Toward a reliable gaze-independent hybrid BCI combining visual and natural auditory stimuli

Sara Barbosa[1], Gabriel Pires[1, 2] and Urbano Nunes[1,3]

[1]Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal
[2]Department of Engineering, Polytechnic Institute of Tomar, Tomar, Portugal
[3]Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal

E-mail: sbarbosa@isr.uc.pt, gpires@isr.uc.pt, urbano@isr.uc.pt

Background: Brain computer interfaces (BCIs) are one of the last communication options for patients in the locked-in state (LIS). For complete LIS patients, interfaces must be gaze-independent due to their eye impairment. However, unimodal gaze-independent approaches typically present levels of performance substantially lower than gaze-dependent approaches. The combination of multimodal stimuli has been pointed as a viable way to increase users' performance.
New method: A hybrid visual and auditory (HVA) P300-based BCI combining simultaneously visual and auditory stimulation is proposed. Auditory stimuli are based on natural meaningful spoken words, increasing stimuli discrimination and decreasing user's mental effort in associating stimuli to the symbols. The visual part of the interface is covertly controlled ensuring gaze-independency.
Results: Four conditions were experimentally tested by 10 healthy participants: visual overt (VO), visual covert (VC), auditory (AU) and covert HVA. Average online accuracy for the hybrid approach was 85.3%, which is more than 32% over VC and AU approaches. Questionnaires' results indicate that the HVA approach was the less demanding gaze-independent interface. Interestingly, the P300 grand average for HVA approach coincides with an almost perfect sum of P300 evoked separately by VC and AU tasks.
Comparison with existing methods: The proposed HVA-BCI is the first solution simultaneously embedding natural spoken words and visual words to provide a communication lexicon. Online accuracy and task demand of the approach compare favorably with state-of-the-art.
Conclusions: The proposed approach shows that the simultaneous combination of visual covert control and auditory modalities can effectively improve the performance of gaze-independent BCIs.

Keywords: Hybrid visual-auditory, Gaze-independent, P300, BCI, Covert control, Meaningful natural spoken words

1. Introduction

Brain-Computer Interfaces (BCIs) might be one of the last communication options for patients in the locked-in state (LIS). These interfaces rely on the interpretation of brain signals, from which it is possible to decode user's intentions, and convert them into commands to control, for example, communication devices and mobility systems (Pires et al., 2012; Lopes et al., 2013; Nijboer et al., 2008). BCIs based on different interaction paradigms allow patients with severe motor impairment to restore useful communication and interaction (Kübler et al., 2009; Sellers et al., 2010).

Locked-in state patients are described as being fully awaked and conscious but unable to perform fine motor movements and having affected speech (Laureys et al., 2005). Patients in locked-in state may show different levels of motor impairment. According to Baeur et al. (Bauer et al., 1979), locked-in state can be classified into classical, incomplete and complete LIS. The classical LIS presents total immobility except for vertical eye movements and blinking, the incomplete LIS can present recovery of some voluntary movements in addition to eye movement, and the complete or total LIS (CLIS) consists of total immobility including horizontal and vertical ocular movements. The locked-in syndrome may emerge from several neurological diseases such as stroke, muscular dystrophy, high spinal cord injury or Amyotrophic Lateral Sclerosis (ALS) (Kübler and Birbaumer, 2008). The latter is characterized by a progressive degeneration of motor neurons. Patients in this condition progressively lose the ability to control voluntary movements and in some cases enter in a complete locked-in state, in which they lose the ability to control any muscle of their bodies including the eyes. Direct gaze, adjust focus or perform eye-blinks is difficult or even impossible, being extremely challenging to control gaze-dependent interfaces (Kübler and Birbaumer, 2008; Schreuder et al., 2010; Fuchino et al., 2008). Despite their quadriplegia and aphonia (Kuehlmeyer et al., 2012), CLIS patients are still aware of themselves and the surroundings, being able to produce and control brain signals and therefore control brain-computer interfaces (BCIs). BCIs can be based on different neural signals, of which the electroencephalogram (EEG) is the most common. EEG-based BCIs can be controlled by neural mechanisms internally induced (e.g. modulation of sensory motor rhythms) or externally evoked (e.g. steady-state visual evoked potentials and P300 event related potential). BCIs based on these neural mechanisms have already been successfully tested with people with severe motor disabilities (Lesenfants et al., 2014; Combaz et al., 2013; Pichiorri et al., 2015). However, BCIs based on the P300 event related potential have been the most used in clinical settings (Mak and Wolpaw, 2009). In contrast to steady-state visual evoked potentials, which are gaze-dependent (Lesenfants et al., 2014), and to sensorimotor rhythms that require long periods of training with no guaranteed successful operation (McFarland and Wolpaw, 2008), P300-based BCIs require minimal training and can be gaze-independent. The P300 is an event related potential (ERP) evoked by a rare and relevant target event presented in an oddball paradigm. P300 is characterized by a positive peak that occurs around 300 ms after the onset of the target stimulus (Kübler et al., 2009; Farwell and Donchin, 1988). The control of P300-based BCIs require selective attention and working memory, thus depending of the complexity of the paradigm (Käthner et al., 2014).

The classic P300-speller and variants have been successfully used in home settings by ALS patients retaining eye movements (Nijboer et al., 2008; Sellers et al., 2010; Kübler and Birbaumer, 2008; Käthner et al., 2015b). Yet, CLIS patients are unable of gazing peripheral stimuli, outside the foveal vision due to their visual impairment (Bauer et al., 1979; Schreuder et al., 2010). Hence, other interfaces rather than the classic visual speller matrix (Farwell and Donchin, 1988) have to be considered. Several P300 gaze-independent BCIs have been proposed, namely visual covert (Treder and Blankertz, 2010; Pires et al., 2011a), auditory and tactile interfaces (Halder et al., 2013; Rutkowski and Mori, 2014). However, both visual covert and auditory P300 BCIs achieve significantly lower accuracies than visual interfaces controlled overtly (i.e., directing gaze) and therefore need to be improved to become a reliable option for CLIS patients (Kübler and Birbaumer, 2008; Riccio et al., 2012; Kaufmann et al., 2013). Tactile based interfaces (Severens et al., 2014) have also been recently pointed as another possible solution. However, tactile interfaces may not be the most appropriate solution for CLIS patients because the skin mechanoreceptors information might not reach the cortical areas (Murguialday et al., 2011). Since CLIS patients usually have an intact hearing system (Sorrentino and Remmert, 2014), the auditory approach might be one of the last options for communication (Höhne et al., 2011).

The design of P300-based interfaces for CLIS patients has to take into consideration their particular skills. Several studies report that these patients may suffer of reduced attention spans and cognitive impairment, preventing them to successfully control highly demanding tasks (Kübler et al., 2009; Lesenfants et al., 2014; Birbaumer, 2006; Sellers et al., 2003). For example, studies with auditory versions of the classic matrix speller and 2-step approaches showed that an additional attention and working memory effort is required to memorize and associate auditory stimuli (column, row or group of characters) with the position of the characters in the matrix speller and in the 2-step group speller (Kübler et al., 2009; Käthner et al., 2015b; Halder et al., 2013; Höhne et al., 2011; Sellers et al., 2003; Hill et al., 2014; Ikegami et al., 2014), thereby leading to unsatisfactory levels of BCI control. Less demanding BCI-tasks have been proposed for LIS and CLIS patients, based on simpler 4-choice (Sellers and Donchin, 2006) and 2-choice (binary) paradigms (Hill et al., 2014; De Vos et al., 2014; Halder et al., 2010; Pokorny et al., 2013; Hill et al., 2004; Hill and Schölkopf, 2012; Miner et al., 1998; Birbaumer et al., 2012). Also, in medical care centers non-BCI interaction/communication with LIS patients is mostly made by means of “yes/no” questions, using for example two eye blinks to express a “no” or only one eye-blink for the “yes” case. Spelling is possible by showing or saying all the letters of the alphabet and waiting for the indication of a “yes” to select a letter. Several simple BCI paradigms were developed based on this “yes/no” principle.

We propose a P300-based gaze-independent BCI, which combines auditory stimuli and visual stimuli detected covertly. The proposed bimodal interface comprises a reduced lexicon composed of the following words: *yes*, *no, hunger*, *thirst*, *urinate, air* and *position*. This 7-word paradigm intends to provide the possibility of a basic “yes and no” conversation with a partner/caregiver, providing at the same time the possibility of the patient to express some basic needs by his/her own initiative. The auditory stimuli are based on natural meaningful spoken words to minimize memory and cognitive demand and to maximize the stimuli discrimination and intuitiveness. The combination of auditory and visual modalities intends to decrease task difficulty and increase stimuli perception, in order to achieve a more discriminable P300 ERP. Although a BCI based on a reduced lexicon can be quite limitative when compared with a speller, it may represent one way to achieve a feasible BCI to be controlled by CLIS patients. Also, the low life expectancy at this advanced stage of the disease (Vianello et al., 2011) motivated us to develop a simple operating interface that would be easy to learn and adapt to.

1.1. State-of-the-art: auditory and hybrid P300-based BCIs

1.1.1. Auditory BCI

BCIs based on auditory stimuli represent an alternative solution for CLIS patients, since there is no need for eye movements to operate such interfaces. Though still relatively recent and few explored with LIS and CLIS patients, different types of auditory paradigms have been already researched and operated by healthy participants and by several LIS patients. Halder *et al.* (Halder et al., 2013) proposed a five-by-five matrix speller, with 25 letters, where different auditory stimuli are assigned to rows and columns. The stimuli consist of the spoken numbers from 1 to 5. A gaze-dependent visual BCI was also tested for comparison. The mean accuracy in online tests was 95.4% (for 3.41 repetitions) for the visual and 62.9% (for 9.01 repetitions) for the auditory BCI. Höhne *et al.* (Höhne et al., 2012) used three spatial directions for sound presentation (left ear, right ear and both ears) and compared three different stimuli: tones, spoken syllables and singed syllables. The offline results showed a slightly higher performance for singed and spoken syllables with average performance results of 63.4% for tones, 64.7% for spoken syllables and 65.8% for singed syllables. Overall, participants rated the natural stimuli as more ergonomic than the artificial stimuli. Schreuder *et al.* (Schreuder et al., 2010) used five speakers at ear's height, spaced evenly with 45º angle between them, covering the 180º frontal area, placed approximately one meter from the subject. The auditory stimuli consisted of five tones: 440 Hz, 494 Hz, 554 Hz, 622 Hz and 699 Hz and with that configuration they achieved a classification accuracy of 74.1%.

Most of the results achieved so far in auditory BCI are encouraging, though they have been obtained mostly in experiments with healthy participants. Although the results cannot be inferred for advanced LIS patients, the achievements with healthy participants are definitely relevant for initial assessment of the interface and for the development of more advanced interfaces. Kübler *et al.* (Kübler et al., 2009) tested the auditory BCI used in (Halder et al., 2013) with four LIS patients able to perform ocular movements. The results were quite discouraging with a high maximum accuracy of 25%, much lower than the accuracy achieved by the same four LIS patients using the five-by-five matrix with visual stimuli (Nijboer et al., 2008). Though, the healthy control group performed just slightly worse with the auditory than with the visual BCI (Furdea et al., 2009). The results showed that the performance for LIS patients cannot be directly inferred from the performance obtained from healthy participants, and that the design of the BCI paradigms should be adapted to characteristics of the patients. The authors also concluded that there must be a considerably improvement on auditory BCIs before they become a more reliable option. Pokorny *et al.* (Pokorny et al., 2013) compared the performance achieved by healthy subjects and minimally conscious state (MCS) patients, recurring to two streams of short beep tones as stimuli. Healthy subjects had an offline performance above chance level and sufficient for communication purposes, however only a small percentage of the MCS patients performed above chance level. The authors concluded that the promising performance results achieved from the healthy participants do not present a guarantee of good results with patients. However, the developed paradigm might offer the opportunity to assess unresponsive patients and eventually, a mean of communication.

