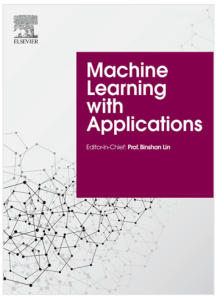
Journal Pre-proof



Single-trial stimuli classification from detected P300 for augmented Brain-Computer Interface: A deep learning approach

Jessica Leoni, Silvia Carla Strada, Mara Tanelli, Alessandra Brusa, Alice Mado Proverbio

PII: S2666-8270(22)00074-3
DOI: https://doi.org/10.1016/j.mlwa.2022.100393
Reference: MLWA 100393

To appear in: *Machine Learning with Applications*

Received date : 14 April 2022
Revised date : 26 July 2022
Accepted date : 26 July 2022

Please cite this article as: J. Leoni, S.C. Strada, M. Tanelli et al., Single-trial stimuli classification from detected P300 for augmented Brain-Computer Interface: A deep learning approach. *Machine Learning with Applications* (2022), doi: https://doi.org/10.1016/j.mlwa.2022.100393.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Journal Pre-proof

Revised manuscript (Clean Version)

Single-Trial Stimuli Classification from Detected P300 for Augmented Brain-Computer Interface: a Deep Learning Approach

Jessica Leoni, Silvia Carla Strada[1]

Politecnico di Milano, Dipartimento di Elettronica Informazione e Bioingegneria (DEIB), Milan, Italy

jessica.leoni@polimi.it, silvia.strada@polimi.it

Mara Tanelli[2]

Politecnico di Milano, Dipartimento di Elettronica Informazione e Bioingegneria (DEIB), Milan, Italy

Istituto di Elettronica e Ingegneria dell'Informazione e delle Telecomunicazioni, Torino, Italy

mara.tanelli@polimi.it

Alessandra Brusa, Alice Mado Proverbio

Milan Center for Neuroscience, Department of Psychology, Università di Milano-Bicocca, Milan, Italy

a.brusa@campus.unimib.it, mado.proverbio@unimib.it

[1]Corresponding author

[2]The work of Mara Tanelli was partially supported by the project YOU-Share, funded by Fondazione Cariplo

Journal Pre-proof

Journal Pre-proof

Abstract

The purpose of advanced Brain-Computer Interfaces (BCIs) is to connect the human brain with an external device without using the muscular system. To do this, they must effectively process mental activity and infer information on the users' intentions and directives. This work proposes a novel and explainable BCI system capable of recognizing P300 deflection in single-trial EEGs with higher accuracy compared to the literature gold standard. Moreover, the proposed deep-learning approach allows us to go beyond the mere P300 detection, which is, to our best knowledge, the current state of the art. Indeed, we first identify the P300-related signal in the single-trial EEG signal, and then, we further discriminate the ERPs associated to the detected P300 between visual and auditory stimuli-related. To do this, we employ a CNN-LSTM neural network, which manages a 3D data representation of the acquired EEG signals. The performance of the approach is tested on experiments carried out on 22 subjects, revealing a 82.4% F1-score in P300 identification and 82.4% discriminating between visual and auditory stimuli. The employed algorithmic procedure also reports the most relevant each EEG channels in determining the predictions, adding interpretability to the proposed AI-based tools. These results pave the way for more sophisticated BCIs, capable of extending the set of available actions for the patients. The project was pre-approved by the Research Assessment Committee of the Department of Psychology (CRIP) for minimal risk projects, under the aegis of the Ethical Committee of University of Milano-Bicocca, on May 27th, 2019, protocol number RM-2019-193.

Keywords: Brain-Computer Interface (BCI); EEG; Deep learning; Classification

1. Introduction and Background

Brain-computer interfaces (BCIs) address systems that directly bridge the human brain with external devices without requiring peripheral muscular activ-

Journal Pre-proof

Journal Pre-proof

ity (Wolpaw et al. (2002)). BCIs were first introduced for medical purposes. Indeed, they were employed as visual or auditory spellers to allow patients affected by severe muscular diseases, such as locked-in syndrome and amyotrophic lateral sclerosis, to communicate with the external world. Recently, the widespread of this technology paved the way also for non-medical applications (Müller et al. (2008)). As result, to date, BCIs are vastly employed, as mental state control (Blankertz et al. (2010)), drunkness recognition (Malar et al. (2011)), autonomous vehicles control (Waibel (2011)), gaming (Royer et al. (2010)), and authentication (Nakanishi et al. (2013)).

Regardless of the specific application, a BCI aims to identify specific patterns in a patient's cognitive processes and translate them accordingly to provide the corresponding machine commands. Although several methods can be leveraged to measure brain processes, over 80% of BCI publications rely on the electroencephalogram (EEG) (Mason et al. (2007)). In detail, it describes the cognitive processes reporting the average activity of the dendritic currents measured over time by electrodes placed on the scalp according to standardized configurations. Therefore, it is an economical, non-invasive, and high temporal resolution technique compared to alternative approaches such as functional near-infrared spectroscopy, functional magnetic resonance imaging, and electrocorticogram (De Venuto & Mezzina (2021)).

It follows that by leveraging an ad-hoc defined experimental setup, it is possible to elicit specific patterns in the subject's brain that can be detected by the BCI system from the EEG analysis. The procedure of the experimental setup depends on the target application; however, three are main elicited patterns that are effective in BCI applications: motor imagery (MI), steady-state visual evoked potential (SSVEP), and P300 (Chan et al. (2015)).

MI deals with the mental rehearsal of physical movement tasks, such as raising a hand. Motion planning induces a synchronous increase or decrease of the neuronal population, affecting the energy content of EEG frequency bands depending on the moved limb. The BCI can detect these phenomena, called event-related synchronization (ERS) or desynchronization (ERD), recognizing

Journal Pre-proof

the specific motion and performing the associated action. Usually, this action consists of moving a real or a virtual object according to the users' direction of motion. It follows that the primary applications of MI-based elicitation schema include wheelchair (Huang et al. (2012)) and screen cursor control (Huang et al. (2009)).

SSVEPs, instead, are elicited by triggering a subject with a visual stimulus, *e.g.* flashing light, repeated at a specific frequency. They induce an increase of EEG energy at the corresponding frequency. According to this paradigm, several visual stimuli are presented to the subject, flickering with different frequencies. A priori is established a mapping between the set of available actions and the flickering frequencies. The user must focus on the visual stimulus whose frequency corresponds to the intended action. As a result, the EEG energy increases accordingly; the BCI recognizes the SSVEPs-induced peak, gets the corresponding frequency, and performs the associated action (Zhu et al. (2010)).

Finally, P300 represents one of the most significant event-related brain potentials (ERP) components, consisting of time-locked responses to a specific class of stimuli. Since it is easy to trigger and measure compared to other ERP components, P300 is the target of most ERP-based BCIs (Allison et al. (2020)). Also, although the overall ERP trend depends on the specific stimulus that elicited it, the P300 component shape is deterministic. In detail, P300 consists of a positive deflection occurring 300 ms after the recognition of a target stimulus (Polich (2012)). According to psychophysiological literature, as (Proverbio & Zani (2003)) and (Polich (2020)), the P300 component would reflect context updating processes (Fonken et al. (2020)), elicited by stimulus-driven attention, if anterior, and voluntary attention allocation, if posterior (Polich (2007)). Moreover, recent evidence suggest that P300 responses directly reflect context updating and learning (Polich (2020)). Again, the P300 component would also reflect working memory processes (Linden (2005)), categorization's certainty (Polich (2007)), conscious processing, stimulus recognition and coding (Dehaene & Changeux (2011)), stimulus arousal, and valence (Proverbio et al. (2020)), stimulus familiarity (Herron et al. (2003)). In other words, the greater the P300

Journal Pre-proof

amplitude evoked by a stimulus, the more distinctive and consciously vivid would be its mental representation, as well as its subsequent memory. It is evident why the P300 component of ERPs represents a precious tool for BCI technology, as it directly reflects the brain activation supporting the conscious mental representations of a stimulus. Also, it is highly effective as a plethora of experimental paradigms can elicit it.

