope processing in the human left and right auditory cortices *Cereb. Cortex* **14** 731–40

[26] Ross B, Borgmann C, Draganova R, Roberts L E and Pantev C 2000 A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones *J. Acoust. Soc. Am.* **108** 679–91

[27] Cheng M, Gao X, Gao S and Xu D 2002 Design and implementation of a brain–computer interface with high transfer rates *IEEE Trans. Biomed. Eng.* **49** 1181–6

[28] Kelly S P, Lalor E C, Finucane C, McDarby G and Reilly R B 2005 Visual spatial attention control in an independent brain–computer interface *IEEE Trans. Biomed. Eng.* **52** 1588–96

[29] Allison B Z, McFarland D J, Schalk G, Zheng S D, Jackson M M and Wolpaw J R 2008 Towards an independent brain–computer interface using steady state visual evoked potentials *Clin. Neurophys.* **119** 399–408

[30] Zhang D, Maye A, Gao X, Hong B, Engel A K and Gao S 2010 An independent brain–computer interface using covert non-spatial visual selective attention *J. Neural Eng.* **7** 016010

[31] Linden R, Picton T, Hamel G and Campbell K 1987 Human auditory steady-state evoked potentials during selective attention *Electroencephalogr. Clin. Neurophysiol.* **66** 145–59

[32] Ross B, Picton T W, Herdman A T and Pantev C 2004 The effect of attention on the auditory steady-state response *Neurol. Clin. Neurophysiol.* **22** PMID: 16012602

[33] Bidet-Caulet A, Fischer C, Besle J, Aguera P-E, Giard M-H and Bertrand O 2007 Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex *J. Neurosci.* **27** 9252–61

[34] Müller N, Schlee W, Hartmann T, Lorenz I and Weisz N 2009 Top-down modulation of the auditory steady-state response in a task-switch paradigm *Front. Human Neurosci.* **3** article 1

[35] Skosnik P D, Krishnan G P and O'Donnell B F 2007 The effect of selective attention on the gamma-band auditory steady-state response *Neurosci. Lett.* **420** 223–8

[36] Lopez M-a, Pomares H, Pelayo F, Urquiza J and Perez J 2009 Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain–computer interfaces *Neurocomputing* **72** 3617–23

[37] Kim D-W, Hwang H-J, Lim J-H, Lee Y-H, Jung K-Y and Im C-H 2011 Classification of selective attention to auditory stimuli: toward vision-free brain–computer interfacing *J. Neurosci. Methods* **197** 180–5

[38] Schalk G, McFarland D, Hinterberger T, Birbaumer N and Wolpaw J 2004 BCI2000: a general-purpose brain–computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

[39] Hill N J, Schreiner T, Puzicha C and Farquhar J 2007 BCPy2000 http://bci2000.org/downloads/BCPy2000

[40] Wolpaw J R, Ramoser H, McFarland D J and Pfurtscheller G 1998 EEG-based communication: improved accuracy by response verification *IEEE Trans. Rehabil. Eng.* **6** 326–33

[41] Lubar J F, Swartwood M O, Swartwood J N and Timmermann D L 1995 Quantitative EEG and auditory event-related potentials in the evaluation of attention-deficit/hyperactivity disorder: effects of methylphenidate and implications for neurofeedback training *J. Psychoeduc. Assess.* **ADHD Special** 143–60

[42] Fuchs T, Birbaumer N, Lutzenberger W, Gruzelier J H and Kaiser J 2003 Neurofeedback treatment for attention-deficit/hyperactivity disorder in children: a comparison with methylphenidate *Appl. Psychophysiol. Biofeedback* **28** 1–12

[43] ASHA 2005 (Central) Auditory processing disorders *Technical Report* The American Speech-Language-Hearing Association

[44] Lambert S 1989 Simultaneous interpreters: one ear may be better than two *TTR: Traduction, Terminologie, Rédaction* **2** 153–62