Journal of Neuroscience Methods 379 (2022) 109661

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Visuo-auditory stimuli with semantic, temporal and spatial congruence for a P300-based BCI: An exploratory test with an ALS patient in a completely locked-in state

Gabriel Pires [a,b,*], Sara Barbosa [a], Urbano J. Nunes [a,c], Edna Gonçalves [d]

[a] Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal
[b] Polytechnic Institute of Tomar, Department of Engineering, Tomar, Portugal
[c] University of Coimbra, Department of Electrical and Computer Engineering, Coimbra, Portugal
[d] University Hospital Center of São João, Porto, Portugal

ARTICLE INFO

Keywords:
Hybrid Visual-Auditory stimuli
Semantic, temporal and spatial congruence
Gaze-independent
P300 event-related potential
Brain-computer interface (BCI)
Head-Related-Transfer-Function
Meaningful natural spoken words
Amyotrophic lateral sclerosis (ALS)
Completely locked-in state (CLIS)

ABSTRACT

Background: Brain-computer interfaces (BCIs) are a promising tool for communication with completely locked-in state (CLIS) patients. Despite the great efforts already made by the BCI research community, the cases of success are still very few, very exploratory, limited in time, and based on simple 'yes/no' paradigms.
New method: A P300-based BCI is proposed comparing two conditions, one corresponding to purely spatial auditory stimuli (AU-S) and the other corresponding to hybrid visual and spatial auditory stimuli (HVA-S). In the HVA-S condition, there is a semantic, temporal, and spatial congruence between visual and auditory stimuli. The stimuli comprise a lexicon of 7 written and spoken words. Spatial sounds are generated through the head-related transfer function. Given the good results obtained with 10 able-bodied participants, we investigated whether a patient entering CLIS could use the proposed BCI.
Results: The able-bodied group achieved 71.3 % and 90.5 % online classification accuracy for the auditory and hybrid BCIs respectively, while the patient achieved 30 % and chance level accuracies, for the same conditions. Notwithstanding, the patient's event-related potentials (ERPs) showed statistical discrimination between target and non-target events in different time windows.
Comparison with existing methods: The results of the control group compare favorably with the state-of-the-art, considering a 7-class BCI controlled visual-covertly and with auditory stimuli. The integration of visual and auditory stimuli has not been tested before with CLIS patients.
Conclusions: The semantic, temporal, and spatial congruence of the stimuli increased the performance of the control group, but not of the CLIS patient, which can be due to impaired attention and cognitive function. The patient's unique ERP patterns make interpretation difficult, requiring further tests/paradigms to decouple patients' responses at different levels (reflexive, perceptual, cognitive). The ERPs discrimination found indicates that a simplification of the proposed approaches may be feasible.

1. Introduction

Brain-Computer Interfaces (BCI) may represent one of the last communication options for patients in the locked-in state (LIS), a condition that results from the progression of amyotrophic lateral sclerosis (ALS) or other neurological disorders (Birbaumer, 2006). LIS is characterized by the paralysis of all voluntary muscles except those responsible for blinks, vertical and horizontal eye movements, and can progress to a completely locked-in state (CLIS), in which there is a complete paralysis including eye movements (Chaudhary et al., 2021; Bauer et al., 1979). While ALS patients in mild stages (Pires et al., 2012), advanced stages and even LIS (Sellers et al., 2010; Kübler and Birbaumer, 2008; Wolpaw et al., 2018) have been shown to control gaze-dependent BCIs with performances close to those of able-bodied, completely LIS patients require gaze-independent interfaces due to their eye impairment. A metanalysis to investigate the effectiveness of BCI for ALS patients tested until June 2013 (Marchetti and Priftis, 2015) showed that there was no clear evidence of the effectiveness of BCI for

* Corresponding author at: Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal.
E-mail address: gpires@isr.uc.pt (G. Pires).

https://doi.org/10.1016/j.jneumeth.2022.109661
Received 23 November 2021; Received in revised form 25 May 2022; Accepted 4 July 2022
Available online 8 July 2022

ALS patients and that BCI was unsuccessful for CLIS patients. The study showed also that most BCI approaches were based on electroencephalogram (EEG) using three neural mechanisms, namely, slow cortical potentials (SCPs) (Birbaumer et al., 1999), sensorimotor rhythms (SMRs) (Wolpaw and McFarland, 2004), P300 event-related potentials (ERPs) (Farwell and Donchin, 1988) and only a single study using steady-state visual evoked potential (SSVEP) (Middendorf et al., 2000). Most interfaces relied on visual stimulation and/or visual feedback. Similar conclusions had also been reached previously in a metanalysis in (Kübler and Birbaumer, 2008), which showed the unsuccessful use of BCI by CLIS patients and no correlation between BCI performance and motor impairment in non-CLIS patients. In recent years a great effort has been made in the design of gaze-independent BCIs based on visual-covert, auditory and tactile modalities, or a combination of these using hybrid/multimodal approaches (see a short survey in (Barbosa et al., 2016)). The hypothesis that CLIS patients may suffer from reduced attention spans (Kübler et al., 2009), lack of voluntary cognitive activity and attention impairment (Murguialday et al., 2011; Kübler and Birbaumer, 2008) has pushed researchers to simplify BCI paradigms (for example, decreasing the number of choices to reduce the required working memory) and to design paradigms that can enhance selective attention. On the other hand, the knowledge that CLIS patients retain the auditory system preserved (Murguialday et al., 2011) but may have affected skin's mechanoreceptors, making unfeasible vibrotactile BCIs, has pointed to auditory-based BCIs as one of the most promising solutions. Unfortunately, until now no successful cases of auditory-based BCIs with CLIS patients have been reported, although traceable auditory ERPs using EEG and electrocorticography (ECoG) have been reported while assessing the impairment of auditory and cognitive functions in participants (Han et al., 2019; Bensch et al., 2014; Murguialday et al., 2011; Pokorny et al., 2013).

More recent studies point to other research directions. A BCI based on functional near-infrared spectroscopy (fNIRS) has been tested with one CLIS patient (Gallegos-Ayala et al., 2014) and 4 CLIS patients (Chaudhary et al., 2017). The oxygenation measured in the frontal area showed discrimination between "yes" and "no" patients' thoughts, reaching accuracies above 70 %. The use of a BCI based on SSVEP was reported in Okahara et al. (2018). One patient in CLIS controlled the SSVEP-BCI covertly attaining an accuracy above 83 %. The authors refer that this achievement greatly depended on the good care provided by caregivers, avoiding drying of the cornea. The SSVEP approach has been almost non-existent in clinical studies and, therefore, this result was unexpected, as it depends solely on the afferent visual pathway. In (Guger et al., 2017) a P300 vibrotactile BCI (wrist stimulation) was successfully used in 2 of 3 CLIS patients, attaining 70 % and 90 % binary accuracy. This result also contradicts the findings reported in Murguialday et al. (2011) indicating the absence of ERPs from vibrotactile stimulation. The study in (Han et al., 2019) with one CLIS patient showed for the first time the successful control of an endogenous EEG-based BCI using motor and calculus imagery (attaining accuracies above 87.5 % with only 5 s). The auditory and cognitive functions of the patient were previously assessed through oddball-auditory stimuli (white noise and high/low tones to evaluate mismatch negativity and P300 components). The cases of success already exist but are still very few, very exploratory and limited in time, and the findings are not consensual (Spüler, 2019), which is a good motivation to continue researching new approaches or replicating experiments to infer the effective feasibility of the proposed approaches. The importance of reporting successful and unsuccessful results, mainly when the number of participants is very low, has been stressed in Lotte et al. (2020), as it serves to accumulate evidence and reach significant general conclusions with statistical relevance.

This paper follows on from a previous study (Barbosa et al., 2016), in which we evaluated a P300-based gaze-independent BCI, using three modalities, namely, visual-covert, auditory, and hybrid visuo-auditory combining visual-covert and auditory stimulation, operated by able-bodied participants. These 3 conditions were also compared to visual-overt control. The proposed hybrid paradigm consisted of a small lexicon of meaningful words presented visual and auditorily with temporal and semantic congruence. The hybrid solution provided the best performance and was considered the least demanding in terms of workload. As a relevant finding, we observed that the waveform evoked by the hybrid paradigm corresponded to an exact additive sum of the ERP of the visual-covert task plus the ERP of the auditory task tested separately. This showed the combined effect of the two afferents pathways contributing to greater discrimination of the target events, leading to an increase in accuracy above 30%, reaching 85.3%. This finding agreed with the hypothesis that audiovisual integration and congruence improve both visual and auditory perception and as an effect of super-additive interaction (Kim et al., 2015; Calvert, 2001). Since the independent visual-covert and auditory tasks did not show to be feasible enough for communication purposes and the hybrid approach was lower than overt visual control, this motivated us to research ways that could further enhance the hybrid paradigm.

The contribution of this paper is threefold: (1) Spatial modulation of auditory stimuli (spoken words) in a 3-dimensional (3D) space, through the implementation of head-related transfer functions (HRTF), which are provided binaurally to the participants; (2) In addition to the temporal and semantic congruence of visual and auditory stimuli presented in Barbosa et al. (2016), we added a spatial congruence between the visual layout and the 3D spatial location of the sound, extending the BCI stimulation to a 3-level congruence: semantic, temporal, and spatial, which is for the best of our knowledge the first time it has been done. We researched the effect of this additional spatial congruence on the stimuli perception and BCI performance in tests with a group of able-bodied participants; and (3) Considering the good results of the proposed approaches with the able-bodied group, an exploratory test was conducted with a CLIS-ALS patient to infer the feasibility of the BCI approach.

2. Spatial auditory BCIs

The type of auditory stimuli may have a considerable impact on BCI performance (Hill et al., 2014; Höhne et al., 2012). Tone sounds are described by some subjects as difficult to differentiate due to their unintuitive identification and unpleasantness (Höhne et al., 2011). Singed or spoken syllables, numbers and words have been considered to overpass the unpleasant task of tracking tone sounds, showing increased classification and overall acceptance (Hill et al., 2014) (Halder et al., 2013).

BCIs based on spatial-auditory stimuli have also been researched. Some authors considered three directions by using the left, right and both ears stimulation, using simple earphones/headphones (Höhne et al., 2012; Höhne and Tangermann, 2014; An et al., 2014). Five directional auditory stimuli (left, center-right, center, center-left and right) were also delivered binaurally with ear/headphones recurring to interaural time difference (ITD) and interaural level difference (ILD) to generate spatial sounds (Käthner et al., 2013; Simon et al., 2014). Other studies have used speakers in different locations, which required more bulky and complex setups (Schreuder et al., 2010; Schreuder et al., 2011). The BCI performances obtained with P300 auditory stimuli are in general much lower than those obtained with visual P300-based BCIs controlled overtly. For example, all the above approaches using headphones to deliver directional auditory cues did not overpass the 70 % accuracy, which is low considering that the experiments were performed with able-bodied participants.