The properties of the auditory stimuli have shown to be relevant factors and have to be considered when projecting high performance interfaces. Käthner *et al.* (Käthner et al., 2013) observed the differences on the performance when considering different values for interstimulus interval (ISI). The results showed that for lower values of ISI, the performance dropped considerably, indicating a more difficult task. For 160 ms, the achieved average performance was 17%, whereas for ISI values of 560 ms, the achieved mean performance was 66%. In (Höhne et al., 2011), participants reported that auditory stimuli based on tone sounds are difficult to differentiate, due to unintuitive identification of the stimuli and some users even described them as unpleasant. This fact might be a drawback in the acceptance of auditory BCIs by their main target users (Höhne et al., 2012). Solutions have been proposed to replace tones or syllables by meaningful words. Guo *et al.* (Guo et al., 2009) used eight spoken Chinese words for an auditory BCI. The stimuli consisted of the sequence of 8 digits (1, 2 …8) spoken in Chinese. With this configuration the authors achieved a mean offline performance of 89.1% considering trials with 15 repetitions. The higher performance compared with the other approaches relying on tone sounds stimuli, shows that this can be a viable approach to increase the reliability of auditory BCIs.

Table 1. Relevant studies of auditory-based BCIs: results of offline and online sessions.

Authors	Participants	Stimuli properties (number of choices)	Performance (repetitions) Offline classification	Online accuracy
Halder, et al., 2013	14 healthy	*Numbers:* 1 to 10 *(25-choice)*	-	62.9% (maximum:15)
Höhne, et al., 2012	9 healthy	*Tones:* 308 Hz left, 524 Hz right and 708 Hz binaural *(9-choice)*	63.4% (14)	-
		Spoken syllables: left, right, binaural *(9-choice)*	64.7% (14)	
		Singed syllables: left, right, binaural *(9-choice)*	65.8% (14)	
Schreuder, et al., 2010	5 healthy	*Tones:* -90º 440 Hz, -45º 494 Hz, 0º 554 Hz, 45º 622 Hz and 90º 699 Hz *(5-choice)*	74.1% (15)	-
Kübler, et al., 2009	4 LIS Patients	*Numbers:* 1 to 10 *(25-choice)*	37.4% (15)	12.1% (15)
Pokorny, et al., 2013	10 heathy	*Tones:* Standard 396 Hz, Deviant 297 Hz *(2-choice)*	83.5% (10)	-
	12 MSC		53.5% (10)	
Käthner, et al., 2013	20 healthy	*Tones:* 440 Hz, 494 Hz and 699 Hz, Directions: -90º, -60º, 0º, 60º and 90º, Different ISIs *(25-choice)*	-	560 ms - 66% (10) 400 ms - 65% (10) 320 ms - 51% (10) 240 ms - 42% (10) 160 ms - 17% (10)
Guo, et al., 2009	8 healthy	*Numbers:* 1 to 8 *(8-choice)*	89.1% (15)	-
Halder, et al., 2010	20 healthy	*Tones* — Pitch: 1000 Hz 75dB, 100 Hz 75dB, pink noise 75dB *(2-choice)*	92.5% (25)	-
		Tones — Loudness: 1000 Hz 65dB, 1000 Hz 75dB, pink noise 75dB *(2-choice)*	79.2% (25)	
		Tones — Direction: 100 Hz 75 dB left, 100 Hz 75 dB right, pink noise 75 dB stereo *(2-choice)*	90.8% (25)	
Hill, et al., 2014	14 healthy	*Words:* Male "no", random "nope", Female "yes", random "yep" *(2-choice)*	-	77 % (20) – cued
		Tones: 512 Hz left and 768 Hz right *(2-choice)*		73 % (20) – cued
	2 ALS	*Words:* Male "no", random "nope", Female "yes", random "yep" *(2-choice)*		84.4 % (20) – free 67.5 % (20) – cued
Käthner, et al., 2015a	1 ALS	*Tones:* 1000 Hz, 100 Hz and pink noise *(2-choice)*	-	91.6 % (20), 79 % (10), 50 % (7)

The low online classification accuracies attained with auditory BCIs suggest that these paradigms demand high selective attention and cognitive workload. As referred previously, this is particularly relevant since LIS and CLIS patients may suffer of reduced attention spans and cognitive impairment (Birbaumer, 2006). Such impairments can prevent them to successfully control highly demanding tasks and therefore simpler solutions must be designed for them. Several binary auditory paradigms have been proposed and evaluated to deliver simpler solutions. Halder *et al.* (Halder et al., 2010) implemented a binary oddball paradigm by using three stimuli. They have also evaluated three conditions by varying the pitch, loudness and direction. The results showed that the loudness differences were not as easy to discriminate as the pitch and the direction. The pitch condition achieved the higher performance for 14 subjects. In (Hill et al., 2014), Hill *et al*. compared two binary paradigms, one based on spoken words and one based on tones. The healthy subjects had a higher performance with spoken words. Two ALS patients tested the word-stimuli condition and obtained a performance well above chance level but using a high number of repetitions. Käthner *et al.* (Käthner et al., 2015a) compared the performance of a LIS patient using an eye-tracking, an electrooculography system and a binary auditory BCI. The auditory approach was tested with three tone sounds and the online results showed an effective level of performance for 20 repetitions. Participant considered to use the binary auditory BCI once he was no longer able to control ocular movements. When comparing the addressed binary choice paradigms with the multiclass paradigms tested with LIS patients (Kübler et al., 2009), it is evident that the consideration of simpler paradigms for these patients is crucial to attain performance results feasible enough to develop a successful BCI. Table 1 summarizes the offline and online results achieved in the above auditory-based BCIs.

1.1.2. Hybrid BCI

Although binary auditory paradigms have demonstrated feasible online accuracies, they might be limitative when compared with paradigms with higher number of choices. There have been increasing efforts to achieve better results with the use of multichoice auditory stimuli in BCI. One approach is to combine the auditory stimuli with other modalities. For example, Rutkowski *et al.* (Rutkowski and Mori, 2014) developed a P300-based hybrid BCI with auditory and tactile stimuli delivered simultaneously by vibrotactile exciters, benefiting from bone-conduction effect for audio. Experimental results with healthy users validated the BCI. A most common approach has been the combination of visual and auditory stimuli. In (Cui et al., 2014) the participants were asked to perform gaze-dependent (overt control) visual and auditory tasks. The participants had to focus a target image and simultaneously an audio stimulus. The P300 components showed to have a greater amplitude in the hybrid task, reflecting positively on the classification accuracy. For the visual approach a 86.7% average accuracy was achieved, whereas for the hybrid approach a 96% average accuracy was achieved. This result is however influenced by the overt control of the visual component, and therefore the control is unsuitable for CLIS patients.

An *et al.* (An et al., 2014) compared four different gaze-independent conditions related to sensory modalities: visual speller, auditory speller, simultaneous auditory and visual stimuli and alternating auditory and visual cues. In the simultaneous auditory and visual stimuli task, participants were asked to concentrate on both visual and auditory stimuli. Better offline accuracies were observed for visual speller (85.4%) and simultaneous auditory and visual stimuli (83.0%) conditions comparing to the auditory speller (62.4%). For the simultaneous auditory and visual stimuli, an effective mean online accuracy of 92% was achieved using 6 repetitions. Table 2 summarizes the offline and online results achieved in the above hybrid BCIs. Results show that the combination of multimodal stimuli increases the overall performance of the BCI, however the proposed approaches still lack of validation with LIS and CLIS patients.

Table 2. Relevant studies of hybrid (auditory and visual) based BCIs: results of online sessions.

Authors	Participants	Gaze	Stimuli properties (number of choices)	Online accuracy (repetitions)
Cui, et al., 2014	3 healthy	Dependent	*Face image + spoken syllables:* Cued by number and arrow + "a", "i", "u", "e", "o", en", "da" and "go" *(8-choice)*	96% (2)
An, et al., 2014	15 healthy	Independent	*Visual symbols + spoken syllables:* unique geometrical shape and color visual symbols + "ti" and "to" left, "ti" and "ot" middle, "ti" and "to" right, 3 pitches *(6-choice/step, 2-step)*	92% (6)
			Alternated visual symbols and spoken syllables: unique geometrical shape and color visual symbols and "ti" and "to" left, "ti" and "ot" middle, "ti" and "to" right , 3 pitches *(6-choice/step, 2-step)*	87.7% (10)
Our method HVA-BCI	10 healthy	Independent	*Word image + spoken words*: Words within a specific lexicon + Natural spoken words within a specific lexicon *(7-choice)*	85.3% (6)

2. Materials and Methods

2.1. Participants

Ten healthy participants (P1-P10) took part in the study (6 male, 4 female, mean age 26.2 years, SD 5.8 years, range 21-42). All of them signed an informed consent to participate in the study. None of the participants possessed hearing impairment, and all had a normal or corrected to normal vision. Five participants (P1-P5) had previous experience with P300 BCI, one participant had experience with steady state visual evoked potential BCI (P8) and the others had never controlled a BCI before (naïve users). Table 3 shows their identification codes, ages, gender and BCI experience.

Table 3. Gender, age and BCI experience of participants.

	Participant									
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
Age	24	23	26	42	24	26	26	25	25	21
Gender	F	F	F	M	M	M	M	M	F	M
BCI experience	VO, VC and AU	VO	VO	VO and VC	VO, VC and AU	Naïve	Naïve	SSVEP	Naïve	Naïve

F – Female, **M** – Male, **VO** – Visual Overt P300, **VC** – Visual Covert P300, **AU** – Auditory P300, **SSVEP –** Steady State Visual Evoked Potential

2.2. Data acquisition and experimental setup

EEG and EOG data were recorded with two 16-channel gUSBamp acquisition systems (g.tec medical engineering GmbH, Austria). EEG was acquired with active Ag/AgCl electrodes and EOG was captured with disposable pre-gelled Ag/AgCl electrodes. The signals were labeled and classified using g.tec Highspeed Simulink^{TM} framework, which was also used to trigger visual and auditory stimuli. Visual stimuli were designed in Simulink^{TM} while auditory stimuli were designed in Presentation$^{®}$ (Neurobehavioral Systems, version 16.5). Simulink^{TM} and Presentation$^{®}$ running on the same computer were synchronized through TCP/IP communication. The experimental setup is shown in Fig. 1A. Data processing and storage were performed on a TOSHIBA laptop (Intel Core i7 2.30 GHz, 8.00 GB RAM, Microsoft Windows 7 Ultimate).