SSVEP and P300 are often employed in spellers, *i.e.,* BCIs allowing the user to communicate letter after letter. However, SSVEP range of different stimuli is limited, as the characteristic frequencies must be non-harmonic and compliant with the monitor refresh rate (Volosyak et al. (2009)). On the other hand, P300-based spellers are robust to this issue and also compliant with different stimuli. Indeed, despite most of them relying on visual stimuli (Farwell & Donchin (1988)), auditory spellers are proposed to allow communication also for patients incapable of controlling eye muscles (Schreuder et al. (2010)). Therefore, considering BCI-based communication systems, P300 represents the most effective pattern elicited in the subject's brain.

However, regardless of the stimulation, P300 is characterized by a low signal-to-noise ratio (SNR) (Schomer & Da Silva (2012)). Therefore, BCIs usually rely on P300 obtained from grandaveraged ERPs. Indeed, synchronously averaging the multiple EEG signals concerning the same stimulus allows for improving SNR, as the background brain activity, which acts as a zero-mean Gaussian noise, is attenuated downstream of the process. Even if grandaverage improves BCIs robustness, it adds a temporal overhead in the elicited pattern recognition process. It follows that it may not be compliant with real-time applications requirements, burdening the end-user's communication process.

Therefore, approaches are needed to allow a BCI to identify not averaged P300 from a single stimulus repetition accurately. This achievement is essential in designing an effective system that enables real-time communication.

Moreover, the literature approaches aims at detecting the ERPs evoked by a specific stimulus, performing the so-called target vs. non-target binary classification. Accordingly, the pool of actions available for the patients is reduced. To

Journal Pre-proof

overcome this limitation, expanding BCIs functionalities, more stimuli can be leveraged, in order to train the classifiers to detect the ERP and also the eliciting stimulus, *e.g.*, distinguishing it by visual or auditory based on its temporal trend. This allows for expanding the dictionary of possible actions available to the subject, paving the way for more sophisticated and effective BCIs.

1.1. Related Works

The literature presents several approaches to detect the P300 in the EEG signal. Considering those based on single-trial experiments, independent component analysis (ICA) was initially set as one of the most promising techniques, reducing the processing time and improving the information transfer rate. Based on this method, Li et al. designed a BCI system capable of detecting the P300 with a 76.67% accuracy (Li et al. (2009)).

However, considering the unfavorable SNR, the EEG non-stationarity over time, and the inter- and intra-subject variations, the advent of machine-learning provides algorithms that soon overcame traditional approaches, given their computational efficiency and the capabilities of data-driven modeling of complex phenomena (Müller et al. (2008)). It follows that recent literature approaches mainly rely on machine learning-based approaches. The most widely used includes linear classifiers, linear discriminant analysis (LDA), and support vector machines-based (SVM) approaches, which still are, the most popular choice in real-time EEG based-BCIs (Lotte et al. (2018)). Considering visual speller, Shrinkage LDA achieves remarkable performances, assessing an online accuracy of about 70% (Blankertz et al. (2011)). Regarding auditory spellers, instead, the state-of-the-art performances were reported by Lelievre, who achieved 71.4% accuracy using SVM. (Lelievre et al. (2013)). Despite the linear classifiers' undisputed effectiveness, their classification capabilities are limited when the data is not linearly separable. Accordingly, non-linear approaches such as the discriminative canonical pattern matching are proved to outperform LDA results, assessed as a promising algorithm for BCI systems based on P300 (Xiao et al. (2019)).

Journal Pre-proof

However, the development of artificial neural networks (ANN) soon revealed their compliance with BCI applications, being nowadays the most adopted choice. The first attempt was presented by Cecotti et al. (Cecotti & Graser (2010)) who, leveraging convolutional neural networks (CNN), outperformed the BCI competition winners, considering the same P300-speller dataset. Also, Kshirsagar et al., in a recent work, proposed a weighted ensemble of CNNs to detect P300 from online single-trial BCI, assessing 92.64% accuracy (Kshirsagar & Londhe (2020)). From then, several networks architectures were proposed, like recurrent neural networks (RNN) (Tal & Friedman (2019)), and long-short term memory networks (LSTM) (Joshi et al. (2018)), despite CNN proves to represent the most effective solution (Vareka (2021)). Indeed, although great performance is achieved by RNN-based approaches (Fedjaev (2017)), they fail in taking advantage of EEG spatial characteristics, leveraging only temporal dependencies information. CNNs are suitable for the EEG data since they can account for their spatio-temporal structure. Also, CNN achieves outstanding performances in features extraction, preventing from selecting them in a manual, handcrafted process (Lotte et al. (2018)). The first CNN-based approaches, such as Cecotti's, leverage a 2D matrix representation of the EEG data since these networks are designed to handle image-like structures. Each row corresponds to a time instant and each column to a channel. In 2017 a novel 3D data representation was presented, allowing to represent more effectively the EEG information to the CNN (Carabez et al. (2017)). Each matrix encodes the EEG channels' spatial distribution according to the proposed data structure, while the third dimension represents time. Therefore, the EEG signal can be represented as a series of 2D matrices representing the electrodes configuration on the patient scalp, tracing from time to time the voltage measured by each channel in the respective matrix cell. CNN is also used in recent work by De Venuto et al. (De Venuto & Mezzina (2021)), which leverages an autoencoder-(1D)CNN to extract the features of the signals acquired by a 6-channels EEG and detect the presence of the P300, achieving a 70.0% F1-Score. Leveraging this effective EEG data representation allows for achieving high performances in P300 detection,

Journal Pre-proof

especially considering hybrid networks architectures, such as the so-called ConvLSTM network (Joshi et al. (2018)). This configuration proves to be suitable for EEG-based BCI, requiring minimal channels pre-processing, and efficiently processing spatial and temporal information, combining the advantages of the two standard networks architectures.

However, poor interpretability often affects ANN-based approaches, representing the main barrier in adopting these systems (Molnar (2020)). Indeed, linear classifiers are easy to be interpreted, allowing for investigating their decision-making process and promoting a better understanding of the monitored processes (Du et al. (2019)). Therefore, to provide insight into the neural networks decision-making process, techniques have been proposed, such as the permutation importance (Altmann et al. (2010)). Such metrics can rank the provided features based on their importance in determining predictions.

In addition, the P300-based BCI systems presented so far address binary classification problems, aiming to recognize the presence or absence of the target response in the single-trial EEG signal. Although the produced spellers improve locked-in patients' life quality, bridging them to the outside world and extending the pool of recognized stimuli would further increase BCI capabilities, providing them with a more comprehensive set of possible actions. Recent studies, such as the one proposed by Wirth et al. (Wirth et al. (2020)), demonstrate that, as ERPs are lock-in stimulus-specific responses, it is possible to classify the induced response based on the eliciting stimulus. In detail, they distinguish two different types of error-related potential, *i.e.,* a particular ERP elicited when the subject perceives a mistake in executing a task, with 68% accuracy.

Therefore, the literature overview concerning machine learning-based BCI systems reveals that these techniques effectively detect the P300, despite the unfavorable SNR, which paves the way for further distinguishing the eliciting stimulus, extending the pool of BCI provided actions. However, new approaches are required to increase these systems' explainability and accuracy, especially concerning classifying the domain of belonging for the eliciting stimuli.

Journal Pre-proof

1.2. Problem Statement

In this work, we propose an accurate and explainable deep-learning-based BCI system able to detect not averaged P300 in single-trial EEG and further distinguish them respect to the sensorial domain of belonging, *i.e.,* visual or auditory. In detail, three innovative contributions are provided:

1. A hierarchical classifier capable of recognizing the sensorial domain of the eliciting stimulus related to the detected P300;
2. An experimental methodology to collect a single-trial EEG dataset referred to P300 elicited by more than one kind of stimulus;
3. A procedure to explain the classifier decision-making process.