Considering 3D audio stimuli via head-related transfer function (HRTF), simpler and quickly adaptable binaural systems can be developed (Sundareswaran et al., 2003, Gardner, 1999; Frauenberger and Noisternig, 2003; Pan et al., 2006). This approach can lead to more flexible setups compared to the spatially distributed speakers, allowing the easy evaluation of different stimuli directions. HRTF is a transfer function that characterizes the sound received by the subject's ear from a specific point in the space. By using a pair of HRTFs, one for each ear, it is possible to produce binaural directional sounds (Wenzel et al., 1993). The auditory stimulus is created by headphones, but the user perceives it as a sound coming from an arbitrary point in the space. When listening to the directional stimulus, the location cues for each ear are reproduced and the listener should perceive the sound source location specified by the HRTFs. If the listener's own HRTFs are used to synthesize the location cues, the binaural synthesizes will work extremely well (Gardner, 1999). HRTFs have been used in virtual audio reality (VAR) display, to generate spatialized sounds and convey them to a listener. This system allows the exploitation of a virtual environment using only the sense of hearing. This technology has been used, for example, in interfaces for visually impaired and blind users (Frauenberger and Noisternig, 2003) and virtual-reality games (Gardner, 1999; Pan et al., 2006). The potential of the HRTF for BCI applications emerges from the fact that by being able to choose any location of the sound, several different setups can be easily and quickly evaluated to develop interfaces that can be designed specifically for each subject. Also, the setup (use of simple earphones/headphones to deliver spatial auditory cues) is much more flexible and pleasant than the one possible with the use of several speakers. Table 1 summarizes the offline and online results regarding the performance of BCIs based on spatial auditory stimulation.

3. Multisensory integration and congruence

Several functional neuroimaging studies have demonstrated brain multisensory integration and the positive effects of temporal, spatial and semantic congruence of audio-visual stimuli (Calvert, 2001; Kim et al., 2015). In our previous study (Barbosa et al., 2016), we showed that multisensory integration increases stimuli perceptual discrimination enhances ERPs and increases BCI classification performance. The same effect has been reported in other BCI studies (Lu et al., 2019; Thurlings et al., 2014; Gondan et al., 2005). The multimodal stimuli can be congruent at different levels. For example, in Barbosa et al. (2016) we tested a semantic congruence between visual and auditory stimuli, whereas in Lu et al. (2019) the BCI was based on stimuli with semantic and spatial (left/right) congruence. Additionally, the effect of spatially congruent vs. incongruent stimuli was researched in a visuo-tactile BCI in Thurlings et al. (2014) and in a visuo-auditory BCI in Gondan et al. (2005). Both studies showed the benefit of spatial congruence in the enhancement of ERPs in addition to the effect of bimodal stimulation. An audio-visual BCI with semantic and spatial congruence was tested successfully as a clinical diagnosis for patients with a disorder of consciousness (DOC) (Wang et al., 2015). The BCI combined drawn numbers presented on the left/right sides of the screen with spoken numbers delivered by speakers placed on the left/right. Despite the above, most experiments with CLIS and DOC patients have been conducted with unimodal auditory stimulation using tone sounds (typically low and high pitch) (Han et al., 2019; Bensch et al., 2014) (Murguialday et al., 2011; Lulé et al., 2013), which motivated us to explore new multimodal congruent approaches.

4. Materials and methods

4.1. *Able-bodied participants*

The same ten able-bodied participants (P1-P10) as in Barbosa et al. (2016) took part in the present study (6 male, 4 female, mean age 26.9 years, SD 5.51 years, range 22–42). None of the participants had hearing impairment, and all had a normal or corrected-to-normal vision. All of them signed informed consent to participate in the study.

4.2. *CLIS patient*

The patient is a male, 55 years of age, ALS functional rating scale-revised (ALSFRS-R) score = 0, entering in a CLIS, resulting from ALS disease with spinal cord onset, diagnosed 95 months before the experiment. The patient had no movement control, except for a very slight and slow horizontal eye movement that he used to communicate with the family (yes/no answers). This ocular movement was starting to vanish, and he couldn't close his eyes spontaneously. The patient used an eye-tracker for some time to communicate but lost the ability to control it about 22 months before this experiment. This was his first contact with BCI. Table 2 summarizes the main clinical characteristics of the patient. The study was approved by the Ethical Committee for Health of the Hospital Center of S. João, Porto, complying with the code of Ethics of the Declaration of Helsinki. Informed consent was signed by a legal representative.

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

Authors	Participants	Delivery system (directions)	Stimuli properties (number of choices)	Performance (repetitions)	
				Offline	Online
(Höhne et al., 2012)	9 able-bodied	Headphones (3)	Tones *(9-choice)*	63.4 % (14)	–
			Spoken syllables *(9-choice)*	64.7 % (14)	–
			Singed syllables *(9-choice)*	65.8 % (14)	–
(Höhne and Tangermann, 2014)	9 able-bodied	Headphones (3)	Spoken alphabet: "A-Z", "-", pause, read and del *(30-choice)*	~73 % (14)	34.7% (dynamic stopping: 5–15)
(An et al., 2014)	15 able-bodied	Headphones (3)	Singed syllables *(6-choice/step, 2-step)*	62.4 % (10)	–
(Käthner et al., 2013)	20 able-bodied	Headphones (5)	Tones and different ISIs *(25-choice)*	–	560 ms - 66 % (10) 400 ms - 65 % (10) 320 ms - 51 % (10) 240 ms - 42 % (10) 160 ms - 17 % (10)
(Simon et al., 2014)	11 able-bodied	Headphones (5)	Animal Tones *(25-choice)*	90.2 % (10)	69.6% (10)
(Schreuder et al., 2010)	5 able-bodied	5 Speakers (5)	Tones *(5-choice)*	74.1 % (15)	–
(Schreuder et al., 2011)	14 able-bodied	6 Speakers (6)	Tones *(6-choice/step, 2-step)*	–	86.1 % (dynamic stopping: 4–15)

Table 2
Summary of the clinical characteristics of the patient.

Characteristic	Patient
Age	55
Gender	Male
ALSFRS-R	0
Diagnosis / onset	ALS / Spinal
Time since onset	95 months
Ventilation / nutrition	Tracheotomy / artificial
State	LIS to CLIS
Movement control	Residual horizontal eye movement (vanishing)
Hearing	Yes
Tactile sensitivity	Yes
Understands what is asked	Yes
Attention spans	< 1 h (unsure)
Medication	Antidepressant and anxiolytic drugs

4.3. Data acquisition

Two 16-channel gUSBamp acquisition devices (g.tec medical engineering GmbH, Austria) were used to acquire EEG and EOG signals, sampled at 256 Hz, following the same procedure as in Barbosa et al. (2016). EEG was acquired with 16 active Ag/AgCl electrodes in a monopolar configuration (Fz, Cz, C3, C4, CPz, Pz, P3, P4, PO7, PO8, POz, Oz, T7, CP5, T8 and CP6) placed according to the extended international 10–20 System. The right or left earlobe was selected for reference and the AFz channel for ground. Vertical and horizontal EOG signals were captured with disposable pre-gelled Ag/AgCl electrodes in a bipolar configuration. Signals were filtered by a band-pass filter between 0.5 and 30 Hz and a notch filter at 50 Hz to eliminate the powerline interference. Data were acquired, processed, and classified in real-time in a High-speed Simulink™ framework.

4.4. Experimental procedure

Participants tested the BCI with two different conditions, one corresponding to purely spatial auditory stimuli delivered binaurally (AU-S) and the other corresponding to hybrid visual and auditory spatial stimuli (HVA-S). In the HVA-S condition, there is a semantic, temporal, and spatial congruence between visual and auditory stimuli. The BCI consisted of a P300 oddball paradigm (1 Target and 6 standards (non-Targets)), comprising seven stimuli that intended to form a small communication lexicon, namely, ‘SIM’ (Yes), ‘NÃO’ (No), ‘FOME’ (Hunger), ‘SEDE’ (Thirst), ‘AR’ (Air), ‘POSIÇÃO’ (Position) and ‘URINAR’ (Urinate), as shown in Fig. 2. All stimuli had a duration of 550 ms (100 ms longer than in (Barbosa et al., 2016)), and an interstimulus interval (ISI) of 100 ms (stimuli onset asynchrony = 650 ms). The duration of visual stimuli was selected to match the same duration of auditory stimuli (explained in Section 4.5.1.). The ISI was kept the same as in (Barbosa et al., 2016) as it showed to match the desired effect. The able-bodied participated in Experiment I and the CLIS patient participated in Experiment II, which occurred at different moments. Experiment I aimed to validate the methods and approach proposed for the BCI, while Experiment II aimed to infer its usability for a CLIS patient.

4.4.1. Experiment I

Each participant performed 2 sessions (the first one with HVA-S and the second one with AU-S, the order of the tests was the same for all participants). In each session, participants first underwent a calibration to obtain data to train the classifier, and then they controlled the BCI online. In the HVA-S condition, participants were comfortably seated in front of a 22-inch computer screen at approximately 50 cm. The screen height was adjusted so that the participants’ eyes were on the same level as the center of the screen and ensured that they could perceive the stimuli covertly (without eye-gazing). The auditory stimuli were delivered through earphones. In the AU-S condition, the only difference was that the screen was removed. Participants were asked to mentally count the target stimuli and ignore all other standard stimuli. Before calibration, the task was carefully explained, and participants had the opportunity to test and clarify any doubts about the paradigm. Participants were also asked to indicate the location of spatial sounds to verify that they matched the expected region.

Before the online operation, participants performed a P300 calibration which consisted of 14 trials each one with 10 sequences/repetitions (about 13 min long), collecting 140 target epochs and 840 standard epochs to train the classifier. During the calibration, participants were informed of which target to attend before each trial. Visual and auditory cues of the target stimulus were provided in the HVA-S condition while an auditory cue of the target stimulus was provided in the AU-S condition. The auditory cue was the sound of the word of the respective stimulus. The ocular movement detector (OMD) described in Barbosa et al. (2016) was used in the HVA-S condition to reject the EEG epochs contaminated with ocular movements. This required an additional 185 s calibration. In the online task, each trial consisted of 6 sequences. The participants had to complete a set of 15 trials (selection of 15 words). The set of words was pre-defined and was the same for all the participants for both conditions. To allow participants to rest and perform ocular movements between trials, an inter-trial interval (ITI) of 8 s was considered.

For the online BCI session, participants had to select 15 words of a predefined sequence (copy task) whose sequence was different from that used in calibration, but the same for both conditions. For the HVA-S condition, the words of the sequence were provided visually at the top of the screen, while in the AU-S condition the operator orally delivered the auditory words. The purpose was to have neutral cues that could not bias the result, i.e., similar to a free task (without cues). At the end of each HVA-S trial, the selected word was visually provided, but no auditory feedback was presented to the participant at the end of each trial for the AU-S condition, only after completing the 15 selections. This option was taken mainly in order to not mix up the auditory cues, the auditory feedback and the auditory stimuli, avoiding confusing the participant during the experiment. The time between selections was 27.3 s ($6 \times 7 \times 0.650$ s) plus the ITI of 8 s. This experiment took about 1.5 h to complete, including setup time and a period for contact and familiarization with the stimuli and interface arrangement. At the end of each condition test, participants were asked to answer a simple questionnaire (Quest I) based on one part of the NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988). After completing the two conditions, participants were asked to answer a second questionnaire (Quest II) to compare the pleasantness of the two conditions.