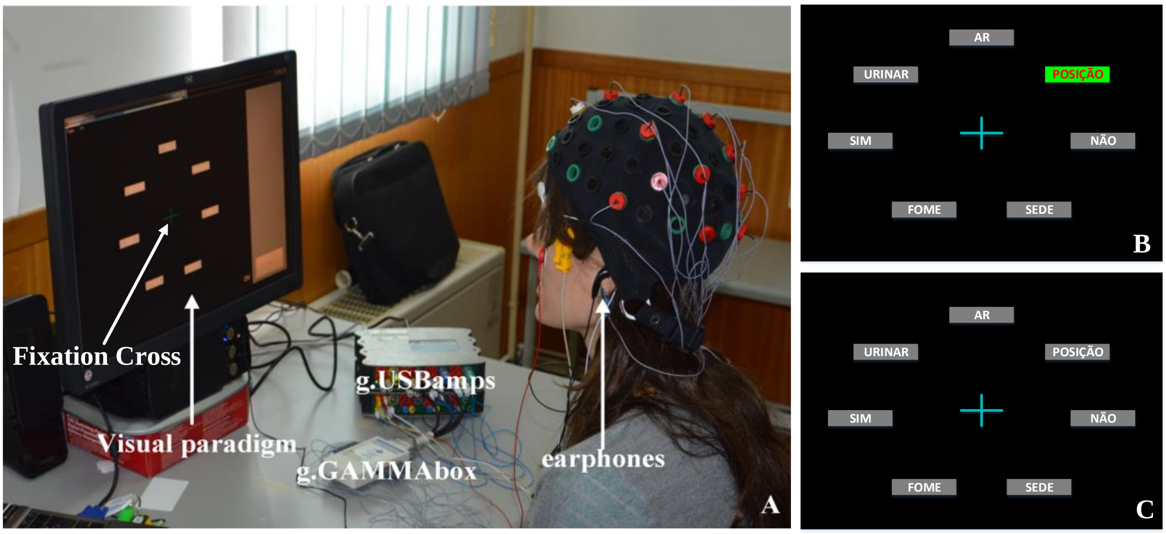


Figure 1. (A) Experimental setup; Arrangement of the seven visual stimuli, the blue cross is used as the fixation point for the covert attention tasks; **(B)** Highlighted event "POSIÇÃO" (the foreground/background color is changed from white/gray to red/green); **(C)** Visual matrix with events off.

EEG signals were recorded monopolarly using 18 electrodes, placed accordingly to the extended international 10-20 System. The right or left earlobe was selected for reference and the AFz channel for ground. Sixteen channels were selected: Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, Oz, T7, CP5, T8 and CP6. The first 12 channels were selected respecting the visual task (Pires et al., 2012) and the last four channels were selected considering the auditory cortex location (Furdea et al., 2009; Purves et al., 2004). EOG data were recorded for monitoring of eye movements during the covert attention tasks. Four EOG electrodes were mounted in a bipolar configuration, two of them were used to monitor the horizontal movements (h-EOG) and the others for the vertical movements (v-EOG). EEG and EOG signals were sampled at 256 Hz and filtered by a band-pass filter between 0.5 and 30 Hz and a notch filter at 50 Hz. Data processing and further analysis was performed in Matlab, the data were epoched between the stimulus onset and 1 second after.

2.3. Experimental design

2.3.1. BCI conditions

Three P300-based gaze-independent conditions were tested, one with visual stimuli, one with auditory stimuli and a hybrid condition, combining the two types of stimuli. A gaze-dependent condition was also tested for control, to verify if the participants could operate a P300-based BCI and to assess their maximum performance. Therefore, each participant tested 4 conditions: visual paradigm controlled overtly (VO – Visual Overt), visual paradigm controlled covertly (VC – Visual Covert), hybrid visual and auditory paradigm controlled covertly (HVA – Visual and Auditory) and auditory paradigm (AU). Each test consisted of a calibration session and a subsequent online session. The order of the tests for each participant was pseudo-randomized with a restriction, the participants had always to perform the paradigm HVA prior to the paradigm AU. This restriction was set so that the participants became familiar with the auditory stimuli before the purely auditory task. After the calibration and online sessions for each one of the four tests, the participants had an interval in which they were asked to answer a simple questionnaire (Quest I) based on the NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988) in order to assess several task parameters such as mental, physical and temporal demand, performance, effort and frustration. This interval also allowed participants to rest between different tests. After completing the four conditions, participants were asked to answer a second questionnaire (Quest II) to compare the pleasantness of the four conditions.

2.3.2. Tasks design

A set of seven words comprising a small communication lexicon was used in visual and auditory oddball tasks, namely, SIM (Yes), NÃO (No), FOME (Hunger), SEDE (Thirst), AR (Air, regarding the breathlessness sensation), POSIÇÃO (Position, regarding the need of position changing due to pain) and URINAR (Urinate). The words were chosen with the help of caregivers and aimed to provide the possibility of a "YES/NO" conversation with a partner, allowing at the same time that CLIS patients could express some basic needs on their own initiative. The low complexity of the task, based on a small set of symbols, and the use of stimuli based on visual words and spoken words with meaning, are intended to increase involvement of patients with the task, improve their selective attention and reduce their working memory demands.

Each task consisted of a 7-stimuli oddball paradigm, where stimuli could be presented purely visual (VO and VC), purely auditory (AU) or simultaneously visual and auditory (HVA). Each trial consisted of 6 sequences (7 random stimuli per sequence), as presented in Fig. 2. On each sequence there was one target stimulus and 6 standard stimuli. The design of the oddball sequences was the same for the four test conditions, except for the modality of the stimuli. Participants were asked to attend the target stimuli and to mentally count them ignoring all other standard events. The system returned the selected symbol after one trial (6 sequences). Participants had to complete a set of 15 trials (selection of 15 words). The set of words was the same for all participants and for the 4 conditions. The time between each selection (inter-trial interval - ITI) was 8 s. The 8 s value was selected to allow the participants to rest and perform ocular movements between trials. To gather data to train the classifier, participants performed a calibration session before the online operation, which consisted of 14 trials each one with 10 sequences (calibration takes about 11 min). At the end of the calibration, 140 target epochs and 840 standard epochs were gathered.

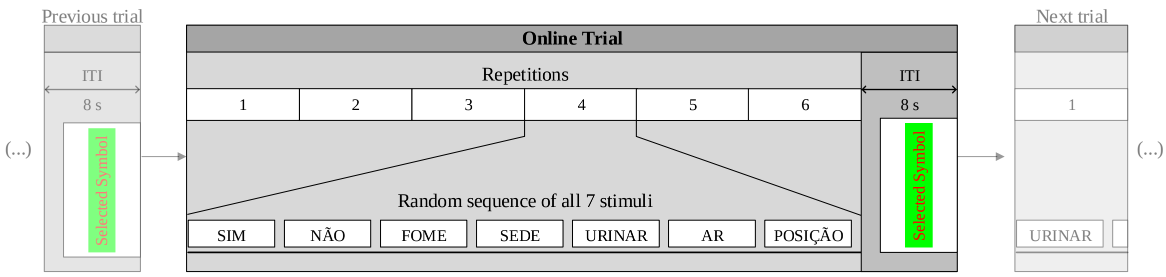


Figure 2. Illustration of an online trial and inter-trial interval (in which it is returned the selected symbol).

Figure 3 shows the temporal diagram of the stimuli of one sequence for each modality. All stimuli have a duration of 450 ms and an interstimulus interval of 100 ms (stimuli onset asynchrony of 550 ms). These times were set based on the minimum requirements for auditory stimuli. The 450 ms was set as the minimum time required to pronounce the words comprising the lexicon. The ISI of 100 ms was set as the minimum interval required to distinguish two consecutive words. In order to provide the same parameter settings for the 4 conditions, we decided to keep these values also for the unimodal visual tasks, despite knowing that these values could be considerably shortened (Pires et al., 2012). For VO and VC tasks, only the visual stimuli are presented to the participant, and in task AU, only auditory stimuli are triggered. In the HVA condition, visual and auditory stimuli are triggered simultaneously. Before each trial, participants were informed of which target to attend. The target stimulus to attend was provided visually for VO, VC and HVA at the top of the screen and orally for the AU task. Feedback (selected word) was visually provided at the end of each trial for the VO, VC and HVA conditions. For the AU condition no feedback was presented to the subject at the end of each trial, only after completing the 15 selections. The time between selections was 23.1s ($6 \times 7 \times 0.550$ s) plus the ITI of 8 s. Participants performed the four conditions in a single experiment, which took about 3 h to complete. This period also included the EEG and EOG mounting, EOG calibration and a period for contact and familiarization with the stimuli and interface arrangement.

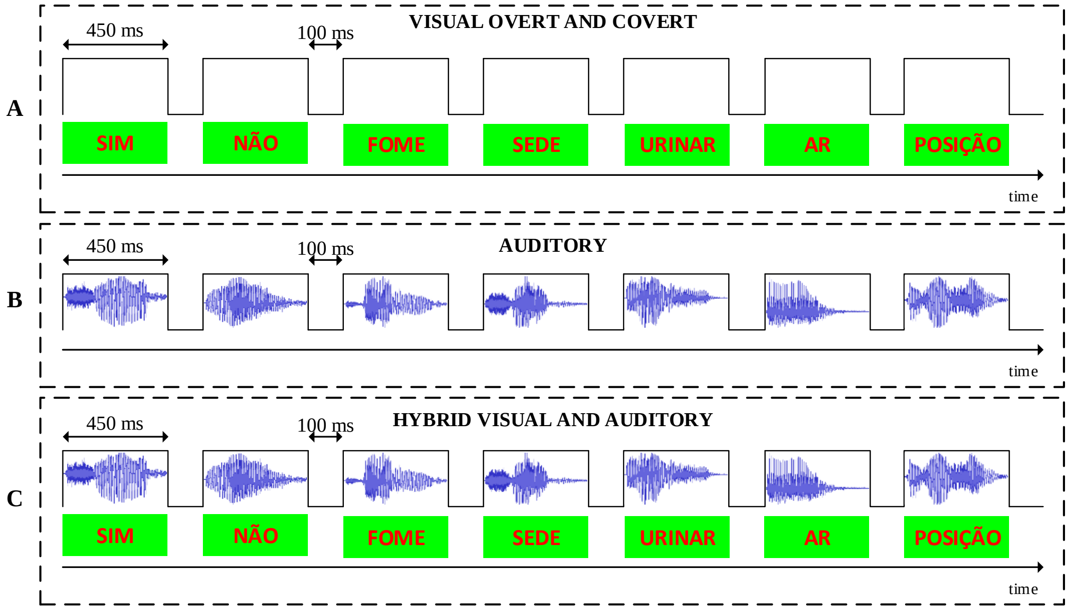


Figure 3. Temporal diagram of stimuli presentation for each modality. All stimuli have a duration of 450 ms and an interstimulus interval of 100 ms (stimuli onset asynchrony of 550 ms). (A) In VO and VC conditions, the stimuli are visual (B) In the AU condition, the stimuli are spoken words; (C) In the HVA condition, the visual and auditory stimuli are presented simultaneously using the same duration and ISI.

Visual tasks

During the three visual-based tasks (VO, VC and HVA) the participants sat in a comfortable chair, approximately 50 cm from a 22-inch screen. Screen height was adjusted so that the participants' eyes were at the same level as the center of the screen. Figure 1 shows, respectively, the layout of the visual paradigm during a stimulus flashing (Fig. 1B) and without stimulus (Fig. 1C). In the example of Fig. 1B, the visual stimulus corresponds to the highlight of "POSIÇÃO by a color change, of both the box (from gray to green) and the word (from white to red). A blue cross, in the center of the screen, is used as the fixation point for the covert attention tasks. Symbols were placed at the visual region of the fovea, ensuring that participants could perceive flashing events without having to gaze the target stimulus. For conditions VC and HVA the participants were asked to fix their gaze in the blue cross. For the VO condition the blue cross was still present in the screen however participants were told to freely gaze the target symbol.