Concerning the first innovative contribution, starting from state-of-the-art solutions, we produce a BCI able to go beyond the traditional P300 detection for the very first time. To this extent, we design a hierarchical classifier combining two CNN-LSTM networks. The first one aims to identify, in the single-trial EEG, samples referred to a P300, similar to the literature approaches. But then, a second neural network further distinguishes the detected ERP associated with the P300 concerning the eliciting stimulus, which can be visual or auditory. To the best of the authors' knowledge, this is the first time a BCI system can manage more than one stimulus and recognize the sensorial domain of belonging. Indeed, state-of-the-art approaches rely on a single stimulus and identify the elicited P300 in the EEG signal. Therefore, we also designed a novel experimental procedure to collect a dataset referred to P300 elicited by stimuli belonging to different sensorial domains. This was kay to produce a dataset on which resorting to evaluate the performances of our BCI system. Accordingly, we collected data considering 22 volunteers triggered by visual and auditory stimuli and leveraged it to assess our BCI system's performance. In this process, two indicators were considered: the F1-Score in detecting the P300 and recognizing the ERP's sensorial domain and the interpretability of the classifier' decision-making process, which was measured according to the so-called permutation importance (Altmann et al. (2010)). Moreover, the proposed BCI is

Journal Pre-proof

evaluated according to a subject-specific and an intra-subject procedure, investigating if providing training data concerning multiple subjects may improve the learning process. The last innovative contribution concerns the interpretability of the classifier's decision-making process provided by the combination of the hierarchical classifier structure with the leveraged 3D data structure, inspired by that proposed by Carabez et al. (Carabez et al. (2017)). Leveraging a hierarchical structure allows dividing the classification task in two steps: P300 detection and eliciting stimulus' sensorial domain recognition. The permutation importance method was applied to the two CNN-LSTM networks separately; therefore, for each one a features ranking is provided, based on their relevance in the respective neural network predictive process. The 3D data structure was key to provide interpretability, as let the features be the EEG channels. Accordingly, the ranking returned by the permutation importance provides information about the importance of each EEG channel in determining the predictions, providing insights into the brain regions most involved in the ERPs recognition process.

This study paves the way for more sophisticated BCIs, both for medical and non-medical applications, allowing to extend the pool of recognized stimuli, *i.e.,* the set of actions available for the users.

2. Method

This section details the methodology proposed to design a BCI capable of detecting P300 in single-trial EEG, further distinguishing the belonging ERP based on the sensorial domain of the eliciting stimulus. First, we describe the pre-processing phase, aimed at discharging negligible EEG channels and improving SNR. Then, we present the 3D EEG data representation. Finally, we detail the architecture of the novel hierarchical classifier produced to further classify the sensorial domain of the eliciting stimulus.

Journal Pre-proof

Journal Pre-proof

2.1. Acquisition Experimental Setup

The main novelty of the proposed approach was to further distinguish the detected P300 components of ERPs based on the sensorial domain of their eliciting stimulus (*i.e.*, visual vs. auditory).

Indeed, as reported in Section 1, literature approaches only perform target vs. non-target binary classification. It follows that the public EEG dataset repositories, *e.g.,* (OpenBCI), concern experiments in which a single stimulus is repeatedly provided to a subject. Therefore, as we need an ad-hoc dataset to evaluate our system performances, we designed an experimental setup including both visual and auditory stimulation. The produced dataset collected data referred to 22 healthy subjects, 11 men and 11 women, aged between 19 and 30 years. All of them proved to be right-handed, according to the proposed Oldfield Inventory test (Oldfield (1971)). To take part in the experiment, volunteers gave their written and informed consent under the Declaration of Helsinki (BMJ 1991; 302: 1194), and with the approval of the Ethics Committee of the University of Milan-Bicocca (prot. N°. RM-2019-193).

During the experiment, each subject sat comfortably in an acoustically shielded and faradized cubicle, wearing an elastic cap equipped with 126 electrodes, arranged according to the international standard defined by the Oostenveld 10-5 system (Oostenveld & Praamstra (2001)). Participants were asked to fixate a red dot located at the center of a screen, placed 114 cm from their eyes. To each volunteer, 11 different experimental runs were administered randomly mixed. 8 of them, provided a visual stimulation and lasted 2 min and 5 s each, while 3 of them provided an auditory stimulation and lasted 1 min and 50 s each. The whole stimulus set comprised 360 images belonging to 9 categories (40 images per category) and 120 auditory stimuli belonging to 3 categories (40 sound files per category). Each stimulus was presented for 1500ms while the inter-stimulus interval (ISI) randomly varied between 500±100ms. In detail, visual stimuli consisted of static pictures presented at the center of a white background. They might belong to 9 different categories, based on their depicted content, namely: faces of adults, infants, and animals, dressed bodies, tools, ev-

Journal Pre-proof

eryday objects, letters, words, and checkerboards. Landscape images were used as visual target stimuli. Auditory stimuli consisted in short fragments belonging to 3 sub-categories: emotional vocalizations, words, and piano music. Natural sounds were used as auditory target stimuli. Each audio clip lasted 1500ms. Audio stimuli were normalized and leveled in intensity. Volunteers were required to press a response key as accurately and quickly as possible whenever an infrequent target related to nature was detected (*e.g.*, pictures of natural landscapes or sea waves sounds, depending on the sensorial domain involved). Responses were provided by pressing a response key with the index finger of either the left or right hands. Hand order was alternated throughout the recording session. The hand order and task conditions were counterbalanced across subjects. For each experimental run, the target number varied pseudo-randomly between 3-5. Fictitious, rare targets were used to avoid contaminating the evoked potentials of interest (non-targets) with the motor potential artifacts linked to the motor response (Proverbio et al. (2020, 2011)). Therefore, the rare targets acted as fictious fillers for keeping the subjects' attention on the stimulation.

2.2. EEG Pre-Processing

EEG signals were acquired and analyzed via EEProbe recording software (ANT Neuro system, Enschede, The Netherlands). Stimuli presentation and triggering was performed using EEvoke Software for audiovisual presentation (ANT Neuro system, Enschede, The Netherlands). Digital amplifiers Synamps were used. The EEG was continuously recorded from 126 scalp sites at a sampling rate of 512 Hz. Horizontal and vertical eye movements were also recorded. Averaged ears served as the reference lead. The EEG and electro-oculogram (EOG) were amplified with a half-amplitude band pass of 0.016–70 Hz. Electrode impedance was kept below 5 $k\Omega$,. Signals coming from hEOG, vEOG, M1, and M2 electrodes were discarded in that not relevant for classification purposes. Baseline correction was applied to each EEG channel. This procedure consists of subtracting from each channel the average voltage recorded in the 200 ms preceding the stimulation. Since P300 is a low-frequency component, an offline

Journal Pre-proof

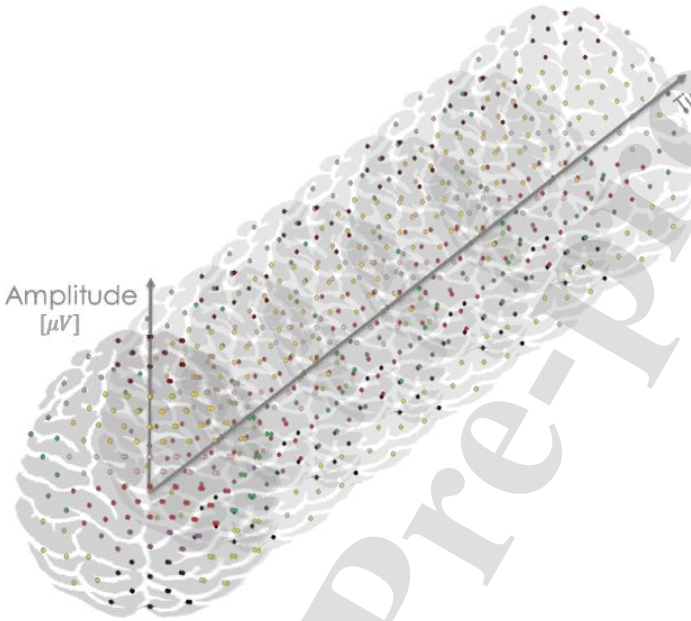
band-pass filter was then applied, between 0.1 Hz and 20 Hz. In addition, artifacts rejection was performed, thresholding channels amplitude to $\pm 50\ \mu V$, according to standard guidelines (Luck (2014); Zani & Proverbio (2003)).