4.4.2. Experiment II

The sessions and procedure were the same as in experiment I, except for the following. The experiment was conducted at the patient’s home. The patient was in a bed in a reclining position, and the screen was placed on an articulated table over the bed and adjusted to ensure covert detection. Instructions and explanations were provided with the help of family members. The patient answered ‘yes/no’ questions with ocular movements. This interaction was difficult and not always consistent as sometimes there was no reaction to questions (evidencing a lack of vigilance and short attention spans), and the ocular movement was very small. During the BCI experiment, the patient also evidenced periods of lack of vigilance. The family and a doctor accompanied the entire course of the experiment. The patient did not perform the OMD calibration, as no traceable EOG was detected during the preparation of the experiments. The patient did not answer any questionnaire.

4.5. BCI paradigm design: Visual and spatial auditory stimuli

The BCI paradigm comprises visual and auditory stimuli. Visual stimuli are provided in Simulink™ while auditory stimuli are provided in Presentation® (Neurobehavioral Systems, version 16.5). Simulink™ and Presentation® are running on the same computer and are synchronized through TCP/IP communication. The EEG epochs are labeled and classified using g.tec Highspeed Simulink™ framework (controlled by gUSBamp acquisition system), which also triggers visual and auditory stimuli. The visual layout is shown in Fig. 2A and was the same as in (Barbosa et al., 2016). The auditory stimuli are different from the ones in Barbosa et al. (2016), as they were recorded differently, encoding spatial location as explained in section 2.5.1.

4.5.1. Auditory stimuli recording

The auditory stimuli match the visual stimuli, consisting of the following Portuguese spoken words, with the phonetic transcription of each speech sound in parentheses: 'SIM' (s'ĩ), 'NÃO' (n'ɐ̃w̃), 'FOME' (f'ɔmə), 'SEDE' (s'edə), 'URINAR' (uɾin'aɾ), 'AR' ('aɾ) and 'POSIÇÃO' (puzis'ɐ̃w̃). The seven spoken words were recorded inside of an ABSLOC.15 soundproof booth from ABSORSOR with the following interior and exterior dimensions: 1140 × 1080 × 2250 mm and 1300 × 1240 × 2500 mm ($W \times L \times H$), respectively. The soundproof booth had a 0.25 s reverberation time and a sound reduction index between 45 and 55 dB. The words from a female voice were recorded and rectified using the Audacity (Audacity Team, version 2.0.5) software. The same stimuli duration of 550 ms was defined for all words. During recordings in the soundproof booth and respective sound analysis with Audacity, it was noticed that the 450 ms used in Barbosa et al. (2016) was too tight for some of the words. Therefore, we opted to increase by 100 ms the stimulus duration to ensure the effectiveness of HRTFs. The recording of stimuli in a soundproof booth was a requirement for the implementation of spatial sounds through HRTF. At the same time, it improved the auditory stimuli in terms of discriminability and sound pleasantness.

4.5.2. Spatial auditory stimuli with head-related transfer functions

The recorded and rectified auditory stimuli were convolved with HRTF, resulting in different spatial locations delivered binaurally via simple earphones or headphones. Head-related transfer functions are usually evaluated by inserting miniature microphones into the ear canals of a human subject or manikin (Gardner, 1999; Carlile, 1996). An audio signal is played by loudspeakers positioned in a precise location and the sound is recorded by the microphones. The recorded left and right signals are processed to derive a pair of HRTFs, which will match the location of the sound source. Different locations will need their pair of HRTFs. The HRTF describes the time delay, amplitude, and tonal transformation for a precise sound source location to the left and right ear of the subject or manikin. Wightman and Kistler (1989) described how the appropriate filters can be generated. The following equations apply only to one ear. For each simulated position, a pair of filters must be constructed. Let $x_1(t)$ represent the electrical signal from a loudspeaker in a free field and $y_1(t)$ the electrical signal from a microphone inserted into the manikin or subject ear canal. In addition, let $x_2(t)$ represent the electrical signal from a headphone and $y_2(t)$ the resultant microphone response. The HRTF goal is to produce $x_2(t)$, given $x_1(t)$ such that $y_1(t)$ is equal to $y_2(t)$. In the frequency domain, let X_1, X_2, Y_1 and Y_2 be the Fourier transforms of $x_1(t)$, $x_2(t)$, $y_1(t)$ and $y_2(t)$, respectively. If we consider L as the loudspeaker transfer function, F as the free-field-to-eardrum transfer function and M as the microphone transfer function, the microphone's response to $x_1(t)$ can be written:

$$Y_1 = X_1 LFM \tag{1}$$

The microphone's response to $x_2(t)$, if we consider H as the headphone-to-eardrum transfer function, can be written as:

$$Y_2 = X_2 HM \tag{2}$$

Setting $Y_1 = Y_2$, we have the following:

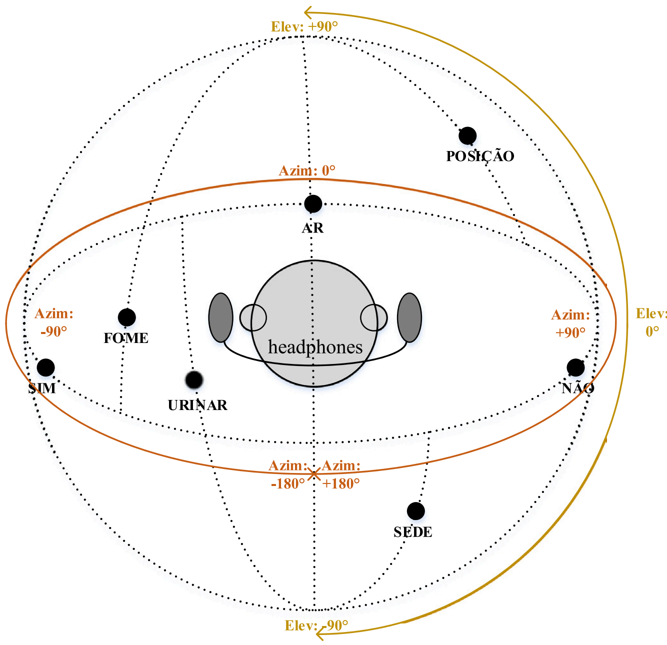
$$X_1 LFM = X_2 HM \tag{3}$$

Solving for X_2 yields to:

$$X_2 = X_1 LF/H \tag{4}$$

Therefore, LF/H is the desired filter transfer function. Passing $x_1(t)$ through this filter results in $x_2(t)$ transduced by the headphones, the audio signal received by the subject at the eardrum is the same signal produced by the loudspeaker in the free field (represented by substituting X_2 in Eq. 2 by the right side of Eq. 4). As mentioned above, the equations apply only to one ear and to a single loudspeaker position. A pair of filters (one for each ear) for each desired sound source location must be designed to synthesize each stimulus. Several aspects influence the HRTF such as the fold and the asymmetry of the pinnae, so ideally, the HRTF for each participant should be evaluated to guarantee the best approximation for each participant. However, since the measurement of HRTFs is a complicated procedure, 3D audio systems usually use HRTFs previously evaluated from a particular human or a manikin. We used the HRTFs pairs available in the database from (Gardner and Martin, 1995). The evaluated measurements consisted of the left and right impulse responses from loudspeakers mounted 1.4 m from a KEMAR dummy-head. Pseudo-random binary sequences were used to obtain impulse responses at a sampling rate of 44.1 kHz. As the HRTFs are not fully transferable between individuals, it was expected that participants could perceive sound localization slightly differently among themselves, but still be able to localize sound coming from the same spatial region. In several VAR studies it is shown that although individual HRTFs provide more accurate spatial localization, their impact compared to generic HRTFs may be minimal (Rummukainen et al., 2021). Although sound localization could be perceived differently for each participant, 3D spatial localization was consistent across BCI sessions for each participant as the earphones were always placed in the same position in the ears. This was verified during pilot experiments and pre-session periods (training and task description), by asking participants of the control group to indicate the location of the sounds, confirming that they could locate sounds as expected. Though, this verification was impossible to do with the CLIS patient due to the lack of communication.

The spatial location of the seven words was selected to avoid the cone of confusion effect (Begault, 2000). Different sound stimuli within the cone of confusion present similar interaural time differences (ITD) and interaural intensity differences (IID). In the absence of spatial cues rather than ITD and IID, if two sound sources are located within the same cone of confusion, the discrimination between the two sources might be confusing or even impossible. To overcome that, we selected elevations and azimuths that would guarantee that the seven different sounds would correspond to discriminable ITD and IID. Fig. 1 shows the selected elevation and azimuth for each one of the seven spoken words and their representation on a 3D view model. Fig. 2B shows the auditory stimuli arrangement in terms of azimuth. The azimuths for the auditory cues were selected to be spatially congruent with the visual stimuli (Fig. 2A). The rationale was to match the visual stimuli at the upper part of the visual layout with sounds located in front of the participant, match visual stimuli at the horizontal line with sounds located on the sides of the participant, and match visual stimuli at the lower part of the visual layout with sounds located behind the participant. For example, 'AR' is located in front of the participant's head. 'NÃO' and 'SIM' corresponded to the right and left sides of the participant's head, respectively, matching the position of the visual stimuli. A small source location shift was introduced for 'FOME', 'SEDE', 'URINAR' and 'POSIÇÃO', to prevent sounds from the back and front of the head from having symmetric azimuths. The words 'URINAR', 'FOME', 'POSIÇÃO' and 'SEDE' had elevation values different than zero, this was introduced to make the sound source localization even more discriminable. Therefore, there is no strict 2D spatial congruence between the visual layout and sound localization, as the visual layout is firstly mapped into the azimuth plane. However, this spatial association was clearly explained to participants who had the opportunity to test the hybrid stimuli. The recorded sounds are available as supplementary material.



Code	Stimuli	Elevation (°)	Azimuth (°)
1	SIM	0	-100
2	NÃO	0	+100
3	FOME	+30	-130
4	SEDE	-40	+148
5	URINAR	-40	-32
6	AR	0	0
7	POSIÇÃO	+30	+60

Fig. 1. Selected elevation and azimuth and spatial location of the auditory stimuli (**Azim** – Azimuth, **Elev** - Elevation).