Auditory and hybrid visual-auditory tasks

The auditory stimuli are the following Portuguese spoken words, recorded and rectified using the Audacity (Audacity Team, version 2.0.5) software: SIM (sˈĩ), NÃO (nˈɐ̃w̃), FOME (fˈɔmɘ), SEDE (sˈedɘ), URINAR (uɾinˈaɾ), AR (ˈaɾ) and POSIÇÃO (puzisˈɐ̃w̃). The phonetic transcription of each speech sound is in parentheses. By using meaningful spoken words we aimed to achieve a less demanding task in discriminating stimuli and to provide a direct relation between stimuli and semantic symbols, thereby trying to alleviate the difficulties reported in (Höhne et al., 2011) and (Höhne et al., 2012). Participants heard the auditory stimuli using earphones and could adjust the audio volume to a comfortable but yet well-perceived level.

2.4. Ocular movement detector

To ensure that covert attention tasks (VC and HVA) were effectively performed without moving the eyes from the blue cross, an ocular movement detector (OMD) based on EOG was developed to reject EEG epochs contaminated with ocular movements as in (An et al., 2014). This allowed us to ensure that participants would experience the same conditions of CLIS patients unable to move the eyes. Before starting the BCI tests, each participant underwent a 185 s calibration of the OMD. Participants were asked to perform a sequence of 11 eye movements while focusing the visual layout (Fig. 4A). The movement labelled with "1" is actually a non-movement where the participant had to focus on the central cross during that trial (see Fig. 4B with illustrative EOG recording while some of the movements were performed). The BCI paradigm was running during the EOG calibration in order to obtain data in similar conditions of that of BCI calibration and online sessions. The OMD detects ocular movements in the horizontal and vertical EOG data. Figure 5 shows the vertical and horizontal EOG signals for three distinct situations: ocular movements, no ocular movements and eye blinking. Healthy participants usually perform eye blinks several times during the experiments, and therefore OMD had to clearly discriminate eye blinks of eye movements. Considering only the amplitude to estimate a threshold detector, eye blinking would be identified as an ocular movement. So, OMD analyses three ranges of values to choose two threshold intervals, namely the maximum and minimum amplitude for 1) no movements (NM_{max} and NM_{min}); 2) ocular movements (OM_{max} and OM_{min}) and 3) eye blinking (EB_{max} and EB_{min}). The two thresholds were defined as follows:

$$Upper\ threshold\ interval = [NM_{max} \quad OM_{max}] \tag{1}$$

$$Lower\ threshold\ interval = [OM_{min} \quad NM_{min}] \tag{2}$$

The values EB_{max} and EB_{min} were evaluated in order to check for overlap of eye movements with the eye blinking. The two threshold intervals correspond to the gray areas represented on the ocular movements in Figure 5. Threshold parameters were set for each participant.

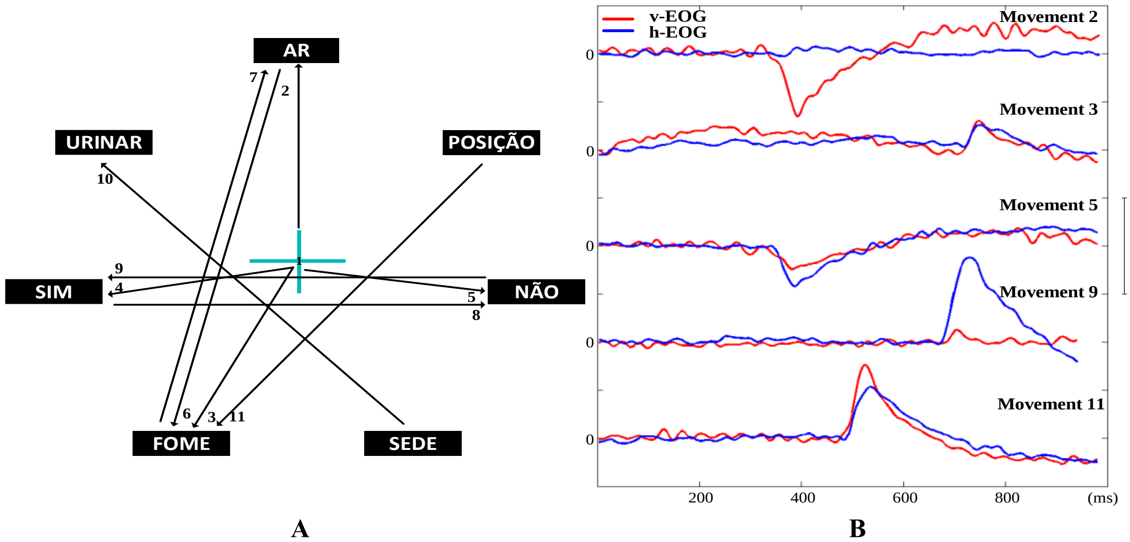


Figure 4. (A) Sequence of the 11 eye movements performed in the EOG calibration; **(B)** example of vertical and horizontal EOG signals for movements 2, 3, 5, 9 and 11.

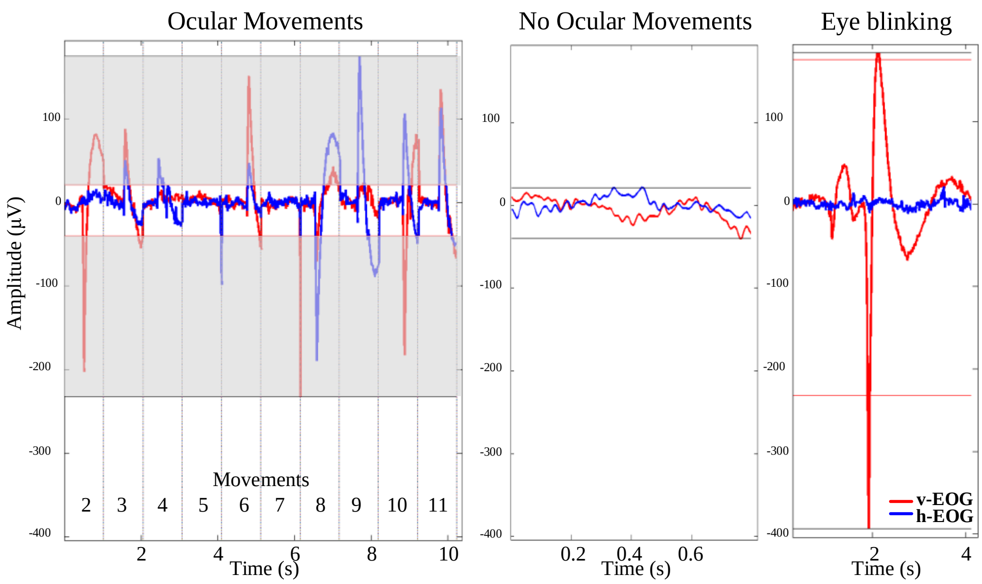


Figure 5. Vertical and horizontal EOG signals for ocular movements, no ocular movements and eye-blinking.

2.5. Classification model and online testing

The main challenge in EEG classification is the extraction of the most discriminant features for the best classification. Classification followed our approach presented in (Pires et al., 2011b) based on statistical spatial filters for feature extraction. The 16 EEG channels are filtered by a statistical spatial filter called CSP-FLD that combines sub-optimally two discrimination criteria, represented in Figure 6 by the "Spatial Filter W" block. From data projected by the spatial filter, the two most discriminative projections are concatenated resulting in a vector of 512 features (time samples). From these, 200 features are selected based on the R-square correlation method. Features are then classified with a Bayes classifier.

Figure 6 shows the overall classification process including the ocular movement detector used to reject EEG data contaminated with ocular movements. If an ocular movement is detected in the calibration session, either in the horizontal or in the vertical EOG, the OMD labels the epochs associated with that ocular movement. Then, three epochs from the EEG training data are rejected: the corresponding epoch in the EEG signal, the previous and the next epoch. The remaining epochs follow to the next phase. From the calibration session, a user-dependent classification model is saved to be used in the online session. This model includes the 200 indices of the selected features, and the classification and spatial filter models.

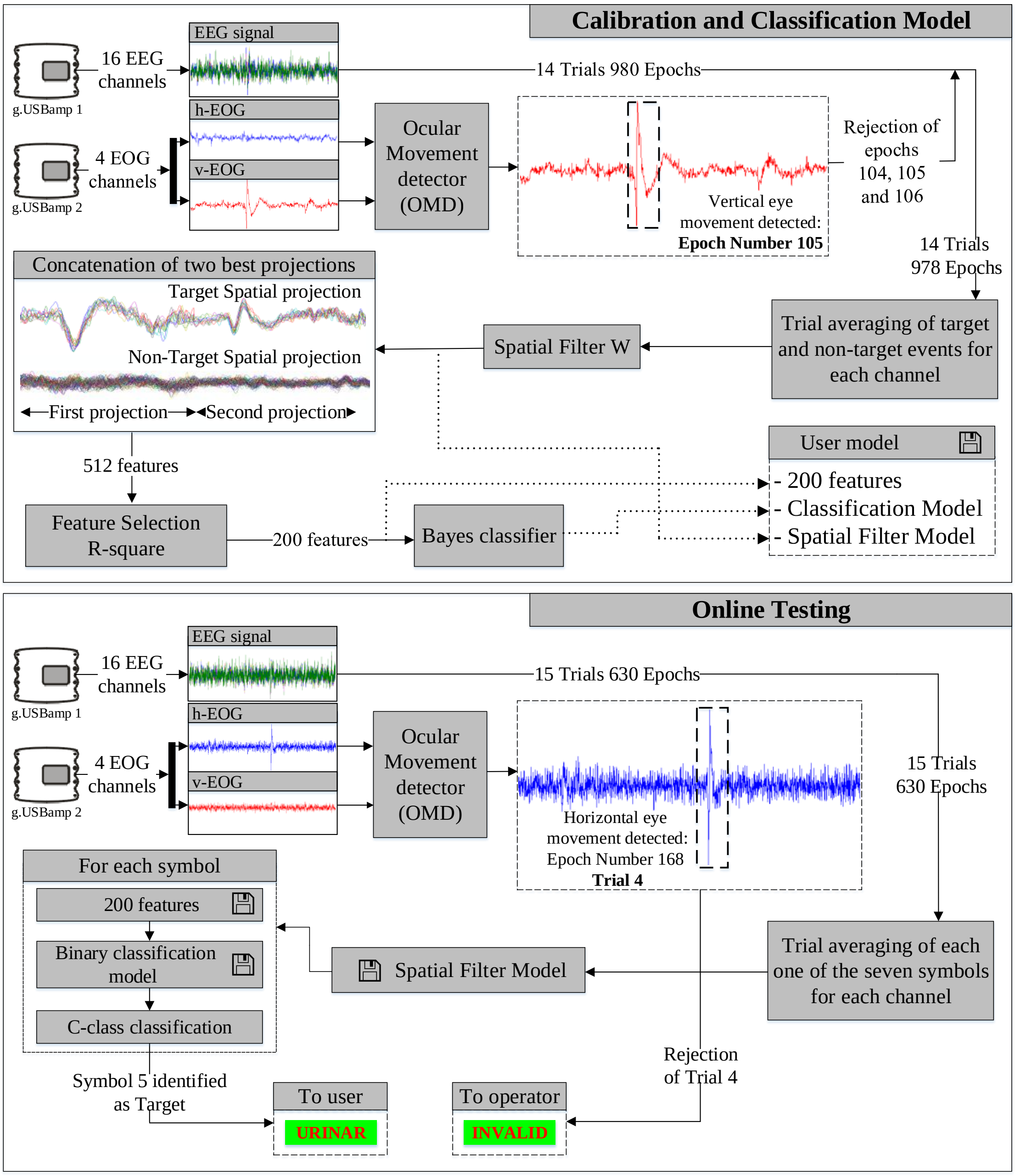


Figure 6. Overall procedure for offline and online classification for conditions VC and HVA. The classification procedure for conditions VO and AU are the same except that OMD is not used.