2.3. 3D EEG Data Representation

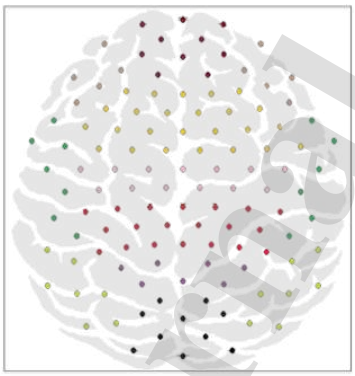
Given the outstanding performances achieved allowing a neural network to automatically extract features, compared to handcrafted processes (Lotte et al. (2018)), in the proposed BCI an ANN performs this task. Nevertheless, data representation affects ANN performances in extracting features. As mentioned in Section 1, 3D data representation outperforms the 2D one, accounting for both spatial and temporal EEG dependencies. Therefore, we encode data in a series of 21x21 matrices, as reported in Figure 1. In detail, each matrix represents the 2D projection of the electrodes cap. Accordingly, the pixel corresponding to an electrode collects the measured voltage in the instant the matrix is referred; pixels that do not correspond to an electrode are zero-padded. It follows that considering an ordered series of matrices allows for accounting also for EEG temporal evolution. As introduced in Section 1, resorting to this data structure was key to provide classifier's decision-making process interpretability. Indeed, according to it, each feature consists of an EEG channel. Also, it effectively represents the spatio-temporal information characterizing the single-trial EEG datum. Indeed, EEG temporal information determines the signal morphology, which the neuroscientist primarily considers associating an ERP to the corresponding eliciting stimulus. Considering the spatial information, it provides insights into the activity of the respective brain area, which is fundamental to inferring the sensorial domain of the eliciting stimulus.

2.4. Hierarchical Classifier

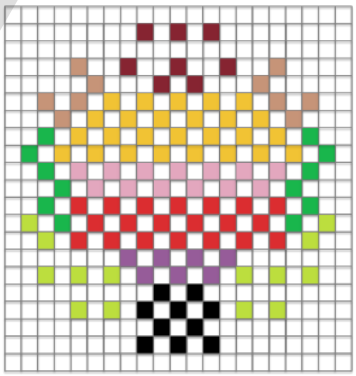
The main contribution provided in our work consists of designing an explainable BCI able to distinguish further its ERPs referred to the detected P300, basing on the sensorial domain of the eliciting stimulus. Therefore, a two-step hierarchical classification architecture was designed; a first classifier is trained to



(a) Temporal Evolution



(b) 2D Electrodes Representation



(c) EEG Frame Structure

Figure 1: EEG 3D Data Representation. This Figure represents the 3D data representation employed in this work. Accordingly, EEG can be considered as a sequence of frames referred to the brain activity. Each frame reports pixels indicating the voltage measured by the respective electrode on the scalp. Therefore, mapping the 2D scalp representation in a pixels grid allows to represent also the spatial information contained in EEG data.

Journal Pre-proof

Journal Pre-proof

detect single repetitions of P300. Then, a second classifier further distinguishes the ERPs referred to the detected P300, as elicited by a visual or an auditory stimulus. To the best of the authors' knowledge, this is the first time a hierarchical architecture has been leveraged in the P300 detection process. It was key to overcome the literature approaches, further distinguishing the sensorial domain of belonging for the stimuli that elicited the detected ERPs. Moreover, it provides modularity to our system. Indeed, provided that a consistent dataset is collected, additional binary splits can be included, further discriminating the eliciting stimuli. The hierarchical classifier architecture is reported in Figure 2.

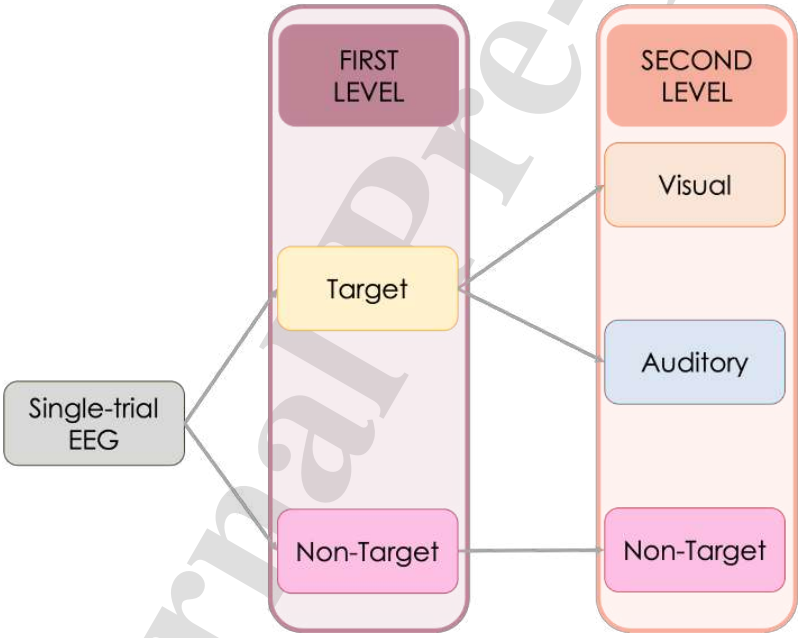


Figure 2: Hierarchical Classifier Architecture. This Figure shows the hierarchical classifier designed in this work. It allows detecting the P300 in single-trial EEG data and recognizing the sensorial domain of the eliciting stimulus.

Both spatial and temporal dependencies characterize EEG data. Despite producing an effective 3D representation to enhance its structure, leveraging

Journal Pre-proof

ad-hoc classifiers is also key. As reported in Section 1, CNN achieves optimal performances in features extraction and P300 classification, given their effectiveness in capturing spatial information. Also, LSTM proves to recognize P300, despite being capable of focusing on temporal information. Therefore, we decide to combine the two architectures, producing a CNN-LSTM network. In detail, the produced hierarchical classifier comprises two instances of CNN-LSTM networks, each corresponding to a split node. This is the first time that the combination of these two networks' structure has been used for P300 detection; however, it has been diffusely used in motor imagery recognition task achieving outstanding performances (Garcia-Moreno et al. (2020); Li et al. (2022)), given its capability in resorting both on spatial and temporal EEG dependencies. Accordingly, this network structure allows mimicking the neuroscientists procedure in detecting and recognizing the deflection caused by eliciting stimuli belonging to a different sensorial domain, enhancing the a posteriori interpretation of the classifier's decision-making process.