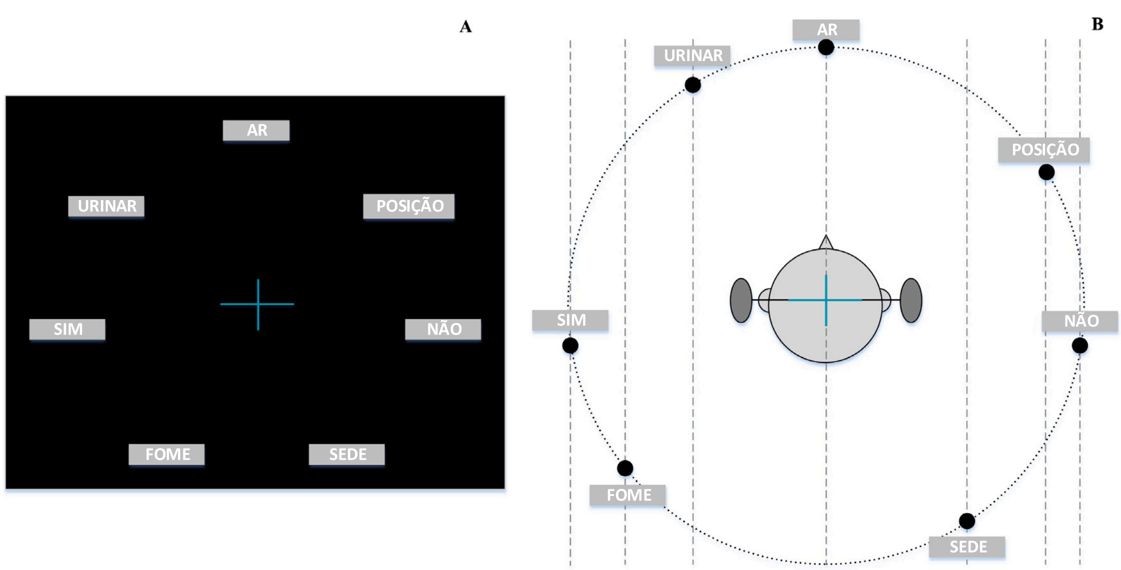


Fig. 2. (A) Spatial location of the visual stimuli and (B) of the auditory stimuli (view above the head). Matching between visual layout and spatial sound location to achieve spatial congruence.

4.6. Classification model

For each event (target and non-target) a 1-second epoch was extracted after the stimulus onset. For each trial, the average of the 6 repetitions was obtained for the 16 EEG recorded channels. A statistical spatial filter called C-FMS that combines sub-optimally two discrimination criteria was applied to the 16 channels (Pires et al., 2011). The two most discriminative projections of the spatial filter are concatenated into a single vector, and the best 200 features are selected through an r-square measure and then classified by a Bayes classifier. Both statistical filter and Bayes classifier models were obtained from calibration data. The overall classification approach followed the same pipeline as in Barbosa et al. (2016) to have a direct comparison of results.

5. Results

5.1. Experiment I: Online accuracy

Each online session was preceded by a calibration session for that same condition. This was repeated for the two conditions. The number of repetitions per trial was pre-set to 6 so that we could make direct comparisons with the results in Barbosa et al. (2016). Table 3 shows the percentage of correct choices (P_{ac}) and the information transfer rate (ITR) for the conditions with non-spatial (HVA and AU) and spatial (HVA-S and AU-S) stimuli for the able-bodied participants. We observe, for both HVA and AU conditions, higher mean accuracies when considering spatial stimuli. The mean accuracy for HVA-S increased from 85.3 % to 90.5 % (paired *t*-test, $p = 0.1$) and for the AU-S from 52.7 % to 71.3 % (paired *t*-test, $p = 0.016$). Removing participant P8 who

Table 3
Online session classification accuracies and ITR (able-bodied group) for the two conditions (HVA and AU), for non-spatial and spatial sound localization, with 6 repetitions. "Non-spatial" columns were directly obtained from Barbosa et al. (2016).

		HVA				AU			
		Non-spatial		Spatial		Non-spatial		Spatial	
		P_{ac}	ITR	P_{ac}	ITR	P_{ac}	ITR	P_{ac}	ITR
Participants	**P1**	93.3	5.68	100	5.95	93.3	5.68	100	5.95
	P2	86.7	4.72	92.3	4.70	6.7	0.10	66.7	2.18
	P3	78.5	3.74	86.7	4.02	33.3	0.41	40	0.61
	P4	92.3	5.52	92.9	4.77	46.7	1.08	80	3.33
	P5	92.9	5.52	100	5.95	100	6.99	100	5.95
	P6	57.1	1.78	66.7	2.18	33.3	0.41	33.3	0.35
	P7	85.7	4.60	100	5.95	80	3.90	93.3	4.84
	P8	100	6.99	72.7	2.67	46.7	1.08	80	3.33
	P9	80.0	3.90	93.3	4.84	53.3	1.50	80	3.33
	P10	86.7	4.72	100	5.95	33.3	0.41	40	0.61
Mean		**85.3**	4.72	**90.5**	4.70	**52.7**	2.16	**71.3**	3.05
Std		11.8	1.40	11.9	1.38	29.7	2.47	25.3	2.11

appears to be an outlier as his performance does not follow the general trend of the participants, the mean accuracy for the HVA condition increased from 83.7 % to 92.4 % (a statistically significant difference of 8.7 %, paired *t*-test $p < 0.001$). The HVA-S condition remains the auditory-based condition with higher accuracies. Nevertheless, the AU-S led to a mean accuracy higher than 70 %, with all participants achieving an equal or better result with the spatial stimuli. The HVA-S accuracy is 19.2 % higher than AU-S (paired *t*-test $p = 0.011$).

5.2. Experiment I: Amplitude, latency and width of P300 ERP

The amplitude and latency of the P300 waves were analysed at 'Pz' electrode. The amplitude measures the value of the P300 peak and the latency is the instant of the P300 peak. These two measurements were obtained from the P300 epochs evoked by the target events gathered during the calibration sessions of the 10 able-bodied participants, making a total of 140 epochs per condition for each participant. The signal portion from 200 ms to 600 ms was analysed, which corresponds to the range where the P300 occurs. The mean amplitudes of the P300 peaks were respectively 3.5 μV and 6.3 μV for the AU-S and HVA-S conditions, respectively. The difference was statistically significant (paired *t*-test, $t[9] = 4.454$, p-value < 0.001), indicating a stronger P300 for the HVA-S condition. A double peak is observed for the HVA-S condition which reflects the cumulative congruent effect of visual and auditory stimulation. Regarding latency, the P300 peaks occurred at 426 ms and 411 ms, respectively for AU-S and HVA-S, but the difference was not statistically significant. However, considering only the first of the double-peak which corresponds to the visual stimulation, the HVA-S latency decreases to 333 ms, being the difference statistical significant (paired *t*-test, $t[9] = 3.4405$, p-value $= 0.0037$). The higher latency for the AU condition may be due to the increased time required for stimuli perception and processing due to the discrimination difficulty in spoken words rather than the modality (Halder et al., 2013).

The individual averages and grand average of Target and non-Target ERP waveforms recorded at 'Pz' electrode are shown in Fig. 3. To evaluate the discriminative time points, we computed point-wise *t*-tests (two-tailed) comparing Target and non-Target responses across all able-bodied participants for the 16 electrodes recorded, as shown in Fig. 4. Significant differences are shown in a color map for an alpha criterion ≤ 0.01. For the AU-S condition, the most discriminative features fall within the 325–600 ms time window but do not cover the visual cortex region (PO7, PO8, POz, Oz). Regarding HVA-S responses, the discriminative features appear in three time windows, 270–315 ms, 440–550 ms and 600–700 ms. This shows the effect of the P300 visual stimulation (270–315 ms) and the cumulative effect of the P300 visual and auditory stimulation (440–550 ms). The 600–700 ms time window seems to reflect the N400 component related to semantic processing. Overall, the results show a stronger discriminative power for the HVA-S condition which is in line with the online classification results.

5.3. Task load results

The 10 able-bodied participants were asked to answer two questionnaires at different moments to assess several subjective task load

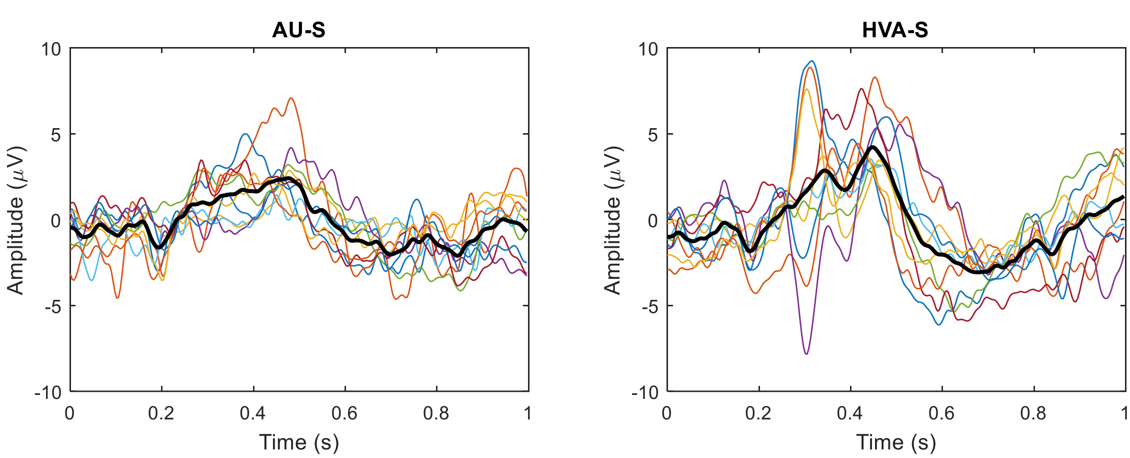


Fig. 3. P300 ERP waveforms recorded at "Pz" electrode. Grand average (bold line) and individual average (thin lines) ERP waveforms for AU-S and HVA-S conditions (able-bodied participants).

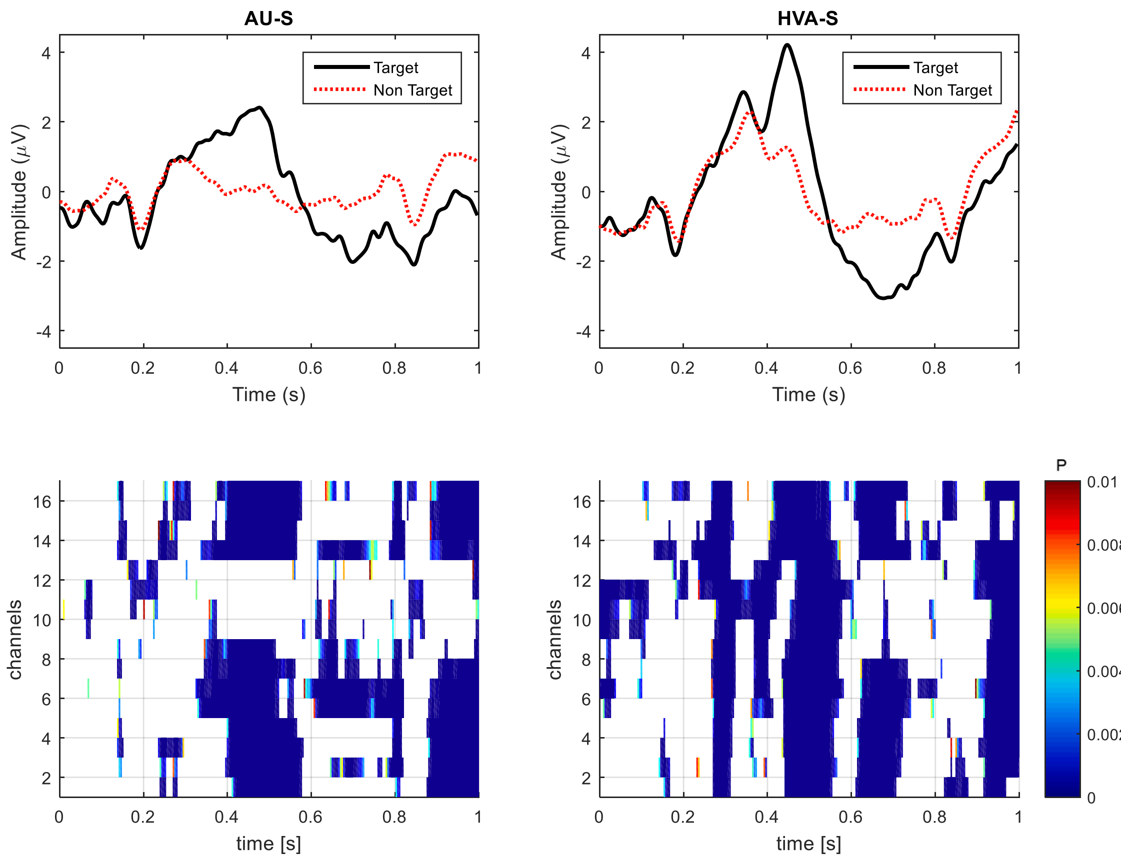


Fig. 4. Comparison of Target and non-Target responses for AU-S and HVA-S conditions (able-bodied participants). Top: Grand average of Target and non-Target ERP waveforms recorded at ‘Pz’ electrode. Bottom: Color map of point-wise *t*-tests comparing target and non-Target responses across all able-bodied participants for the 16 electrodes recorded. Significant differences are shown for an alpha criterion of 0.01 (non-white areas). Electrode 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.