Classification models obtained from calibration are used to classify online each stimulus event as target or non-target. The binary classification model is applied to the EEG responses of the seven events, returning a probability value for each event. The symbol associated with the highest probability is chosen as being the symbol mentally selected by the user (block identified as "C-class classifier" where C is the number of classes, i.e., the number of symbols). If an ocular movement is detected in the EOG data, the corresponding trial is marked as invalid and not considered for assessment of the result although the participant receives the classification feedback. This procedure was followed so that the participant was not distracted and to not influence the experiment flow.

2.6. Online metrics

To compare the performance between the 4 BCI conditions and these with the state-of-the-art results, we computed the online classification accuracy, the information transfer rate (ITR) and the effective symbols per minute (eSPM). Online classification accuracy is the most important metric since it defines the feasibility of the interface. An interface with low accuracy is highly prone to errors, which is a demotivating factor for BCI usage. The interface speed was fixed to approximately 2.5 SPM (discarding the ITI time) for all participants (corresponding to 6 sequences per trial) and was settled as a tradeoff between accuracy and speed previously determined during pilot experiments. In clinical settings, the speed of the BCI would have to be adjusted to the needs and performance of patients (Kübler et al., 2015). A more realistic measure of speed is provided, the effective SPM (eSPM), which embeds the accuracy information (5). The information transfer rate (ITR), or bit rate, is also an important metric for state-of-the-art comparison since it embeds simultaneously the number of encoded symbols, the online accuracy and the SPM (Wolpaw et al., 2000). We evaluated the ITR in bits per minute (bpm) and the eSPM as follows:

$$ITR = SPM \times \left[log_2(N_s) + P_{ac}log_2(P_{ac}) + (1 - P_{ac})log_2\frac{(1 - P_{ac})}{(N_s - 1)}\right] \quad (3)$$

$$SPM = \frac{60}{N_{rep} \times (N_{ev} \times SOA) + 1} \quad (4)$$

$$eSPM = SPM \times P_{ac} \quad (5)$$

where N_s is the number of symbols, P_{ac} is the online accuracy, N_{rep} is the number of repetitions, N_{ev} is the number of events ($N_{ev} = N_s$), SOA is the stimulus onset asynchrony and the value 1 represents the time needed to record the epoch of the last event of the trial. The SPM was computed discarding the 8 s ITI.

3. Results

3.1. EOG calibration

Participants performed the EOG calibration task before starting their experimental session. The two EOG threshold intervals obtained individually for each one of the ten participants had the following mean values: (13.8 ± 3.7) for NM_{max}, (-13.2 ± 4.9) for NM_{min}, (165.7 ± 44.8) for OM_{max}, (-175.3 ± 52.4) for OM_{min}, (303.1 ± 205.3) for EB_{max} and (-403.5 ± 160.0) for EB_{min}. Participants presented ocular movements thresholds that do not overlap with the eye blinking thresholds.

In the HVA calibration phase an overall percentage of 2.00% of target and 2.05% of non-target EEG epochs were excluded from the calibration data as being affected by ocular movements. For the VC condition that percentage was 2.36% for the target and 2.21% for the non-target epochs. The low rate of epoch rejection can be explained by the long ITI of 8 s, which helped participants to rest and perform ocular movements between trials. For the online session, in the HVA condition 6.00% of trials were rejected whereas for the VC case 9.33% of the trials were excluded from the online results.

3.2. *Online accuracy and information transfer rate*

Each online session was preceded by a calibration session for that same condition. This was repeated for the four conditions. As already referred, the number of repetitions per trial was pre-set to 6 in preliminary experiments, as a tradeoff between classification rate and SPM. For this number of repetitions, it was expected a ceiling effect for the VO condition since participants usually control BCI with fewer repetitions, typically less than 4 with rates above 90%. VO condition was used mainly as a control condition to ensure that participants were able to use effectively a P300-based BCI. Participants had to select a sequence of 15 words. Table 4 shows the percentage of correct choices, P_{ac} , the ITR and the eSPM ($N_s = N_{ev} = 7$, $N_{rep} = 6$ and SOA $= 550$ ms) for the four conditions.

Table 4. Online session accuracy, ITR and eSPM for the four conditions (VO, VC, HVA and AU), with six repetitions

		VO			**VC**			**HVA**			**AU**		
		P_{ac}	ITR	eSPM	P_{ac}	ITR	eSPM	P_{ac}	ITR	eSPM	P_{ac}	ITR	eSPM
Participants	**P1**	100	6.99	2.49	92.3	5.52	2.30	93.3	5.68	2.32	93.3	5.68	2.32
	P2	86.7	4.72	2.16	53.3	1.50	1.33	86.7	4.72	2.16	6.7	0.10	0.17
	P3	100	6.99	2.49	14.3	0.00	0.36	78.5	3.74	1.95	33.3	0.41	0.83
	P4	86.7	4.72	2.16	66.7	2.56	1.66	92.3	5.52	2.30	46.7	1.08	1.16
	P5	100	6.99	2.49	78.5	3.74	1.95	92.9	5.52	2.31	100	6.99	2.49
	P6	80	3.90	1.99	20.0	0.04	0.50	57.1	1.78	1.42	33.3	0.41	0.83
	P7	80	3.90	1.99	78.5	3.74	1.95	85.7	4.60	2.13	80	3.90	1.99
	P8	100	6.99	2.49	33.4	0.41	0.83	100	6.99	2.49	46.7	1.08	1.16
	P9	100	6.99	2.49	40.0	0.71	1.00	80.0	3.90	1.99	53.3	1.50	1.33
	P10	100	6.99	2.49	46.7	1.08	1.16	86.7	4.72	2.16	33.3	0.41	0.83
Mean		93.3	5.92	2.32	52.4	1.93	1.30	85.3	4.72	2.12	52.7	2.16	1.31
Std		8.9	1.41	0.22	27.7	1.88	0.65	11.8	1.40	0.30	29.7	2.47	0.74

For the VC and AU conditions, we observe close mean accuracies, with results slightly lower for the VC case, however the paired t-test show that this difference is not statistically significant (paired t-test, p = 0.97). For the HVA condition we observe online performances of 32.9% and 32.6% higher than for the VC (paired t-test, p = 0.001) and AU condition (paired t-test, p = 0.005), respectively. The HVA condition was the gaze-independent solution that approached more the results of the VO condition, with a difference of 8% (paired t-test, p = 0.04). For the VC condition, the ITR was 0.23 bpm and the eSPM was 0.01 symbols per minute lower than for the AU condition. The HVA condition achieved more 2.79 bpm and 0.82 eSPM than the VC condition and more 2.56 bpm and 0.81 eSPM than the AU condition. Since the number of repetitions per trial is constant for all participants, the variation of the eSPM and ITR depends only on the variation of the accuracy. Therefore, the eSPM and ITR exhibit the same statistical behavior observed for the accuracy analysis.

A confusion matrix with the online classification accuracies is analyzed aiming to infer if visual adjacent distractors (Pires et al., 2011b) and similar phonetics could have produced systematic errors.

Visual Overt

Selected Targets (columns); Target to select (rows)

	1	2	3	4	5	6	7
1	94	3	0	0	3	0	0
2	0	95	0	0	0	5	0
3	0	0	95	0	0	0	5
4	0	0	0	100	0	0	0
5	0	0	0	0	100	0	0
6	5	0	0	0	0	95	0
7	5	5	10	0	0	0	80

Visual Covert

Selected Targets (columns); Target to select (rows)

	1	2	3	4	5	6	7
1	47	7	10	7	17	12	0
2	10	35	5	15	20	5	10
3	5	10	45	20	5	5	10
4	0	15	5	60	15	0	5
5	15	5	5	5	45	10	15
6	0	5	0	20	20	50	5
7	5	15	5	5	10	5	55

Hybrid Visual and Auditory

Selected Targets (columns); Target to select (rows)

	1	2	3	4	5	6	7
1	65	0	0	12	3	10	10
2	0	90	0	5	0	0	5
3	0	0	90	0	5	5	0
4	0	5	0	75	0	15	0
5	0	5	5	0	80	10	0
6	0	0	0	0	5	85	10
7	5	15	0	5	10	0	65

Auditory

Selected Targets (columns); Target to select (rows)

	1	2	3	4	5	6	7
1	60	3	0	7	13	7	10
2	25	40	5	5	15	5	5
3	5	0	55	15	20	0	5
4	10	10	5	45	0	10	20
5	5	25	0	0	50	15	5
6	5	10	0	5	25	55	0
7	0	15	5	0	0	20	60

Correct selection · Visual adjacent distractors · Similar phonetics · +

Figure 7. Confusion matrices with online accuracy (%) for the four situations (Target numbers correspond to: 1. SIM, 2. NÃO, 3. FOME, 4. SEDE, 5. URINAR, 6. AR and 7. POSIÇÃO)

Figure 7 presents the confusion matrices for the four conditions. For the visual paradigms (VO, VC and HVA) we evaluated the adjacent distractors as the possible justification for errors. Each target stimulus is surrounded by two adjacent stimuli acting as possible distractors. For the VO condition, there were only 2 error occurrences of adjacent distractors, while for the VC condition there were 36 errors, which may be explained by the more demanding stimuli perception required in covert attention. For the AU and HVA conditions, we analyzed whether words with similar phonetics produced systematic errors. We grouped the seven spoken words into three pairs of similar phonetics: 1) SIM (sˈĩ) and SEDE (sˈedɘ), 2) NÃO (nˈɐ̃w̃) and POSIÇÃO (puzisˈɐ̃w̃) and 3) URINAR (uɾinˈaɾ) and AR (ˈaɾ). The percentages of errors that could be explained by adjacent distractors and similar phonetics were 21%, 46%, 49% and 23% respectively for the VO, VC, HVA and AU conditions. Notwithstanding, the remaining errors occurred without a correlation with these two effects. We can hypothesize that these errors have a random nature caused by task difficulty and in some cases caused by perception effects occurring in particular sequences of the auditory stimuli, as reported by some participants (see section 3.5). However, we could not further analyse this, since the sequence of stimuli is random and was not registered during online sessions.