An instance of the employed CNN-LSTM network is reported in Figure 4. In both CNN and LSTM structures, several layers were explicitly introduced to improve network performances. Max pooling 2D and dropout layers are used to reduce data dimensionality, preventing overfitting (Srivastava et al. (2014)). Also, ReLU (Ide & Kurita (2017); Krizhevsky et al. (2012)), and batch normalization (Liu et al. (2018); Ioffe & Szegedy (2015)) layers are leveraged, given their assessed capabilities in speeding up the learning convergence and in reducing the sensitivity to initialization settings. The CNN portion structure is composed of an input layer, which passes input matrices to the convolutional layer, composed of 8 filters of size 3x3. Then, batch normalization and ReLU layers are applied. Each batch size is set to 64 samples. The activations produced by the following max pooling 2D layer are proposed as inputs to the LSTM structure, which learns the temporal data dependencies. This part of the network comprises a sequence input layer, which passes the instances to the following LSTM layer, composed of 512 units. Then a dropout of 0.5 is applied, and instances undergo another LSTM layer of 256 units. Finally, the

Journal Pre-proof

classification is performed, leveraging a linear fully connected layer that weights the input coefficients and adds a bias vector; then, a non-linear softmax layer produces the actual predictions. During the training procedure, the learning rate was set to $1e^{-4}$, and Adam's method was used as the optimizer. Accordingly, the two CNN-LSTM networks composing the hierarchical classifier were trained. The first network aims at detecting P300 in the proposed data; the second one distinguishes ERPs-related instances between elicited by a visual or an auditory stimulus. Resorting to a hierarchical structure was key to divide the classification task in two steps: P300 detection and ERP's sensorial domain recognition. This was essential for enhancing the decision-making process interpretability, allowing for inspecting the two processes separately. To provide the reader a better understanding of the proposed BCI functioning, a scheme is reported in Figure 3. Accordingly, the subject is required to focus, among the available stimuli, on the one corresponding to the intended action. Therefore, the associated ERP is elicited in the brain. The P300 is detected at the first level of the hierarchical classifier. Then, the second one recognizes the sensorial domain of the stimulus. This allows the BCI actuator to perform the action intended by the subject.

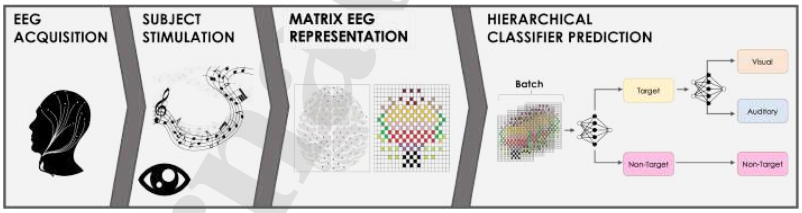


Figure 3: Presented BCI Working Pipeline. This Figure illustrates the functioning of the proposed BCI system.

3. Results

This Section first defines the metrics employed to evaluate the proposed BCI in terms of accuracy and interpretability. Then, the obtained results are

Journal Pre-proof

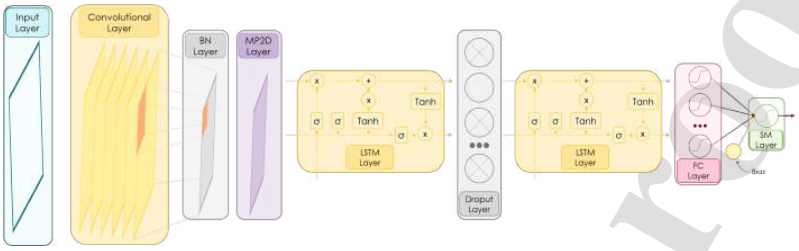


Figure 4: CNN-LSTM Network Architecture. This Figure depicts the structure of a single CNN-LSTM network. The hierarchical classifier employs two networks, one in each split note. It is composed of a series of convolutional layers, *i.e.,* convolutional, composed of 8 3x3 filters, batch normalization (BN), and max pooling 2D (MP2D), with 2 as the pool size. Then, two LSTM layers follow, of 512 and 256 units, interspersed by a dropout one. Finally, the prediction is computed through a fully connected (FC) layer and a soft-max (SM) activation function. CNN network part accounts for electrodes spatial dependencies, while LSTM for temporal ones.

presented and discussed. In detail, evaluation is performed considering two scenarios; in the first one, the system is trained and tested for each volunteer. Therefore, 22 hierarchical classifiers were trained and evaluated according to 10-fold cross-validation procedure. Instead, the second scenario involves training the hierarchical classifier on a dataset extracted according to a volunteers-based stratified procedure. Therefore, the data of all the participants are both in training and the test set. Both scenarios are key to investigate the proposed BCI performances. Indeed, the first one provides insights about the classifier efficacy in learning patterns to predict new data of the same user. On the other hand, the second scenario allows understanding whether merging data from multiple subjects enhance the classifier's predictive capabilities.

3.1. Evaluation Metrics

Each CNN-LSTM network and the overall hierarchical classifier performances are evaluated according to 10-fold cross-validation with respect to accuracy, precision, recall and F1-Score. Let true positives (TP) be the number of instances belonging to the positive class correctly predicted, and false positives (FP) the

Journal Pre-proof

Journal Pre-proof

number of incorrectly predicted ones. Accordingly, true negative (TN) and false negative (FN) can be defined for the negative class. Accuracy can be computed as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

representing the percentage of correct predictions. Precision is instead defined as the accuracy on the positive class:

$$Pre = \frac{TP}{TP + FP}, \tag{2}$$

while recall represents the percentage of instances belonging to the positive class which are correctly detected:

$$Rec = \frac{TP}{TP + FN}. \tag{3}$$

Also, F1-score is computed as the harmonic mean of precision and recall:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{4}$$

We decide to consider both F1-score and accuracy, as the first highlights, the results respect to false negatives and false positives, while the latter concerns

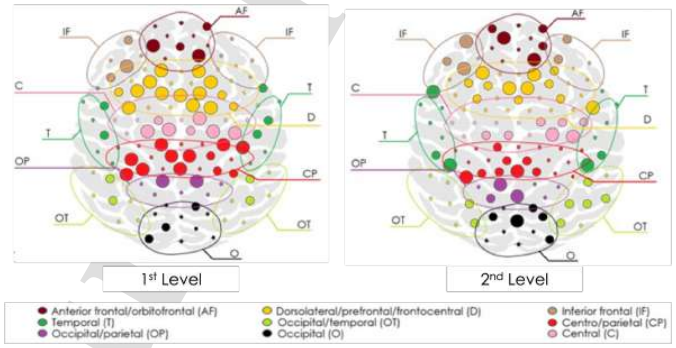


Figure 5: Features importance. This figure shows features importance estimated leveraging permutation importance approach for the P300 vs baseline classifier (left), and for the visual versus auditory ERP classifier (right). It turns out that P300 detection is a more complex task, which requires information provided by more electrodes.

Journal Pre-proof

mostly true positives and true negatives. Also, F1-Score proves to be a more robust metric when considering an unbalanced dataset.

As mentioned in Section 1, the complex network architecture complicates understanding their decision-making processes. Nevertheless, interpretability is becoming mandatory, favoring trust in machine-learning applications (Molnar (2020)), and providing additional knowledge about the investigated problem (Du et al. (2019)). In the ensemble classifiers, insights into the decision-making process are provided by reconstructing a ranking of the employed features based on the impact assumed in determining the predictions. As reported in Section 1, the most employed metric is permutation importance. According to its definition, feature importance can be estimated considering the decrease in performance caused by randomly shuffling its values. The underlying rationale is that a random permutation of a feature's values mimics its removal from the model. So, the higher the performances drop, the more the classifier relies on the considered feature values in its decision-making process. Considering EEG channels, values shuffling also causes the loss of signal temporal dependencies, forcing the signal to be meaningless. Permutation importance is computed for each CNN-LSTM network in the produced hierarchical classifier. In detail, at each split node, this metric is computed considering the CNN-LSTM network fitted on the non-shuffled features in the training set. Iteratively, a specific EEG-channel is shuffled in all the validation set instances, while the others are left unchanged.