parameters such as mental, physical, and temporal demand, performance, effort and frustration and pleasantness of the two conditions. Quest I, based on the NASA-TLX, was answered at the end of each online condition, and Quest II after completing the two conditions. Participants had to grade task load parameters from 0 to 20, where 0 is very low and 20 very high (except for the performance parameter that was scored on the opposite scale, i.e., 0 is perfect and 20 is failure). In Quest II, participants were asked to compare the two conditions regarding pleasantness, where 20 and 0 are highly pleasant and unpleasant, respectively. Fig. 5 shows the average results for each item and the respective standard deviation for the AU-S and HVA-S conditions and includes the AU and HVA results obtained in Barbosa et al. (2016) for direct comparison.

In Barbosa et al. (2016), participants have 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

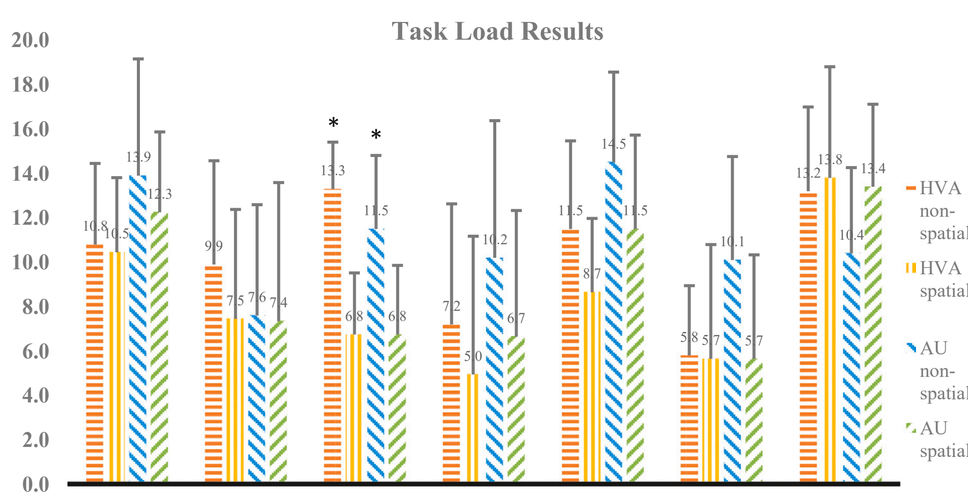


Fig. 5. Task load results comparing spatial and non-spatial HVA-S and AU-S tasks (AU and HVA results report directly to those obtained in Barbosa et al. (2016)). Error bars indicate standard deviation, and ‘* ’ indicate differences between spatial and non-spatial pairs that are statistically significant.

auditory condition, they had difficulties in identifying the target since they expected a phonetic combination due to a kind of 'hearing memory' acquired during the random sequence of the words. The AU and HVA conditions had global (the sum) values of 67.8 and 58.5, respectively. With the introduction of the spatial stimuli (AU-S and HVA-S conditions), we observe that these global values decreased to 50.2 and 43.9, respectively, which corresponds to an increase in subjective appreciation. However, the differences per item, between spatial and non-spatial stimuli, are only statistically significant for 'Temporal demand' ($p < 0.001$). Additionally, the differences between AU-S and HVA-S conditions are lower than between AU and HVA. The AU-S condition has similar results to the HVA-S condition regarding 'Physical' and 'Temporal demand', 'Frustration' and 'Pleasantness'. The only statistically significant difference is for 'Mental demand'. During debriefings, participants P1, P2, P4, P5 and P8 also stated that the AU-S solution was sometimes more pleasant than the HVA-S solution since they did not require focusing on the blue cross.

5.4. Experiment II: CLIS patient results

5.4.1. ERP neurophysiological analysis

The Target and non-Target ERP waveforms recorded at Pz electrode for the CLIS patient are shown in Fig. 6. Both AU-S and HVA-S responses do not evidence a traceable P300 ERP. Notwithstanding, the Target and non-Target ERPs present distinguished patterns in some time windows. To infer the existence of discriminative features we ran the point-wise t-tests (two-tailed) comparing Target and non-Target responses for the 16 electrodes recorded. The significance level was defined for an alpha of 0.05 (for the 0.01 criterion, the significant time windows are almost inexistent). For the AU-S condition, two thin time windows around 50 ms and 200 ms show statistical differences. The 200 ms time window appears mainly on the central cortex and auditory cortex. These most discriminative time windows do not match the ones found in the able-bodied group. These two discriminative areas do not appear in the HVA-S condition. There is a positive peak occurring around 100 ms of the target responses that is not discriminative, while there is a discriminative time window around 500 ms associated with a negative peak. It spans over all electrodes seeming to integrate both visual and auditory information, although the discrimination is stronger in the auditory cortex (T8 and CP6). Notwithstanding this discriminative time window matches the same as the control group, it is difficult to say if it reflects the same neurophysiological process, as here we have a negative deflection while in the control group it corresponds clearly to the P300 component.

5.4.2. Classification results

The offline binary (Target vs non-Target) classification results obtained from the calibration dataset of the CLIS patient are shown in Fig. 7. Results report the balanced accuracy obtained through 10-fold cross-validation using the same classification methodology used for the control group (see Section 4.6.). Results are presented simulating different numbers of epochs (repetitions) per trial. The results are clearly above the chance level, reaching an accuracy of approx. 90 % for 8 repetitions in the AU-S condition. The results are better for the AU-S condition than for the HVA-S condition.

Taking the data-driven classification models using 8 repetitions, we conducted the online BCI experiment first with the AU-S condition and then with the HVA-S condition, each one immediately after the respective calibration. The online classification results were 30 % accuracy with the AU-S condition (above chance level, $1/7 = 0.1429$, considering that BCI has 7 choices) and at the chance level with the HVA-S condition.

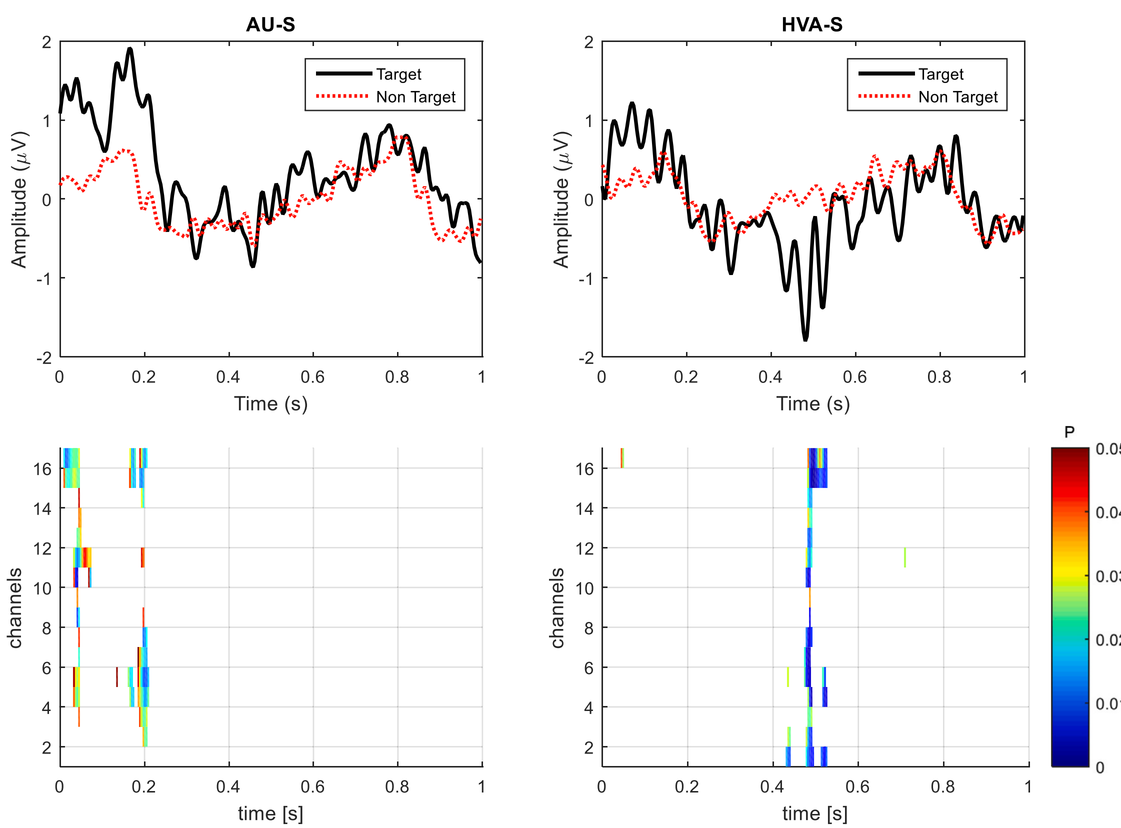


Fig. 6. Comparison of Target and non-Target responses for AU-S and HVA-S conditions (CLIS patient). Top: Grand average of Target and non-Target ERP waveforms recorded at 'Pz' electrode. Bottom: Color map of point-wise *t*-tests comparing target and non-Target responses across all able-bodied participants for the 16 electrodes recorded. Significant differences are shown for an alpha criterion of 0.05 (non-white areas). Electrode 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.