3.3. Amplitude, latency and width of P300 ERP

The amplitude, latency and width of the P300 waves were analyzed. The amplitude measures the value of the P300 peak, the latency is the instant of the P300 peak and the width is estimated as the interval between the zero-crossings before and after the maximum peak occurrence. These three measurements were obtained from the P300 epochs evoked by the target events gathered during the calibration sessions of the 10 participants taking the 16 channels, making a total of 2240 epochs per condition. Only the signal portion from 200 ms to 600 ms was analyzed, which corresponds to the range where the P300 occurs. The mean amplitude of the P300 wave for the condition VO is the highest (23.6 μV), followed by conditions AU (21.6 μV), HVA (20.1 μV) and VC (20.0 μV). Moreover, the VO amplitude is significantly higher than the VC amplitude (paired t-test, $p = 0.02$) and from the HVA condition (paired t-test, $p = 0.03$), but not significantly different from the condition AU (paired t-test, $p = 0.26$). The P300 wave has an average latency of 389.7 ms, 406.8 ms, 407.4 ms and 407.1 ms for the VO, VC, HVA and AU conditions, respectively. The latency for AU and VC conditions are statistically higher than the latency for the condition VO (paired t-test, $p = 0.03$, $p = 0.03$) and when comparing the VC and AU conditions no significant differences are observed (paired t-test, $p = 0.96$). The higher latency for the conditions VC and AU may be due to the increased difficulty in stimuli perception, as more time is needed for the participant to perceive the stimulus (Höhne et al., 2011). When comparing the latencies between the HVA and VC, AU and VO conditions, no significant difference was found (paired t-test, $p = 0.91$, $p = 0.96$, $p = 0.06$). Analyzing the P300 width, the results showed values of 116 ms, 119 ms, 125 ms and 111 ms for the VO, VC, HVA and AU conditions, respectively, however the differences were not statistically significant.

In Fig. 8 the grand average of target and non-target EEG signals is plotted for the four conditions. Channels PO7, Pz and T7 were selected to represent different brain regions with discriminative power between target vs. non-target (see Fig. 10). The ERP waveforms obtained for VO, VC and AU unimodal conditions are very different. The VO condition evokes the most pronounced P300 and N200 ERPs and the AU condition the least pronounced. The HVA waveform is similar to the VC waveform but with a higher amplitude. This suggests that the HVA evoked responses are mostly driven by the visual stimuli. This is emphasized in Fig. 9A, comparing the grand average of the P300 epochs for the four conditions in Pz channel. Most interestingly, the HVA waveform seems to correspond to the addition of the VC and AU waveforms, which is confirmed in Fig. 9B where we sum the P300 responses of the two independent tasks (Visual Covert + Auditory).

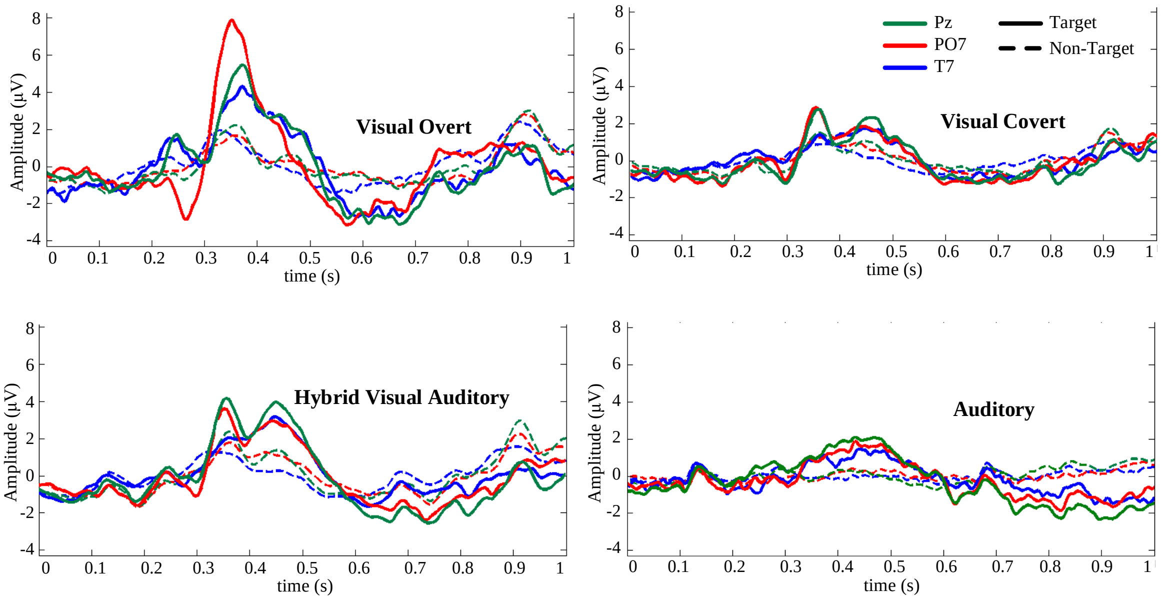
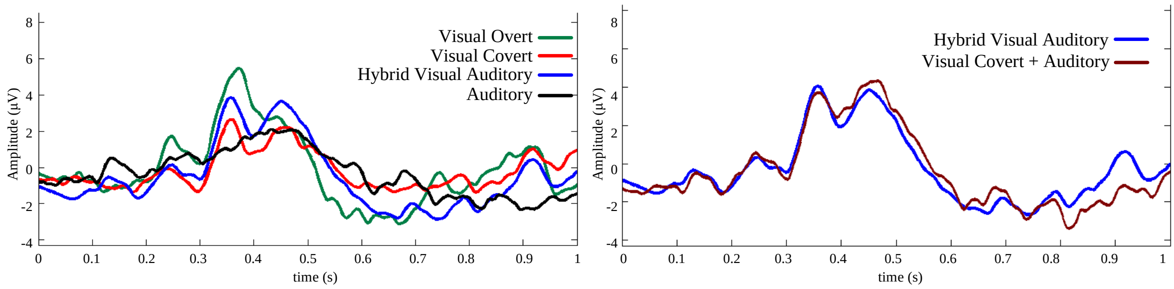


Figure 8. Grand average of target and non-target EEG signals for the channels Pz, PO7 and T7 for the four conditions (obtained from the 10 participants).



A **B**

Figure 9. P300 Target EEG signals for Pz channel for: **(A)** the four conditions (VO, VC, HVA and AU); and **(B)** the comparison between the HVA condition and the additive sum of the visual covert and auditory conditions (obtained from the 10 participants).

The waveforms in specific intervals were further analyzed. For the VC condition, within the interval 300 to 400 ms, target and non-target waveforms are different in Pz and PO7 channels, but similar in T7. For the interval 400 – 500 ms, T7 channel is the one with the highest waveform difference between target and non-target signals. For the AU condition, a higher difference between the target and non-target amplitudes occurs in the interval from 400 to 500 ms and from 800 ms to 1 s. The highest amplitude difference between target and non-target is observed in channel Pz. This behaviour is in accordance with the R-square values obtained in Fig. 10. This suggests that different perceptual/cognitive processes are triggered in visual and auditory tasks.

3.4. Discriminative ERP components

In order to evaluate the most discriminative ERP latencies, we analyzed the R-square values between the target and non-target epochs of the 10 participants, for the 16 channels. The R-square levels usually correlate with the classification rates (Pires et al., 2011b). Figure 10 shows the R-square color maps for the four conditions. The discriminative components match the P300 location, despite being in different latencies. The P300 latency for the VO condition is located around the 300 – 400 ms interval, whereas for the remaining three conditions, the P300 has latencies around 400 – 500 ms. For the auditory stimulation we found other discriminative ERP components around 800 ms. Nevertheless, the auditory P300 component had R-square values slightly higher than those observed in the VC condition, which is in agreement with the online results. VC, AU and HVA conditions have a wider discrimination interval than VO condition. In HVA, we observe an overlap of the discrimination values for the VC and AU conditions, suggesting that participants had an effective simultaneous attention to the visual covert and auditory stimulation.

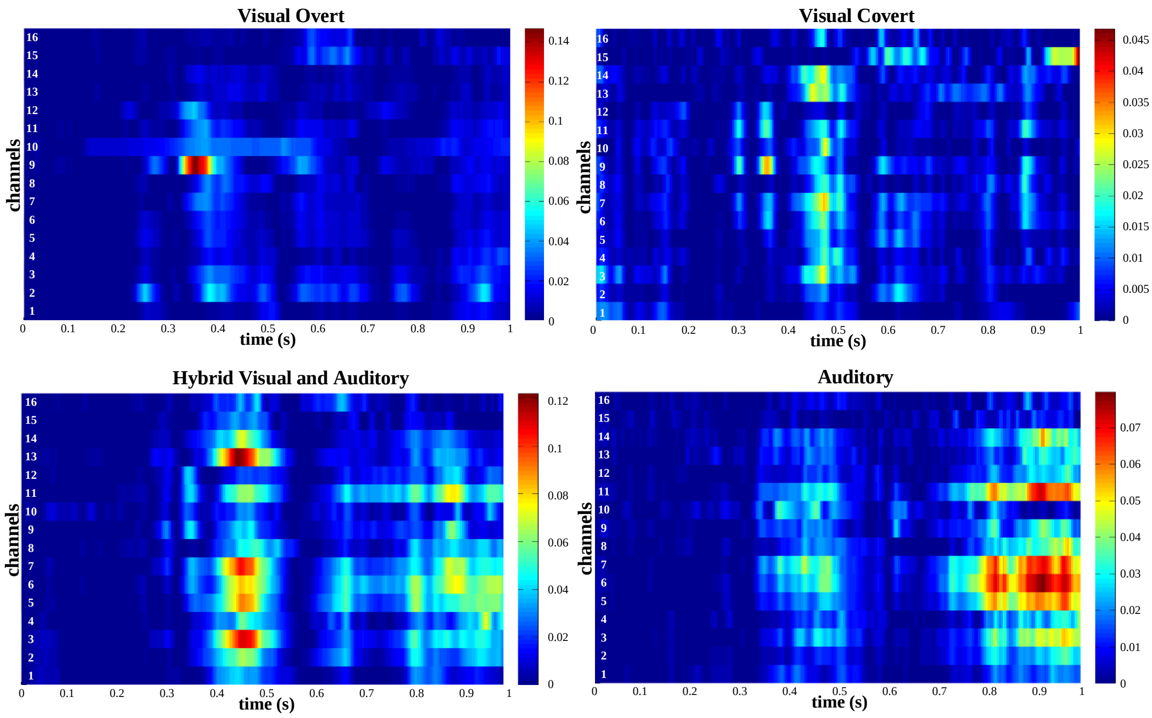


Figure 10. R-square between Target and Non-Target taking the 10 participants for the four conditions. (Channel numbers correspond to: 1 – Fz, 2 – Cz, 3 – C3, 4 – C4, 5 – CPz, 6 – Pz, 7 – P3, 8 – P4, 9 – PO7, 10 – PO8, 11 – POz, 12 – Oz, 13 – T7, 14 – CP5, 15 – T8 and 16 – CP6)

Topographical maps in Figure 11 were analyzed specifically for the two intervals identified as the ones with the most discriminative P300 latencies for the VO condition (300 – 400 ms) and for the remaining three conditions (400 – 500 ms). The purpose of this analysis is to evaluate if there exists spatial correlation between the hybrid condition and the VC and AU conditions. The topographical maps for the VO condition present a higher discriminability between target and non-target events in the 300 – 400 ms interval, which is supported by the already presented R-square results. For the remaining three conditions, significant differences are found in the 400 – 500 ms interval. For the interval 300 – 400 ms, conditions VC, AU and HVA presented the most discriminative features in the frontal lobes. For the interval 400 – 500 ms, the most discriminative features for the HVA condition are located in the left temporal and parietal lobes. Conditions VC and AU also follow this trend, however with lower discriminative levels. The topographical maps of the HVA condition appear to correspond to a superposition of the topographical maps of the VC and AU conditions, for both intervals 300 – 400 ms and 400 – 500 ms. A brain left lateralization is visible, which can be explained because the left side of the brain controls analytic thought, logic, verbal and rational thought, being these some of the abilities required to perceive visual written and spoken words (Seghier and Price, 2011; Tervaniemi and Hugdahl, 2003).