The produced set is provided to the CNN-LSTM network, which predicts the output classes. The performances drop is estimated using the loss rather than the accuracy, being more robust. The loss is calculated using binary categorical cross-entropy, as each CNN-LSTM network splits the data in two classes:

$$Loss_i = -\frac{\sum_{k=1}^{N} y_k log\hat{y_k} + (1-y_k)log(1-\hat{y_k})}{N} \tag{5}$$

where $i = 1, ..., n_{features}$, Y collects the actual labels, $\hat{Y}$ the predictions, and N is the validation set size. Thanks to the employed 3D data structure, detailed in Section 2, each feature corresponds to an EEG channel. Therefore permutation

Journal Pre-proof

importance provides an EEG channel ranking for each network, providing in-depth information about which electrodes are most informative in detecting the P300 and recognizing the sensorial domain of the associated eliciting stimuli. This information is key to interpreting the hierarchical classifier decision-making process regarding the most involved brain regions.

Table 1: CNN-LSTM Subject-Specific Evaluation. This Table reports the performance, in terms of accuracy F1-score, recall, and precision, at each level of the hierarchical classifier with respect to 10-fold cross-validation. Training and testing were conducted according to the subject-specific scenario.

ID	Gender	1st Level				2nd Level			
		Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
1	F	82.0	81.5	73.0	92.2	95.6	95.4	96.8	94.0
2	F	81.7	81.4	75.9	87.7	89.0	82.5	99.9	70.2
3	M	78.9	78.7	81.0	76.5	78.0	75.5	75.9	75.1
4	F	85.0	84.4	82.3	86.7	97.0	96.8	94.5	99.3
5	F	92.7	91.4	84.3	99.8	98.5	97.3	95.4	99.3
6	F	84.9	84.9	81.6	87.3	95.6	95.4	96.8	94.0
7	M	72.1	68.6	87.0	56.6	99.8	98.0	97.0	99.1
8	F	86.1	85.7	77.5	95.8	87.0	85.5	81.6	89.9
9	F	89.7	89.3	86.3	92.5	95.1	95.1	96.7	93.6
10	M	77.1	76.0	76.3	75.8	83.1	92.4	89.9	95.0
11	M	92.7	92.2	86.6	98.6	98.2	97.9	97.2	98.6
12	F	89.3	87.1	78.8	97.3	96.1	92.2	98.4	86.7
13	M	77.1	77.1	80.4	74.1	89.3	86.8	84.4	89.4
14	M	82.0	80.1	79.8	80.5	94.2	93.6	88.6	99.3
15	M	75.0	74.7	71.2	78.5	99.9	99.9	99.9	99.9
16	M	79.9	79.8	71.9	89.6	98.9	98.3	98.9	97.8
17	M	87.9	83.1	73.5	95.5	96.2	91.0	86.2	96.4
18	F	95.0	92.3	87.1	98.1	98.9	98.7	99.3	98.2
19	F	89.0	84.4	80.8	88.4	99.8	99.2	98.9	99.6
20	F	84.5	81.1	75.2	87.9	88.0	87.9	85.2	90.8
21	M	76.0	75.1	71.8	78.7	83.2	82.7	82.6	82.9
22	M	89.3	85.0	79.2	91.7	99.8	99.3	99.8	98.9
Males		80.7 ± 6.6	78.1 ± 5.7	81.5 ± 11.9	79.1 ± 6.2	92.8 ± 8.0	90.9 ± 8.2	93.9 ± 8.1	92.3 ± 8.0
Females		87.3 ± 4.3	80.3 ± 4.7	92.15 ± 4.9	85.8 ± 3.9	94.6 ± 4.5	94.9 ± 6.0	92.3 ± 8.5	94.6 ± 4.5

3.2. Subject-Specific Evaluation

This evaluation scenario, involves training and testing a hierarchical classifier for each volunteer. Table 1 reports the produced results. This evaluation

Journal Pre-proof

Table 2: CNN-LSTM Intra-Subjects Evaluation. This Table reports the performance, in terms of accuracy F1-score, recall, and precision, at each level of the hierarchical classifier with respect to 10-fold cross-validation. Training and testing were conducted according to the intra-subjects scenario.

ID	Gender	1st Level				2nd Level			
		Acc	F1	Pre	Rec	Acc	F1	Pre	Rec
1	F	84.1	82.7	75.0	92.2	95.6	95.2	93.1	97.3
2	F	81.5	80.5	75.9	85.6	92.0	91.9	85.1	99.8
3	M	79.8	78.9	81.5	76.5	83.0	82.7	76.3	90.2
4	F	80.1	84.8	83.0	86.7	97.0	97.0	95.8	98.2
5	F	92.7	91.4	84.3	99.8	97.7	97.5	98.1	97.0
6	F	84.9	84.4	81.6	87.3	93.9	93.8	94.7	92.9
7	M	79.1	88.0	86.8	89.2	98.8	98.4	97.9	98.9
8	F	86.3	85.8	77.7	95.8	89.1	87.4	83.5	91.6
9	F	90.7	90.5	86.3	95.2	95.1	94.9	94.5	95.4
10	M	77.2	76.5	77.0	76.0	93.2	93.1	87.8	99.1
11	M	92.9	92.3	86.6	98.6	95.7	95.4	92.2	98.9
12	F	88.2	87.7	79.9	97.3	93.4	92.2	91.2	93.2
13	M	79.3	78.1	81.4	75.1	89.7	89.4	84.9	94.5
14	M	82.0	80.1	79.8	80.5	93.9	93.2	91.8	94.6
15	M	75.9	75.3	72.3	78.5	99.9	99.9	99.9	99.9
16	M	79.9	79.8	71.9	89.6	97.8	97.8	97.5	98.2
17	M	85.6	83.6	74.4	95.5	92.5	91.0	93.3	88.3
18	F	93.0	92.6	87.6	98.1	97.9	97.4	96.8	98.1
19	F	89.0	84.4	80.8	88.4	98.9	98.9	98.7	99.1
20	F	84.5	81.1	75.2	87.9	85.2	84.9	82.8	97.1
21	M	76.2	75.5	71.5	79.9	82.1	81.9	79.0	85.0
22	M	86.3	84.5	78.4	91.7	99.1	99.0	99.0	98.1
Males		84.0 ± 5.3	83.6 ± 5.3	79.5 ± 5.0	88.4 ± 8.4	93.7 ± 5.1	92.8 ± 5.9	90.8 ± 7.7	95.1 ± 5.0
Females		86.7 ± 4.2	86.0 ± 4.1	80.7 ± 4.4	92.2 ± 5.2	94.2 ± 4.1	94.1 ± 3.6	92.2 ± 5.8	96.3 ± 2.7

procedure aimed to investigate, considering our restricted sample, the effects of the *BCI illiteracy*. This well-known phenomenon is one of the BCIs open problems, and deals with their inapplicability for a non-negligible users segment, due to the low performances reported (Blankertz et al. (2009)).

The reported results show that, on average, the hierarchical classifier achieves 76.6% F1-Score. In detail, 82.4% is assessed in P300 detection and 82.4% in distinguishing the sensorial domain of the eliciting stimulus. Also, the overall F1-Score is higher than 70.0% for 18 of the considered subjects, which is the

Journal Pre-proof

minimum threshold required to effectively control a BCI system Yu et al. (2021). As reported in Section 1, these results can be compared to the literature only concerning the first classification level, *i.e.*, target vs. non-target. Indeed, the second classification level, *i.e.*, auditory vs. visual ERP, represents one of the innovative contributions provided by our approach, which, to the best of the authors' knowledge, has never been addressed before. Therefore, considering the target vs. non-target classification, our approach achieves performances comparable with most of the literature approaches.

Considering labels representation and available instances for each subject, it turns out that there is a strong unbalance for subjects 3, 7, 13 and 21. Indeed, even though all the volunteers undergo the same experimental protocol, the artifacts rejection phase may discard a different amount of data for each. Accordingly, we can attribute the performance drop to the smaller number of available instances that has not allowed the classifier to learn effective patterns. Also, although the average metrics assessed by the female participants are slightly higher than male ones, no statistical evidence is deducible, as the confidence boundaries overlap. So, we can conclude that our BCI system performances on a single subject mostly depend on the amount of training data provided.