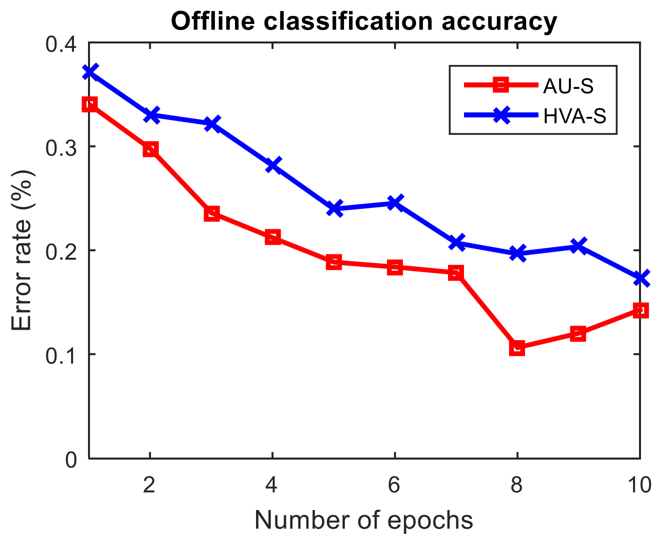


Fig. 7. Offline classification results (Target vs. non-Target) obtained through 10-fold cross-validation from the calibration dataset of the CLIS patient for different numbers of epochs per trial. Results were obtained from 140 Target and 840 non-Target epochs.

6. Discussion

This paper proposes and evaluates a P300-based hybrid BCI that combines visual (covert control) and spatial auditory stimuli where there is a semantic, temporal, and spatial congruence between visual and auditory stimuli. As a follow-up to our previous hybrid-BCI (Barbosa et al., 2016) we wanted to research whether modulating the location of the spatial source of the auditory stimuli could improve the BCI performance of both the purely auditory BCI and the visuo-auditory BCI.

6.1. Experiment I: Able-bodied group

6.1.1. Classification results

A comparison with our previous work shows that the BCI performance of the group of able-bodied participants was improved significantly by about 19 % and 5 %, respectively for the AU-S and HVA-S (visual covert) conditions, attaining 71.3 % and 90.5 % for 6 repetitions. Considering that this is a 7-class BCI, the results compare favorably to the state of the art. The HVA-S accuracy is close to that obtained with visual-overt control in (Barbosa et al., 2016), whose result was 93.3 % and which remains the most effective way of controlling P300-based BCIs. Looking at the standard deviations (SD) of the classification accuracy in Table 3, we observe very high inter-subject variability for the auditory conditions. Yet, the variability is smaller for the spatial auditory stimuli in addition to having improved the average, suggesting the positive discrimination effect of 3D sounds across participants, and the effective use of HRTFs. The results suggest that the responses to auditory stimuli are more user-specific. Regarding the hybrid modality, both HVA and HVA-S present much lower variability than the purely auditory conditions, which indicates that the cumulative effect of multimodal stimulation is effective among participants. The classification variability is consistent with other studies. For example, in Halder et al. (2013), a study with 40 participants, the authors reported a 38% SD of the accuracy for auditory stimuli and 14.7% for visual stimuli, values higher than those obtained in our study. Overall, both hybrid modality and 3D spatial sounds contributed to higher classification accuracy and less inter-subject variability.

Although it is not straightforward to correlate the accuracy increase with the spatial congruence of visual and auditory stimuli (HVA-S vs HVA), as we are not testing congruent vs incongruent conditions (but only congruent vs neutral), it remains clear the positive effect of spatial sounds against the left/right sounds delivered simultaneously, which were used in Barbosa et al. (2016). From a methodological point of view, the use of the HRTF showed to be an effective approach to delivering 3D auditory stimuli simply using earphones or headphones, avoiding complex and bulky setups. Although the same HRTFs were used for all participants, who eventually perceived sound localization slightly differently among themselves, they were still able to localize sound coming from the same spatial region and did not report the occurrence of perceived sound localization shifts during the experiment.

6.1.2. Neurophysiological analysis of P300 ERP

Comparing the AU-S and HVA-S conditions, the neurophysiological analysis showed the cumulative integration effect of visual (covert control) and auditory stimuli, consistent with previous literature. This increased the discriminative time windows between Target and non-Target responses, leading to a 19.2 % improvement in online accuracy for the HVA-S condition when compared to AU-S, and reinforces the relevance of multimodal congruent stimulation. The P300 waveforms in Fig. 3 exhibit inter-subject variability, but the difference between Targets and non-Targets shows that the difference is statistically significant in the P300 peak (see Fig. 4). The increase in amplitude in the HVA-S modality compared to AU-S is also statistically significant. Comparing the P300 latency in HVA-S and AU-S modalities, the difference in the second P300 peak (related to auditory stimuli) is not statistically significant, which shows that the multimodal integration of visual and auditory stimuli varied across participants. The higher latency of the auditory peaks may be more related to the difficulty in discriminating spoken words than to the modality as hypothesized in Halder et al. (2013).

6.1.3. Task load results

Qualitative evaluation based on questionnaires and debriefings points out the positive effect of the integration of 3D spatial modulation of the auditory stimuli. Participants assessed, on average, temporal demand, physical demand, performance, and general effort as lower in spatial conditions, yet the improvement was statically significant only for temporal demand. In addition, participants also reported verbally the easiness in perception/discrimination of the auditory stimuli. While greater discrimination between auditory stimuli was reported to come from spatial location encoding, it is possible that greater duration of spatial sounds by 100 ms may have had a positive impact on discrimination and consequently on classification accuracy. The improvement in HVA-S compared to AU-S was statistically significant for mental demand, which suggests that the stimuli integration reduced the mental effort to discriminate stimuli.

6.2. Experiment II: CLIS patient

The BCI was designed specifically for its possible use by people in CLIS, aiming to provide a small communication lexicon that could go beyond the “yes/no” interaction. Given the good overall results achieved with the able-bodied group, the ultimate goal of this study was to test this BCI with a CLIS patient. This opportunity came up about 2 and a half years after validation with the control group.

The experimental results with the CLIS patient, however, did not show any typical traceable P300 waveform for both AU-S and HVA-S conditions. Nevertheless, some short discriminative time windows between Target and Non-Target responses were identified, albeit with a much lower level of discrimination than with the control group.

For the AU-S condition, these time windows occurred around 50 and 200 ms. The 50 ms peak can be related to the P50 component associated with pre-attentional filtering (gating) of sensory information (usually used as a biomarker for schizophrenia diagnosis, using the pair-clicked paradigm) (Chang et al., 2011). The P200 peak can be related to the P200 component associated with stimulus recognition and perceptual processing modulated by attention and pre-classification preceding the

P300 component (Mazer et al., 2021; Polich, 2007; Crowley and Colrain, 2004). Although a P200 was also present in non-Target events, the amplitude of attended stimuli was higher, which indicates a minimum level of attention of the patient and that he was able to identify the spoken words. In the control group, the P200 component (preceding N200) was also present but with lower amplitude and lower latency (also exhibiting some discrimination power). Compared to other CLIS studies, the only study with semantic auditory stimuli was in Bensch et al. (2014) using ECoG. The authors also reported no significant effects in N100, P200 and P300 for semantic stimulation (oddball and priming). Moreover, the oddball tones evoked N100 and P200 components across all experiments while P300 could not be detected when the patient entered in CLIS. The absence of a P300 component may suggest a cognitive impairment associated with the semantic processing of the stimuli. Additionally, a deficit in working memory required for target comparison and mental representation may play here crucial relevance. In our study, the fact that the patient was taking anxiolytics, could also have exerted a suppressive effect. In Han et al. (2019), the oddball tones led to a traceable P300 around 600 ms with one CLIS patient. Tests performed with 4 CLIS patients (Khalili-Ardali et al., 2021) using tones in a “local effect-global effect” paradigm showed global effects in 2 of the patients (related to memory updating), while paradoxically no pre-attentive components were found. Tests in patients with DOC (Wang et al., 2015) also revealed traceable P300 using oddball tones. None of the above CLIS studies has reported the possibility of a feasible BCI based on auditory stimuli. In our results, the binary offline classification performance obtained from calibration data through cross-validation shows that Target vs. non-Target discrimination was possible, although not going beyond 90 % accuracy, even for more than 10 repetitions of the same event, which shows high inter-trial variability. The best online accuracy, the gold standard metric, was 30 %, which is not enough for an effective interface, although above the chance level.

For the HVA-S condition, there was a discriminative time window at around 500 ms. No discriminative windows appeared around 50 ms and 200 ms as in the AU-S condition, but the discriminative power at 500 ms of the HVA-S condition was higher than the one found in 50 ms and 200 ms in the AU-S condition. The discriminative time window occurs across all channels, which suggests the expected integration of visual and auditory processing. The time window coincides with a stronger negative deflection of Target responses. We hypothesize that it might be associated with a delayed N200 component, which in this case would reflect conscious attention to attended stimuli, as the spoken words and visual stimuli are all different and thereby could not reflect unconscious mismatch negativity. CLIS patients are known to have a delay in higher cognitive processing (Khalili-Ardali et al., 2021) which may explain this delayed N200, exacerbated by the multimodal integration. Around 800 ms, there is a positive peak in both AU-S and HVA-S conditions that could be associated with a delayed P300 component due to semantic processing (Bensch et al., 2014), but there is no discrimination between Target and non-Target responses sustaining this hypothesis. Contrary to expectations, the offline classification accuracy was lower than that obtained in the AU-S condition and the online accuracy was at the chance level. Thus, we must conclude that the patient did not take advantage of the semantic congruence of the multimodal stimuli. The concurrent pathways of reading and listening may have had the opposite effect, considering that CLIS patients may have difficulty integrating stimuli due to impaired cognitive functions (Bensch et al., 2014; Khalili-Ardali et al., 2021). The integration of visual and auditory stimuli has not been tested with CLIS patients, but the study with DOC patients in Wang et al. (2015) reports online BCI results above 60% accuracy (‘yes/no’ paradigm).

The (low) level of vigilance (sustained attention) of the patient may have affected the results of the experiment. To further investigate this issue, we analysed the power spectral density (PSD) of the EEG for the control group and the CLIS patient. EEG was recorded while the participants were controlling the BCI. We took data from non-Target epochs and estimated the PSD with Welch’s method (see Fig. 8). Compared to the control group, the CLIS patient shows an increased power of delta rhythms which is known to be associated with drowsiness and lack of vigilance, although an increase in theta rhythms was also expected in this case (Oken et al., 2006; Paulo et al., 2021). The relation between lower frequencies (delta) and higher frequencies is higher in the CLIS participant than in the control group. This is consistent with the results obtained in Secco et al. (2021), where the PSD of the EEG of CLIS patients was analysed during resting-state. A comparison of our spectra to theirs may not be straightforward, as here EEG was recorded while participants were performing a mentally sustained demanding task, i.e., they were not in a resting state. The authors hypothesized that the lack of higher frequencies was related to the lack of means of communication. On the other hand, it is known that higher frequencies are related to memory and cognitive processing (Klimesch, 1999). In particular, theta and alpha rhythms are known to respond in opposite directions when task demand increases (theta/alpha ratio increases). Although it is not possible to be conclusive about the underlying reasons for the spectra, it is nevertheless clear that the delta/(theta+alpha+beta) ratio is higher in the CLIS patient.