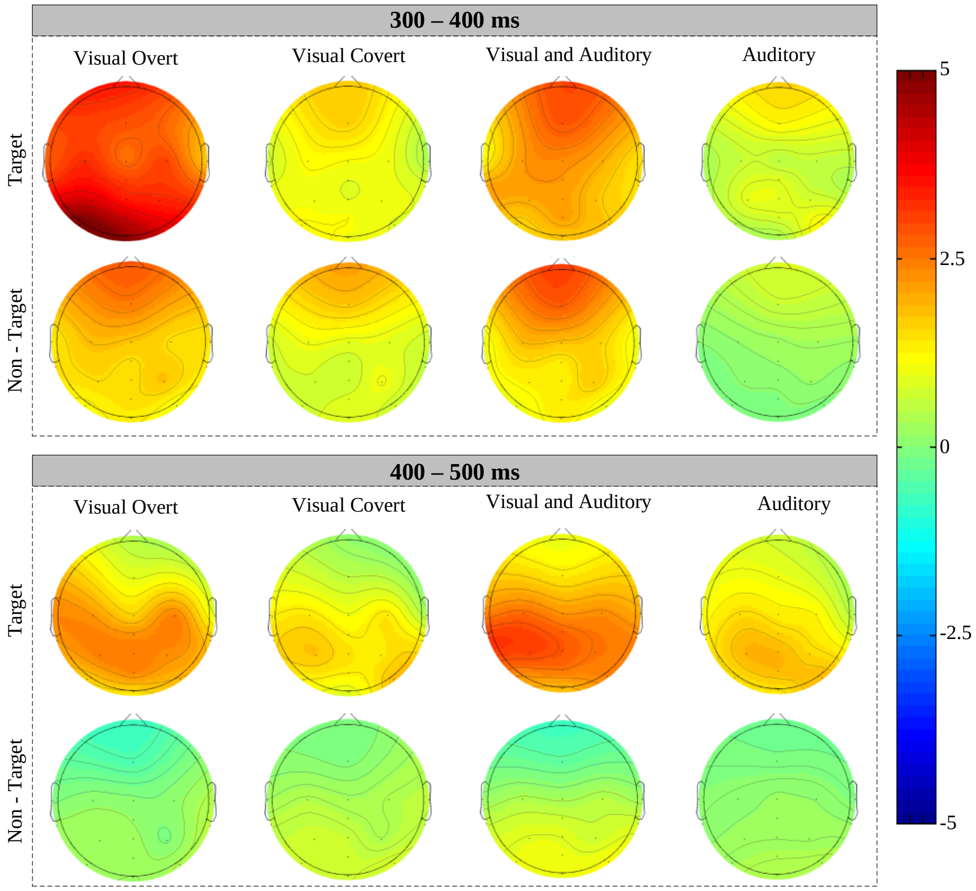


Figure 11. Topographical maps for the Target and Non-Target grand average for the four conditions, for two time intervals: 300 – 400 ms and 400 – 500 ms (obtained from the 10 participants).

3.5. Task load results

The 10 participants were asked to answer two questionnaires at different moments in order to assess several task load parameters such as mental, physical and temporal demand, performance, effort and frustration and pleasantness of the four conditions. Quest I, based on the NASA-TLX, was answered at the end of each online condition, and Quest II after completing the four conditions. Participants had to grade task load parameters from 0 to 20, where 0 is very low and 20 very high and performance parameter where 0 is perfect and 20 is failure. In Quest II, participants were asked to compare the four conditions regarding pleasantness, where 0 and 20 is highly pleasant and unpleasant, respectively. Figure 12 shows the average results of the two questionnaires for the four conditions.

Participants classified condition AU as the one more mental and general effort demanding task and the most frustrating one. Some of the participants reported that during the auditory condition, they had difficulties in identifying the target since they expected for a phonetic combination due a kind of “hearing memory” acquired during the random sequence of the words. For this condition, the global task load result was 67.8 preceded by the VC, with a global result of 70.0. The VO and HVA conditions had global results of 42.6 and 58.5, respectively. From the three gaze-independent approaches, the hybrid HVA presents the lower task load. Also, the participants graded the four conditions in order of pleasantness as VO, followed by HVA, AU and VC.

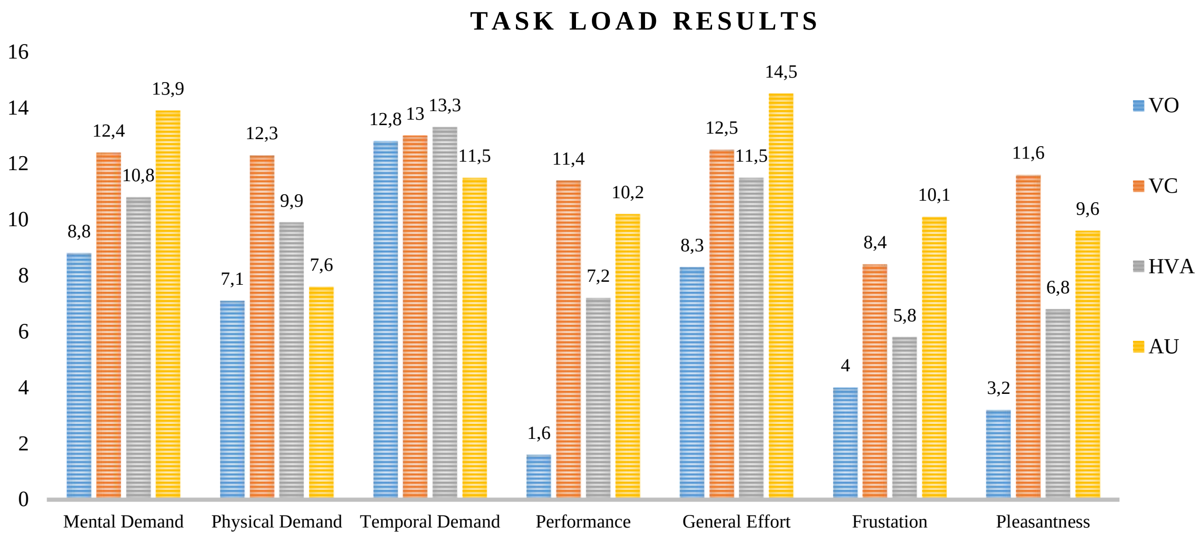


Figure 12. Task load results for interfaces VO, VC, HVA and AU.

4. Discussion

In this paper, we presented a completely gaze-independent hybrid approach that combines visual stimuli with spoken auditory stimuli. We compared four P300-based conditions, namely Visual Overt, Visual Covert, Auditory and Hybrid (Visual Covert and Auditory). The first served as a control condition to validate the classification models and also to confirm BCI literacy of users. As expected this gaze-dependent interface had the best results and there was a ceiling effect on six participants, with online accuracies of 100%. The visual covert and the auditory interfaces had online classification results lower than 60%, and were reported by participants as demanding tasks. The HVA condition presented classification improvements of more than 32% over VC and AU conditions. Eight of the 10 participants achieved more than 80% accuracy.

Participants' feedback regarding the auditory stimuli referred that some words sound awkward and difficult to perceive. This may be due to the limited duration and phonetic similarities between words. The stimuli durations were chosen to not increase too much the trial time and in particular the time of the visual stimuli, which is known to have larger P300 amplitudes for shorter stimuli durations (Allison and Pineda, 2006). Participants P1 and P5, who had previous experience with the four conditions were the only ones who overpassed 75% online accuracy in both VC and AU conditions. The mean accuracy for those participants was 85.4% for the VC condition and 96.6% for the AU condition. This suggests that experience may help users to improve their selective attention abilities. Therefore, familiarization with the auditory cues may be an important key to further improve users' performance. For 7 participants, the individual online accuracies for the AU condition were higher than the ones for the VC condition, revealing that even though most of the participants were not familiarized with the AU condition, it was easier for them to track auditory stimuli than covert visual cues. The low classification accuracies obtained with the VC and AU make us conclude that they are not currently viable alternatives for BCI control. However, for the HVA condition there was an improvement of over 32%, clearly indicating the effectiveness of this bimodal approach, and its consideration as a viable alternative for CLIS patients. Moreover, HVA BCI was reported as a less demanding and more pleasant interface.

The R-square analysis of the AU condition revealed that besides the P300 discriminative components, a component starting around 800 ms had also discriminability. We hypothesize that this can be related to the P600 ERP, associated with sentence processing in auditory listening experiments, particularly when words are difficult to understand (Moore et al., 2009). However, to confirm this hypothesis, further experiments would be required. The temporal range of discriminative R-square values for the HVA condition span over VC and AU conditions, and seem to be additive. This is reinforced by the waveform of the P300 grand average of the HVA task which coincides with the sum of the P300 grand average of the VC task and the P300 grand-average of AU task (independent tasks). The simultaneous attention to the visual covert cues and the auditory stimuli has clear neurophysiologic evidences that it increases users' performance.

The topographical maps obtained from the data of the 10 participants showed a left lateralization tendency, revealing that the participants enrolled in the experiments used the left side of the brain to increase the analytical thought, one of the abilities that we think as being greatly important in the visual and auditory target identification (Seghier and Price, 2011; Tervaniemi and Hugdahl, 2003). The gaze-dependent condition revealed the most discriminative features in the parietal and occipital left lobes, which is in accordance with the findings in (Ikegami et al., 2012). The hybrid condition presented the most discriminative features in the parietal and temporal left lobes (auditory cortex), revealing that the participants had an effective attention to the visual and also to the auditory cues (Furdea et al., 2009; Purves et al., 2004).

The results from the questionnaires indicate higher demand in the perception of auditory stimuli, which justifies higher latency, lower amplitude and consequently lower accuracy when comparing to the gaze-dependent condition. The pilot tests previously made before starting systematic experiments indicated that people preferred meaningful words than simple tones, which motivated our choice. However, results show that spoken words stimuli need further improvements. Several parameters can be adjusted such as duration and intelligibility of the words and their spatial location. The HVA condition was reported as more pleasant than VC and AU conditions, which correlates with the less demanding task. Regarding the task load index of the three gaze-independent conditions, HVA had the lowest value.

Overall, the HVA condition was the gaze-independent solution that lead not only to the highest online performance but was also the one with the highest acceptance between the participants, strongly validating the presented hybrid interface as the one with the highest chances to be well accepted by complete locked-in state patients. Although the reduced lexicon provided by the proposed BCI is quite limited when compared with a BCI-speller, it may represent one way to achieve a feasible BCI adapted to the skills and needs of CLIS patients.