3.3. Intra-Subjects Evaluation

In the second evaluation scenario, the hierarchical classifier is trained on a sub dataset extracted according to a subject-based stratified procedure, and its performances are estimated according to 10-fold cross-validation. Therefore, a single hierarchical classifier is trained, considering 70% data from each volunteer, but is tested on the remaining 30% data of a single subject per time. The obtained results are reported in Table 2.

Considering the overall performances, the hierarchical classifier assesses 78.1% F1-Score, 83.6% in detecting the P300, and 93.5% in distinguishing the sensorial domain of the eliciting stimulus. Considering gender-based results, the performances slightly increase compared to the subject-specific scenario. Also, 20 out

Journal Pre-proof

of 22 subjects achieves an F1-Score greater than 70.0%. Therefore, we can deduce that considering data collected on multiple subjects but according to the same experimental setup increases classifier learning capabilities.

Further consideration can be presented regarding the high performances achieved in recognizing the stimulus' sensorial domain once the P300 is detected. This outcome is consistent with our a priori knowledge, *i.e.,* visual and auditory ERPs have a characteristic trend that significantly differentiates them. Indeed, the differences between visual and auditory ERPs refer to the most active brain areas and their temporal evolution. While early sensorial potentials tend to be larger over occipital sites in response to visual stimuli, and to central sites in response to auditory stimuli (Regan (1989)) the topographical distribution changes over time. At P300 latency range brain dynamics moves over anterior brain regions supporting working memory, (e.g., Brunoni & Vanderhasselt (2014)), and brain potentials tend to be more negative to auditory stimuli and more positive to visual stimuli (Falkenstein et al. (1995)). It is worth noting that an expert ERP investigation analyzing the effects of stimulus category on the amplitude of electrical potentials, based on the same set of stimuli, found that the sensorial modality (visual vs. auditory) was best assessed at midline fronto/central sites (Fz and Cz) at P300 latency level (Proverbio & Tacchini (2022)). The CNN-LSTM network was ad-hoc designed for learning patterns effective in predicting, as CNN enhances spatial features and LSTM focuses on temporal patterns.

3.4. BCI Interpretability

As reported in Section 1, providing in-depth knowledge of a BCI system decision-making process is as important as the predictions themselves. Therefore, we evaluate our system also considering its interpretability, according to permutation importance. Accordingly, for each CNN-LSTM classifier, a ranking is produced based on each electrode's significance in determining the predictions. To provide reliable results, regardless of the considered subject, we compute this metric considering the hierarchical classifier trained according to

Journal Pre-proof

the second evaluation scenario. The achieved results are reported in Figure 5.

Considering the CNN-LSTM classifier aimed at detecting P300, the three most important channels are FFC3h, C2, and FC3; the three less relevant are Oz, OI2h, AFF5h. The channels mainly considered are located in the dorsolateral prefrontal region of the brain, while those neglected in the occipital area. The classifier that distinguishes the ERPs elicited by visual or auditory stimuli considers as the three most important channels F2, FCC1h, and AFp3h, which suggest the relevance of the fronto-central electrodes in revealing the sensorial modality of stimulation, at P300 sensorial stage and with these types of stimuli. In a recent BCI study involving purely visual, purely auditory or audio-visual stimulation, the best accuracy for P300 classification was instead found at Pz site, but no frontal electrode was used in that study, so that it cannot be excluded a more anterior distribution for the P300 potential. Again, in an auditory-tactile BCI study it was found that the best site for detecting P300 was indeed Fz electrode (Brouwer & Van Erp (2010)).

The most crucial electrodes are located in the frontal anterior and dorsolateral prefrontal area. In the dorsolateral prefrontal zone, few of the available channels are considered. This proves the CNN dimensionality reduction capabilities, as it discharges redundant channels. Also, it can be noticed that channels importance is more diffuse in the P300 detection classifier, while a smaller number of channels is considered in distinguishing the ERPs eliciting stimuli sensorial domain. This result further supports the conclusion that it is easier to classify the ERPs than to detect them, given the distinctive trend that characterizes responses elicited by stimuli belonging to different sensorial domains. Instead, to discriminate the sensorial domain of the stimulus, a reduced number of channels is considered, as the classifier needs to recognize only the characteristic spatio-temporal patterns of the observed ERPs.

Journal Pre-proof

4. Conclusion

In this work, a novel BCI is presented. In addition to detect P300 in a single-trial EEG, our system can also recognize the sensorial domain of the stimulus eliciting the considered ERP. The predictions are achieved leveraging a two-step hierarchical classifier specifically designed. A CNN-LSTM network performs the predictions at each split node, accounting for both EEG spatial and temporal dependencies. Networks' capabilities are also enhanced by leveraging a suitable 3D data representation. The proposed system is validated on real data, acquired according to an ad-hoc defined experimental setup. The optimal results obtained in terms of accuracy and interpretability pave the way for more sophisticated BCI, capable of providing more actions to their users. Of course, future works deal with online testing our method performances in terms of accuracy and information transfer rate. Also, we aim at performing a sensitivity analysis on the number of EEG channels to identify the best trade-off between user comfort and BCI system accuracy.

References

Allison, B. Z., Kübler, A., & Jin, J. (2020). 30+ years of p300 brain–computer interfaces. *Psychophysiology*, *57*, e13569. doi: 10.1111/psyp.13569.

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*, 1340–1347. doi: 10.1093/bioinformatics/btq134.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of erp components—a tutorial. *NeuroImage*, *56*, 814–825. doi: 10.1016/j.neuroimage.2010.06.048.

Blankertz, B., Sanelli, C., Halder, S., Hammer, E., Kübler, A., Müller, K.-R., Curio, G., & Dickhaus, T. (2009). Predicting bci performance to study bci illiteracy. *BMC Neurosci*, *10*, P84. doi: 10.1186/1471-2202-10-s1-p84.

Journal Pre-proof

Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L. E., Sturm, I., Curio, G. et al. (2010). The berlin brain–computer interface: non-medical uses of bci technology. *Frontiers in neuroscience*, *4*, 198. doi: 10.3389/fnins.2010.00198.

Brouwer, A.-M., & Van Erp, J. B. (2010). A tactile p300 brain-computer interface. *Frontiers in neuroscience*, *4*, 19. doi: 10.3389/fnins.2010.00019.

Brunoni, A. R., & Vanderhasselt, M.-A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: a systematic review and meta-analysis. *Brain and cognition*, *86*, 1–9. doi: 10.1016/j.bandc.2014.01.008.

Carabez, E., Sugi, M., Nambu, I., & Wada, Y. (2017). Convolutional neural networks with 3d input for p300 identification in auditory brain-computer interfaces. *Computational intelligence and neuroscience*, *2017*. doi: 10.1155/2017/8163949.

Cecotti, H., & Graser, A. (2010). Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, *33*, 433–445.

Chan, A. T., Quiroz, J. C., Dascalu, S., & Harris, F. C. (2015). An overview of brain computer interfaces. In *Proc. 30th Int. Conf. on Computers and Their Applications*.

De Venuto, D., & Mezzina, G. (2021). A single-trial p300 detector based on symbolized eeg and autoencoded-(1d) cnn to improve itr performance in bcis. *Sensors*, *21*, 3961. doi: 10.3390/s21123961.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*, 200–227. doi: 10.1016/j.neuron.2011.03.018.

Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, *63*, 68–77. doi: 10.1145/3359786.

Journal Pre-proof

Falkenstein, M., Koshlykova, N., Kiroj, V., Hoormann, J., & Hohnsbein, J. (1995). Late erp components in visual and auditory go/nogo tasks. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *96*, 36–43. doi: 10.1016/0013-4694(94)00182-K.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, *70*, 510–523. doi: 10.1016/0013-4694(88)90149-6.