6.3. Limitations of the study, overall difficulties and lessons learnt

As noted by other researchers, the results show that a successful approach with a healthy group is no guarantee of success with a clinical group. Validation with control groups is of course essential but may not be meaningful until validation with target groups is achieved. One of the main challenges of the study with the CLIS patient was the interpretation of his unique ERP patterns, which did not match those typical of the control group. Therefore, there is great uncertainty regarding the origin of the ERPs, which may eventually be speculative. To better interpret the ERPs it is therefore required to apply different tests/paradigms that allow the decoupling of patients’ responses at different levels (reflexive, perceptual, cognitive). This systematic evaluation was recently presented in Khalili-Ardali et al. (2021). Another major difficulty was the uncertainty regarding the patient’s spans of vigilance (eventually noticeably short) which interfered in the evaluation of the levels of consciousness and awareness and absolutely in the BCI performance. The use of medication that depresses the Central Nervous System, such as, for example, Diazepam that our patient was taking interferes with the level of vigilance, and ideally these drugs should be withdrawn, but this is not always possible. A “wakefulness detector” during the experiments would be very desirable.

Our study was very exploratory including only one CLIS patient and therefore the validation of the proposed BCI approach for CLIS patients was inconclusive. Nevertheless, this study highlighted many relevant issues essential for further experiments and BCI adjustments. Although the inclusion in the study of ALS and LIS patients could also have been important to previously validate the BCI, most of the related literature shows that the BCI performance only drops significantly when patients enter a CLIS (as referred to in Section 1.). Therefore, the main limitation of the study was the impossibility of conducting a longitudinal study, which would have allowed us to record the patient’s EEG history before entering the completely LIS and continuously monitor reflexive, perceptual and cognitive functions (eventually including other imaging techniques such as fMRI), reaching more meaningful conclusions. The lack of communication with the patient made it impossible to know whether the participant understood the combination of visual and auditory stimuli. As well, it was not known whether the use of HRTFs was effective in the spatial modulation of the auditory stimuli, providing the desired effect.

It was yet to show the possibility of communicating with a CLIS patient beyond simple “yes/no” interaction. Despite the unsuccessful online BCI results, the discrimination found between target and non-target responses shows that it is worth exploring a simplification of the proposed BCI. Prior assessment of patients’ working memory is

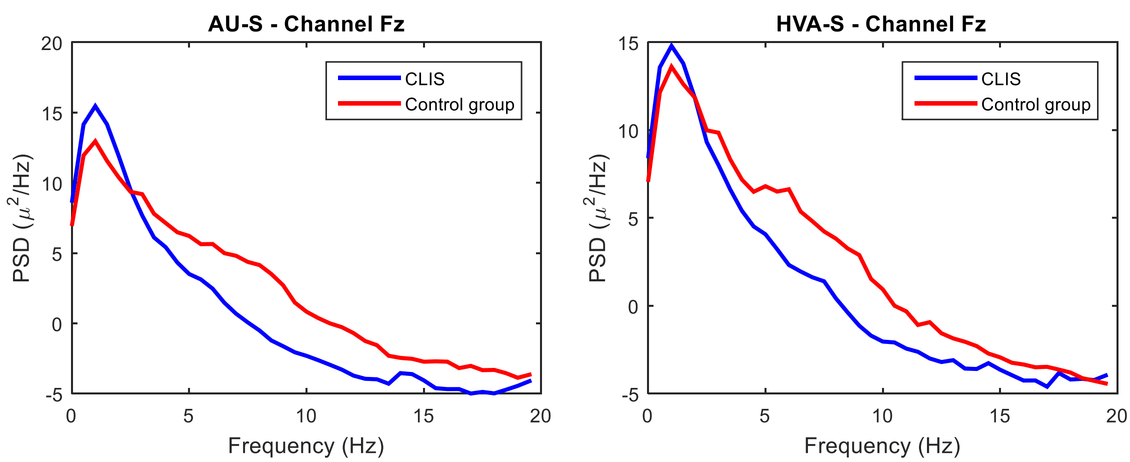


Fig. 8. Power spectral density estimated with Welch's method using EEG data (non-Target epochs) of all participants and EEG data (non-Target epochs) of the CLIS patient in both conditions (AU-S and HVA-S). Welch's method was used with the following parameters: frequency resolution of 0.5 Hz, Hamming window and 15% overlap.

necessary to fully understand whether the use of spoken words is a viable approach and, if so, how many words could be used. The offline binary classification results were also a promising indicator that eventually other machine learning techniques can lead to better online results and eventually could be combined with the detection of the vigilance state. Unfortunately, it was not possible to proceed with further experiments with the CLIS patient to evaluate other approaches. To the best of our knowledge, it was also the first experiment of this kind in Portugal. Medical doctors must be aware of the existence of BCIs and suggest to patients and family members to start using BCI as soon as possible, thus increasing the number of longitudinal studies, leading to a better systematization of procedures and interpretation of results.

The effective use of BCI in clinical and non-clinical applications in real-world scenarios still faces many challenges as discussed in (Xu et al., 2021), related to BCI usability and usefulness. Examples include device wearability, cognitive load, naturalness of human-machine interaction, and session- and user-independent models. Although at this stage, our study was mainly concerned with validating a paradigm that could evidence the possibility of communication with a CLIS patient, it was designed so that communication was natural and intuitive, reducing the workload. On the other hand, we are undertaking many efforts towards the improvement of BCI naturalness, self-paced control and workload (Cruz et al., 2021), as well as researching new methods targeting session and user-independent models (Cruz et al., 2022), which can complement the BCI approach proposed in this study.

CRediT authorship contribution statement

Gabriel Pires: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision. **Sara Barbosa:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft. **Urbano J. Nunes:** Writing - review & editing, Supervision. **Edna Gonçalves:** Writing - Formal analysis, review & editing.

Declaration of Competing Interests

The authors declare no competing interests.

Acknowledgments

This work was financially supported by the Portuguese Foundation for Science and Technology (FCT) through Project Grant B-RELIABLE (PTDC/EEI-AUT/30935/2017) and the Institute of Systems and Robotics, University of Coimbra (ISR-UC) through FCT grant UIDB/00048/2020. The authors would like to thank all participants of the control group, and patient and family members for their valuable participation and contribution. We would also like to thank Aniana Cruz for her help in EEG recordings and the Telecommunications Institute of Coimbra for providing the means to record the spoken words.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jneumeth.2022.109661.

References

An, X., Höhne, J., Ming, D., Blankertz, B., 2014. Exploring combinations of auditory and visual stimuli for gaze-independent brain-computer interfaces. PLOS One 9 (10), e111070. https://doi.org/10.1371/journal.pone.0111070.

Barbosa, S., Pires, G., Nunes, U., 2016. Toward a reliable gaze-independent hybrid BCI combining visual and natural auditory stimuli. J. Neurosci. Methods 261, 47–61. https://doi.org/10.1016/j.jneumeth.2015.11.026.

Bauer, G., Gerstenbrand, F., Rumpl, E., 1979. Varieties of the locked-in syndrome. J. Neurol. https://doi.org/10.1007/BF00313105.

Begault, D.R., 2000. Overview of spatial hearing, Part I: azimuth and elevation Perception. 3-D Sound For Virtual Reality and Multimedia. NASA. https://doi.org/10.2307/3680997.

Bensch, M., Martens, S., Halder, S., Hill, J., Nijboer, F., Ramos, A., Birbaumer, N., Bogdan, M., Kotchoubey, B., Rosenstiel, W., Schölkopf, B., Gharabaghi, A., 2014. Assessing attention and cognitive function in completely locked-in state with event-related brain potentials and epidural electrocorticography. J. Neural Eng. 11, 026006. https://doi.org/10.1088/1741-2560/11/2/026006.

Birbaumer, N., 2006. Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. Psychophysiology 43, 517–532. https://doi.org/10.1111/j.1469-8986.2006.00456.x.

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H., 1999. A spelling device for the paralysed. Nature. https://doi.org/10.1038/18581.

Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123. https://doi.org/10.1093/cercor/11.12.1110.

Carlile, S. (1996). Virtual Auditory-Space: Generation and Applications. In Landes Bioscience Springer R.G. Landes Company https://link.springer.com/book/10.1007/978-3-662-22594-3?

Chang, W.P., Arfken, C.L., Sangal, M.P., Boutros, N.N., 2011. Probing the relative contribution of the first and second responses to sensory gating indices: a meta-analysis. Psychophysiology 48 (7), 980–992. https://doi.org/10.1111/j.1469-8986.2010.01168.x.

Chaudhary, U., Xia, B., Silvoni, S., Cohen, L.G., Birbaumer, N., 2017. Brain–computer interface–based communication in the completely locked-in state. PLOS Biol. 15, e1002593. https://doi.org/10.1371/journal.pbio.1002593.

Chaudhary, U., Mrachacz-Kersting, N., Birbaumer, N., 2021. Neuropsychological and neurophysiological aspects of brain-computer-interface (BCI) control in paralysis. J. Physiol. 599 (9), 2351–2359. https://doi.org/10.1113/JP278775.

Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115 (4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021.