5. Conclusion

The P300-based HVA–BCI here proposed was validated as highly feasible to provide a gaze-independent solution for complete locked-in state patients. The hybrid condition (HVA) was the gaze-independent solution with the highest online performance, and exhibited a P300 signal corresponding to a sum of the P300 of the VC condition and the P300 of the AU condition, which evidences a cumulative effect of the bimodal stimuli. The online accuracy, ITR and effective SPM achieved with this solution compares favourably with the state-of-the-art gaze-independent performances. Moreover, the hybrid condition was classified by the participants as the most pleasant gaze-independent solution.

Further analysis of different properties of visual and auditory stimuli can be considered to find a higher discrimination between targets and non-targets stimuli.

Acknowledgments
This work has been supported by the Portuguese Foundation for Science and Technology (FCT) under the project AMS-HMI12 – Assisted Mobility Supported by Shared-Control and Advanced Human-Machine Interfaces: RECI/EEI-AUT/0181/2012 and by the QREN under the project CENTRO-07-ST24-FEDER-002028: Diagnosis and Assisted Mobility for People with Special Needs.

References

Allison BZ, Pineda JA. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system. Int J Psychophysiol. 2006;59(2):127–40.

An X, Höhne J, Ming D, Blankertz B. Exploring combinations of auditory and visual stimuli for gaze-independent brain-computer interfaces. PLoS One. 2014;9(10):e111070.

Bauer G, Gerstenbrand F, Rumpl E. Varieties of the locked-in syndrome. J Neurol. 1979;221(2):77–91.

Birbaumer N. Brain-computer-interface research: Coming of age. Clin Neurophysiol. 2006;117(3):479–83.

Birbaumer N, Piccione F, Silvoni S, Wildgruber M. Ideomotor silence: The case of complete paralysis and brain-computer interfaces (BCI). Psychol Res. 2012;76(2):183–91.

Combaz A, Chatelle C, Robben A, Vanhoof G, Goeleven A, Thijs V, et al. A comparison of two spelling Brain-Computer Interfaces based on visual P3 and SSVEP in Locked-In Syndrome. PLoS One. 2013;8(9):e73691.

Cui G, Zhao Q, Cao J, Cichocki A. Hybrid-BCI: Classification of auditory and visual related potentials. Soft Comput. Intell. Syst. (SCIS), 2014 Jt. 7th Int. Conf. Adv. Intell. Syst. (ISIS), 15th Int. Symp. 2014. p. 297–300.

Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 1988;70(6):510–23.

Fuchino Y, Nagao M, Katura T, Bando M, Naito M, Maki A, et al. High cognitive function of an ALS patient in the totally locked-in state. Neurosci Lett. 2008;435(2):85–9.

Furdea A, Halder S, Krusienski DJ, Bross D, Nijboer F, Birbaumer N, et al. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology. 2009;46(3):617–25.

Guo J, Hong B, Guo F, Gao X, Gao S. An auditory BCI using voluntary mental response. 2009 4th Int IEEE/EMBS Conf Neural Eng. 2009. p. 455–8.

Halder S, Hammer EM, Kleih SC, Bogdan M, Rosenstiel W, Birbaumer N, et al. Prediction of auditory and visual p300 brain-computer interface aptitude. PLoS One. 2013;8(2):e53513.

Halder S, Rea M, Andreoni R, Nijboer F, Hammer EM, Kleih SC, et al. An auditory oddball brain-computer interface for binary choices. Clin Neurophysiol. 2010;121(4):516–23.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Adv Psychol. 1988;52(C):139–83.

Hill J, Lal TN, Bierig K, Birbaumer N, Schölkopf B. An auditory paradigm for Brain--Computer Interfaces. Adv Neural Inf Process Syst 17. 2004;17:569–76.

Hill NJ, Ricci E, Haider S, McCane LM, Heckman S, Wolpaw JR, et al. A practical, intuitive brain-computer interface for communicating "yes" or "no" by listening. J Neural Eng. 2014;11(3):035003.

Hill NJ, Schölkopf B. An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. J Neural Eng. 2012;9(2):026011.

Höhne J, Krenzlin K, Dähne S, Tangermann M. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. J Neural Eng. 2012;9(4):045003.

Höhne J, Schreuder M, Blankertz B, Tangermann M. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Front Neurosci. 2011;5(August):99.

Ikegami S, Takano K, Kondo K, Saeki N, Kansaku K. A region-based two-step P300-based brain-computer interface for patients with amyotrophic lateral sclerosis. Clin Neurophysiol. 2014;125(11):2305–12.

Ikegami S, Takano K, Wada M, Saeki N, Kansaku K. Effect of the Green/Blue Flicker Matrix for P300-Based Brain-Computer Interface: An EEG-fMRI Study. Front Neurol. 2012;3(July):113.

Käthner I, Kübler A, Halder S. Comparison of eye tracking , electrooculography and an auditory brain-computer interface for binary communication : a case study with a participant in the locked-in state. J Neuroeng Rehabil. 2015a;12(76):1–11.

Käthner I, Kübler A, Halder S. Rapid P300 brain-computer interface communication with a head-mounted display. Front Neurosci. 2015b;9(207).

Käthner I, Ruf C a., Pasqualotto E, Braun C, Birbaumer N, Halder S. A portable auditory P300 brain-computer interface with directional cues. Clin Neurophysiol. 2013;124(2):327–38.

Käthner I, Wriessnegger SC, Müller-Putz GR, Kübler A, Halder S. Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. Biol Psychol. 2014;102(1):118–29.

Kaufmann T, Holz EM, Kübler A. Comparison of tactile, auditory, and visual modality for brain-computer interface use: A case study with a patient in the locked-in state. Front Neurosci. 2013;7(129):1–12.

Kübler A, Birbaumer N. Brain-computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients? Clin Neurophysiol. 2008;119(11):2658–66.

Kübler A, Furdea A, Halder S, Hammer EM, Nijboer F, Kotchoubey B. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. Ann N Y Acad Sci. 2009;1157:90–100.

Kübler A, Müller-Putz GR, Mattia D. User-centred design in brain-computer interface research and developmen. Ann Phys Rehabil Med. 2015;58(5):312–4.

Kuehlmeyer K, Racine E, Palmour N, Hoster E, Borasio GD, Jox RJ. Diagnostic and ethical challenges in disorders of consciousness and locked-in syndrome: A survey of German neurologists. J Neurol. 2012;259(10):2076–89.

Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, et al. The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? Prog Brain Res. 2005;150(05):495–511.

Lesenfants D, Habbal D, Lugo Z, Lebeau M, Horki P, Amico E, et al. An independent SSVEP-based brain-computer interface in locked-in syndrome. J neural Eng neural Eng. 2014;11(3):035002.

Lopes AC, Pires G, Nunes U. Assisted navigation for a brain-actuated intelligent wheelchair. Rob Auton Syst. 2013;61(3):245–58.

Mak JN, Wolpaw JR. Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects. IEEE Rev Biomed Eng. 2009;2:187–99.

McFarland DJ, Wolpaw JR. Sensorimotor rhythm-based brain-computer interface (BCI): model order selection for autoregressive spectral analysis. J Neural Eng. 2008;5(2):155–62.

Miner LA, McFarland DJ, Wolpaw JR. Answering questions with an electroencephalogram-based brain-computer interface. Arch Phys Med Rehabil. 1998;79(9):1029–33.

Moore B, Tyler L, Marslen-Wilson W. The fractionation of spoken language understanding. Percept Speech From Sound to Mean. 2009. p. 235–40.

Murguialday AR, Hill J, Bensch M, Martens S, Halder S, Nijboer F, et al. Transition from the locked in to the completely locked-in state: A physiological analysis. Clin Neurophysiol. 2011;122(5):925–33.

Nijboer F, Sellers EW, Mellinger J, Jordan M a, Matuz T, Furdea A, et al. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol. 2008;119(8):1909–16.

Pichiorri F, Morone G, Petti M, Toppi J, Pisotta I, Molinari M, et al. Brain-computer interface boosts motor imagery practice during stroke recovery. Ann Neurol. 2015;77(5):851–65.

Pires G, Nunes U, Castelo-Branco M. GIBS block speller: toward a gaze-independent P300-based BCI. Conf Proc IEEE Eng Med Biol Soc. 2011a;2011:6360–4.

Pires G, Nunes U, Castelo-Branco M. Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. J Neurosci Methods. 2011b;195(2):270–81.

Pires G, Nunes U, Castelo-Branco M. Comparison of a row-column speller vs. a novel lateral single-character speller: Assessment of BCI for severe motor disabled patients. Clin Neurophysiol. 2012;123(6):1168–81.

Pokorny C, Klobassa DS, Pichler G, Erlbeck H, Real RGL, Kübler A, et al. The auditory P300-based single-switch brain-computer interface: Paradigm transition from healthy subjects to minimally conscious patients. Artif Intell Med. 2013;59(2):81–90.

Purves D, Augustine GJ, Fitzpatrick D, Hall WC, LaMantia A-S, White LE. The auditory system. Neuroscience. Sunderland, Massachusetts: Sinauers Associates, Inc.; 2004. p. 303–9.

Riccio a, Mattia D, Simione L, Olivetti M, Cincotti F. Eye-gaze independent EEG-based brain-computer interfaces for communication. J Neural Eng. 2012;9(4):045001.

Rutkowski TM, Mori H. Tactile and bone-conduction auditory brain computer interface for vision and hearing impaired users. J Neurosci Methods. 2014;244:45–51.

Schreuder M, Blankertz B, Tangermann M. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLoS One. 2010;5(4):e9813.

Seghier ML, Price CJ. Explaining left lateralization for words in the ventral occipitotemporal cortex. J Neurosci. 2011;31(41):14745–53.

Sellers EW, Donchin E. A P300-based brain-computer interface: Initial tests by ALS patients. Clin Neurophysiol. 2006;117(3):538–48.

Sellers EW, Schalk G, Donchin E. The P300 as a typing tool: tests of brain computer interface with an ALS patient. Psychophysiology. 2003;40(s77).

Sellers EW, Vaughan TM, Wolpaw JR. A brain-computer interface for long-term independent home use. Amyotroph Lateral Scler. 2010;11(5):449–55.

Severens M, Van der Waal M, Farquhar J, Desain P. Comparing tactile and visual gaze-independent brain-computer interfaces in patients with amyotrophic lateral sclerosis and healthy users. Clin Neurophysiol. 2014;125(11):2297–304.

Sorrentino SA, Remmert L. Caring for persons with common health problems. In: Elsevier Health Science, editor. Mosby's Essentials Nurs Assist. Missouri; 2014. p. 436.

Tervaniemi M, Hugdahl K. Lateralization of auditory-cortex functions. Brain Res Rev. 2003;43(3):231–46.

Treder MS, Blankertz B. (C)overt attention and visual speller design in an ERP-based brain-computer interface. Behav Brain Funct. 2010;6:28.

Vianello A, Arcaro G, Palmieri A, Ermani M, Braccioni F, Gallan F, et al. Survival and quality of life after tracheostomy for acute respiratory failure in patients with amyotrophic lateral sclerosis. J Crit Care. 2011;26(3):329.e7–14.

De Vos M, Gandras K, Debener S. Towards a truly mobile auditory brain-computer interface: Exploring the P300 to take away. Int J Psychophysiol. 2014;91(1):46–53.

Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, et al. Brain-computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng. 2000;8(2):164–73.