Cruz, A., Pires, G., Lopes, A., Carona, C., Nunes, U.J., 2021. A self-paced BCI with a collaborative controller for highly reliable wheelchair driving: experimental tests with physically disabled individuals. IEEE Trans. Hum. -Mach. Syst. 52 (2), 109–119. https://doi.org/10.1109/THMS.2020.3047597.
Cruz, A., Pires, G., Nunes, U.J., 2022. Spatial filtering based on Riemannian distance to improve the generalization of ErrP classification. Neurocomputing 470, 236–246. https://doi.org/10.1016/j.neucom.2021.10.078.
Farwell, L.A., Donchin, E., 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr. Clin. Neurophysiol. 70, 510–523. https://doi.org/10.1016/0013-4694(88)90149-6.
Frauenberger, C., & Noisternig, M. , 2003. 3D audio interfaces for the blind. In: Proceedings of the 9th International Conference on Auditory Display. pp. 280–283.
Gallegos-Ayala, G., Furdea, A., Takano, K., Ruf, C.A., Flor, H., Birbaumer, N., 2014. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. Neurology 82, 1930–1932. https://doi.org/10.1212/WNL.0000000000000449.
Gardner, W., 1999. 3D audio and acoustic environment modeling. Wave Arts, Inc, 1–9. https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.8.1665.
Gardner, W.G., Martin, K.D., 1995. HRTF measurements of a KEMAR. J. Acoust. Soc. Am. 97 (6), 3907–3908. https://doi.org/10.1121/1.412407.
Gondan, M., Niederhaus, B., Rösler, F., Röder, B., 2005. Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Percept. Psychophys. 67, 713–726. https://doi.org/10.3758/BF03193527.
Guger, C., Spataro, R., Allison, B.Z., Heilinger, A., Ortner, R., Cho, W., La Bella, V., 2017. Complete locked-in and locked-in patients: command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools. Front. Neurosci. 11 https://doi.org/10.3389/fnins.2017.00251.
Halder, S., Hammer, E.M., Kleih, S.C., Bogdan, M., Rosenstiel, W., Birbaumer, N., Kübler, A., 2013. Prediction of auditory and visual p300 brain-computer interface aptitude. PLOS One 8 (2), e53513. https://doi.org/10.1371/journal.pone.0053513.
Han, C.-H., Kim, Y.-W., Kim, D.Y., Kim, S.H., Nenadic, Z., Im, C.-H., 2019. Electroencephalography-based endogenous brain–computer interface for online communication with a completely locked-in patient. J. Neuroeng. Rehabil. 16 (1), 18. https://doi.org/10.1186/s12984-019-0493-0.
Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. 52 (C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.
Hill, N.J., Ricci, E., Haider, S., McCane, L.M., Heckman, S., Wolpaw, J.R., Vaughan, T.M., 2014. A practical, intuitive brain-computer interface for communicating "yes" or "no" by listening. J. Neural Eng. 11 (3), 035003 https://doi.org/10.1088/1741-2560/11/3/035003.
Höhne, J., Tangermann, M., 2014. Towards user-friendly spelling with an auditory brain-computer interface: the CharStreamer paradigm. PLOS One 9 (6), e98322. https://doi.org/10.1371/journal.pone.0098322.
Höhne, J., Schreuder, M., Blankertz, B., Tangermann, M., 2011. A novel 9-class auditory ERP paradigm driving a predictive text entry system. Front. Neurosci. 5 (August), 99. https://doi.org/10.3389/fnins.2011.00099.
Höhne, J., Krenzlin, K., Dähne, S., Tangermann, M., 2012a. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. J. Neural Eng. 9 (4), 045003 https://doi.org/10.1088/1741-2560/9/4/045003.
Käthner, I., Ruf, C. a, Pasqualotto, E., Braun, C., Birbaumer, N., Halder, S., 2013. A portable auditory P300 brain-computer interface with directional cues. Clin. Neurophysiol. 124 (2), 327–338. https://doi.org/10.1016/j.clinph.2012.08.006.
Khalili-Ardali, M., Wu, S., Tonin, A., Birbaumer, N., Chaudhary, U., 2021. Neurophysiological aspects of the completely locked-in syndrome in patients with advanced amyotrophic lateral sclerosis. Clin. Neurophysiol. 132 (5), 1064–1076. https://doi.org/10.1016/j.clinph.2021.01.013.
Kim, H., Hahm, J., Lee, H., Kang, E., Kang, H., Lee, D.S., 2015. Brain networks engaged in audiovisual integration during speech perception revealed by persistent homology-based network filtration. Brain Connect. 5, 245–258. https://doi.org/10.1089/brain.2013.0218.
Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29, 2.
Kübler, A., Birbaumer, N., 2008. Brain-computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? Clin. Neurophysiol. 119 (11), 2658–2666. https://doi.org/10.1016/j.clinph.2008.06.019.
Kübler, A., Furdea, A., Halder, S., Hammer, E.M., Nijboer, F., Kotchoubey, B., 2009. A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. Ann. N. Y. Acad. Sci. 1157, 90–100. https://doi.org/10.1111/j.1749-6632.2008.04122.x.
Lotte, F., Jeunet, C., Chavarriaga, R., Bougrain, L., Thompson, D.E., Scherer, R., Mowla, M.R., Kübler, A., Grosse-Wentrup, M., Dijkstra, K., Dayan, N., 2020. Turning negative into positives! Exploiting 'negative' results in brain–machine interface (BMI) research. Brain-Comput. Interfaces 6, 178–189. https://doi.org/10.1080/2326263X.2019.1697143.
Lu, Z., Li, Q., Gao, N., Yang, J., Bai, O., 2019. A novel audiovisual P300-speller paradigm based on cross-modal spatial and semantic congruence. Front. Neurosci. 13 https://doi.org/10.3389/fnins.2019.01040.
Lulé, D., Noirhomme, Q., Kleih, S.C., Chatelle, C., Halder, S., Demertzi, A., Bruno, M.A., Gosseries, O., Vanhaudenhuyse, A., Schnakers, C., Thonnard, M., Soddu, A., Kübler, A., Laureys, S., 2013. Probing command following in patients with disorders of consciousness using a brain-computer interface. Clin. Neurophysiol. 124, 101–106. https://doi.org/10.1016/j.clinph.2012.04.030.
Marchetti, M., Priftis, K., 2015. Brain-computer interfaces in amyotrophic lateral sclerosis: a metanalysis. Clin. Neurophysiol. 126, 1255–1263. https://doi.org/10.1016/j.clinph.2014.09.017.
Mazer, P., Macedo, I., Paiva, T.O., Ferreira-Santos, F., Pasion, R., Barbosa, F., Almeida, P., Silveira, C., Cunha-Reis, C., Marques-Teixeira, J., 2021. Abnormal habituation of the auditory event-related potential P2 component in patients with schizophrenia. Front. Psychiatry 12. https://doi.org/10.3389/fpsyt.2021.630406.
Middendorf, M., McMillan, G., Calhoun, G., Jones, K.S., 2000. Brain-computer interfaces based on the steady-state visual-evoked response. IEEE Trans. Rehabil. Eng. 8, 211–214. https://doi.org/10.1109/86.847819.
Murguialday, A.R., Hill, J., Bensch, M., Martens, S., Halder, S., Nijboer, F., Schoelkopf, B., Birbaumer, N., Gharabaghi, A., 2011. Transition from the locked in to the completely locked-in state: a physiological analysis. Clin. Neurophysiol. 122, 925–933. https://doi.org/10.1016/j.clinph.2010.08.019.
Okahara, Y., Takano, K., Nagao, M., Kondo, K., Iwadate, Y., Birbaumer, N., Kansaku, K., 2018. Long-term use of a neural prosthesis in progressive paralysis. Sci. Rep. 8 https://doi.org/10.1038/s41598-018-35211-y.
Oken, B.S., Salinsky, M.C., Elsas, S.M., 2006. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin. Neurophysiol. 117 (9), 1885–1901. https://doi.org/10.1016/j.clinph.2006.01.017.
Pan, Z., Cheok, A.D., Yang, H., Zhu, J., Shi, J., 2006. Virtual reality and mixed reality for virtual learning environments. Comput. Graph. 30 (1), 20–28. https://doi.org/10.1016/j.cag.2005.10.004.
Paulo, J.R., Pires, G., Nunes, U.J., Member, S., 2021. Cross-subject zero calibration driver ' s drowsiness detection: exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification. 29, 905–915. https://doi.org/10.1109/TNSRE.2021.3079505.
Pires, G., Nunes, U., Castelo-Branco, M., 2011. Statistical spatial filtering for a P300-based BCI: Tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis. J. Neurosci. Methods 195 (2), 270–281. https://doi.org/10.1016/j.jneumeth.2010.11.016.
Pires, G., Nunes, U., Castelo-Branco, M., 2012. Comparison of a row-column speller vs. a novel lateral single-character speller: Assessment of BCI for severe motor disabled patients. Clin. Neurophysiol.: Off. J. Int. Fed. Clin. Neurophysiol. 123, 1168–1181. https://doi.org/10.1016/j.clinph.2011.10.040.
Pokorny, C., Klobassa, D.S., Pichler, G., Erlbeck, H., Real, R.G.L., Kübler, A., Lesenfants, D., Habbal, D., Noirhomme, Q., Risetti, M., Mattia, D., Müller-Putz, G.R., 2013. The auditory P300-based single-switch brain-computer interface: paradigm transition from healthy subjects to minimally conscious patients. Artif. Intell. Med. 59 (2), 81–90. https://doi.org/10.1016/j.artmed.2013.07.003.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.
Rummukainen, O.S., Robotham, T., Habets, E.A.P., 2021. Head-related transfer functions for dynamic listeners in virtual reality. Appl. Sci. 11 (14) https://doi.org/10.3390/app11146646.
Schreuder, M., Blankertz, B., Tangermann, M., 2010. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLOS One 5 (4), e9813. https://doi.org/10.1371/journal.pone.0009813.
Schreuder, M., Rost, T., Tangermann, M., 2011. Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Front. Neurosci. 5 (October), 112. https://doi.org/10.3389/fnins.2011.00112.
Secco, A., Tonin, A., Rana, A., Majid, A.J., Birbaumer, N., & Chaudhary, U., 2021. RESEARCH ARTICLE EEG power spectral density in locked-in and completely locked-in state patients: a longitudinal study. 0123456789, 473–480.
Sellers, E.W., Vaughan, T.M., Wolpaw, J.R., 2010. A brain-computer interface for long-term independent home use. Amyotroph. Lateral Scler. 11, 449–455. https://doi.org/10.3109/17482961003777470.
Simon, N., Käthner, I., Ruf, C.A., Pasqualotto, E., Kübler, A., Halder, S., 2014. An auditory multiclass brain-computer interface with natural stimuli: usability evaluation with healthy participants and a motor impaired end user. Front. Hum. Neurosci. 8, 1039. https://doi.org/10.3389/fnhum.2014.01039.
Spüler, M., 2019. Questioning the evidence for BCI-based communication in the complete locked-in state. PLOS Biol. 17, 1–5. https://doi.org/10.1371/journal.pbio.2004750.
Sundareswaran, V., Wang, K., Chen, S., Behringer, R., McGee, J., Tam, C., & Zahorik, P. , 2003. 3D audio augmented reality: implementation and experiments. In: Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 1–2. 〈https://doi.org/10.1109/ISMAR.2003.1240728〉.
Thurlings, M.E., Brouwer, A.-M., van Erp, J.B.F., Werkhoven, P., 2014. Gaze-independent ERP-BCIs: augmenting performance through location-congruent bimodal stimuli. Front. Syst. Neurosci. 8 (September), 1–14. https://doi.org/10.3389/fnsys.2014.00143.
Wang, F., He, Y., Pan, J., Xie, Q., Yu, R., Zhang, R., Li, Y., 2015. A novel audiovisual brain-computer interface and its application in awareness detection. Sci. Rep. 5 https://doi.org/10.1038/srep09962.
Wenzel, E., Arruda, M., Kistler, D., Wightman, F., 1993. Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am. 94 (1), 111–123. https://doi.org/10.1121/1.407089.
Wightman, F.L., Kistler, D.J., 1989. Headphone simulation of free-field listening. I: stimulus synthesis. J. Acoust. Soc. Am. 85 (2), 868–878. https://doi.org/10.1121/1.397558.

Wolpaw, J.R., McFarland, D.J., 2004. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. USA 101 (51), 17849–17854. https://doi.org/10.1073/pnas.0403504101.

Wolpaw, J.R., Bedlack, R.S., Reda, D.J., Ringer, R.J., Banks, P.G., Vaughan, T.M., Heckman, S.M., McCane, L.M., Carmack, C.S., Winden, S., McFarland, D.J., Sellers, E.W., Shi, H., Paine, T., Higgins, D.S., Lo, A.C., Patwa, H.S., Hill, K.J., Huang, G.D., Ruff, R.L., 2018. Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. Neurology 91 (3), e258–e267. https://doi.org/10.1212/WNL.0000000000005812.

Xu, M., He, F., Jung, T.P., Gu, X., Ming, D., 2021. Current challenges for the practical application of electroencephalography-based brain–computer interfaces. Engineering 7 (12), 1710–1712. https://doi.org/10.1016/j.eng.2021.09.011.