Fedjaev, J. (2017). Decoding eeg brain signals using recurrent neural networks. *Technical University of Munich, Master Thesis*, .

Fonken, Y. M., Kam, J. W., & Knight, R. T. (2020). A differential role for human hippocampus in novelty and contextual processing: Implications for p300. *Psychophysiology*, *57*, e13400. doi: 10.1111/psyp.13400.

Garcia-Moreno, F. M., Bermudez-Edo, M., Rodríguez-Fórtiz, M. J., & Garrido, J. L. (2020). A cnn-lstm deep learning classifier for motor imagery eeg detection using a low-invasive and low-cost bci headband. In *2020 16th International Conference on Intelligent Environments (IE)* (pp. 84–91). IEEE. doi: 10.1109/IE49459.2020.9155016.

Herron, J. E., Quayle, A. H., & Rugg, M. D. (2003). Probability effects on event-related potential correlates of recognition memory. *Cognitive Brain Research*, *16*, 66–73. doi: 10.1016/S0926-6410(02)00220-3.

Huang, D., Lin, P., Fei, D.-Y., Chen, X., & Bai, O. (2009). Decoding human motor activity from eeg single trials for a discrete two-dimensional cursor control. *Journal of neural engineering*, *6*, 046005. doi: 10.1088/1741-2560/6/4/046005.

Huang, D., Qian, K., Fei, D.-Y., Jia, W., Chen, X., & Bai, O. (2012). Electroencephalography (eeg)-based brain-computer interface (bci): A 2-d virtual wheelchair control based on event-related desynchronization/synchronization

Journal Pre-proof

and state control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*, 379–388. doi: 10.1109/TNSRE.2012.2190299.

Ide, H., & Kurita, T. (2017). Improvement of learning for cnn with relu activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 2684–2691). IEEE. doi: 10.1109/IJCNN.2017.7966185.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456). PMLR.

Joshi, R., Goel, P., Sur, M., & Murthy, H. A. (2018). Single trial p300 classification using convolutional lstm and deep learning ensembles method. In *International Conference on Intelligent Human Computer Interaction* (pp. 3–15). Springer. doi: 10.1007/978-3-030-04021-5_1.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Kshirsagar, G. B., & Londhe, N. D. (2020). Weighted ensemble of deep convolution neural networks for single-trial character detection in devanagari-script-based p300 speller. *IEEE Transactions on Cognitive and Developmental Systems*, *12*, 551–560. doi:10.1109/TCDS.2019.2942437.

Lelievre, Y., Washizawa, Y., & Rutkowski, T. M. (2013). Single trial bci classification accuracy improvement for the novel virtual sound movement-based spatial auditory paradigm. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (pp. 1–6). IEEE. doi: 10.1109/APSIPA.2013.6694317.

Li, H., Ding, M., Zhang, R., & Xiu, C. (2022). Motor imagery eeg classification algorithm based on cnn-lstm feature fusion network. *Biomedical Signal Processing and Control*, *72*, 103342.

Journal Pre-proof

Journal Pre-proof

Li, K., Sankar, R., Arbel, Y., & Donchin, E. (2009). Single trial independent component analysis for p300 bci system. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4035–4038). IEEE.

Linden, D. E. (2005). The p300: where in the brain is it produced and what does it tell us? *The Neuroscientist*, *11*, 563–576. doi: 10.1177/1073858405280524.

Liu, M., Wu, W., Gu, Z., Yu, Z., Qi, F., & Li, Y. (2018). Deep learning based on batch normalization for p300 signal detection. *Neurocomputing*, *275*, 288–297. doi: 10.1016/j.neucom.2017.08.039.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, *15*, 031005. doi: 10.1088/1741-2552/aab2f2.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press. doi: 10.1016/0028-3932(71)90067-4.

Malar, E., Gauthaam, M., & Chakravarthy, D. (2011). A novel approach for the detection of drunken driving using the power spectral density analysis of eeg. *International Journal of Computer Applications*, *21*, 10–14. doi: 10.5120/2525-3436.

Mason, S. G., Bashashati, A., Fatourechi, M., Navarro, K. F., & Birch, G. E. (2007). A comprehensive survey of brain interface technology designs. *Annals of biomedical engineering*, *35*, 137–169. doi: 10.1007/s10439-006-9170-0.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *Journal of neuroscience methods*, *167*, 82–90. doi: 10.1016/j.jneumeth.2007.09.022.

Journal Pre-proof

Nakanishi, I., Baba, S., Ozaki, K., & Li, S. (2013). Using brain waves as transparent biometrics for on-demand driver authentication. *International journal of biometrics*, *5*, 288–305. doi: 10.1504/IJBM.2013.055965.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, *9*, 97–113. doi: 10.1016/0028-3932(71)90067-4.

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, *112*, 713–719. doi: 10.1016/S1388-2457(00)00527-7.

OpenBCI (). Publicly available eeg datasets. https://openbci.com/community/publicly-available-eeg-datasets/. Accessed: 2022-06-02.

Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, *118*, 2128–2148. doi: 10.1016/j.clinph.2007.04.019.

Polich, J. (2012). Neuropsychology of p300. *The Oxford handbook of event-related potential components*, . doi: 10.1093/oxfordhb/9780195374148.013.0089.

Polich, J. (2020). 50+ years of p300: Where are we now? *Psychophysiology*, *57*, e13616–e13616.

Proverbio, A. M., Adorni, R., & D'Aniello, G. E. (2011). 250 ms to code for action affordance during observation of manipulable objects. *Neuropsychologia*, *49*, 2711–2717. doi: 10.1016/j.neuropsychologia.2011.05.019.

Proverbio, A. M., Camporeale, E., & Brusa, A. (2020). Multimodal recognition of emotions in music and facial expressions. *Frontiers in human neuroscience*, *14*, 32. doi: 10.3389/fnhum.2020.00032.

Proverbio, A. M., & Tacchini, M. (2022). Erp markers of visual and auditory perception: a useful tool for bci systems. *Journal of Neural Engineering*, .

Journal Pre-proof

Proverbio, A. M., & Zani, A. (2003). Electromagnetic manifestations of mind and brain. In *The cognitive electrophysiology of mind and brain* (pp. 13–40). Elsevier.

Regan, D. (1989). Human brain electrophysiology. *Evoked potentials and evoked magnetic fields in science and medicine*, .

Royer, A. S., Doud, A. J., Rose, M. L., & He, B. (2010). Eeg control of a virtual helicopter in 3-dimensional space using intelligent control strategies. *IEEE Transactions on neural systems and rehabilitation engineering*, *18*, 581–589. doi: 10.1109/TNSRE.2010.2077654.

Schomer, D. L., & Da Silva, F. L. (2012). *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins. doi: 10.1086/413356.

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*, e9813. doi: 10.1371/journal.pone.0009813.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*, 1929–1958.

Tal, O., & Friedman, D. (2019). Recurrent neural networks for p300-based bci. *arXiv preprint arXiv:1901.10798*, .

Vareka, L. (2021). Comparison of convolutional and recurrent neural networks for the p300 detection. In *BIOSIGNALS* (pp. 186–191). doi: 10.5220/0010248201860191.

Volosyak, I., Cecotti, H., & Graser, A. (2009). Optimal visual stimuli on lcd screens for ssvep based brain-computer interfaces. In *2009 4th International IEEE/EMBS Conference on Neural Engineering* (pp. 447–450). IEEE. doi: 10.1109/NER.2009.5109329.

Journal Pre-proof

Journal Pre-proof

Waibel, M. (2011). Braindriver: A mind controlled car. *IEEE Spectrum*, .

Wirth, C., Toth, J., & Arvaneh, M. (2020). "you have reached your destination": A single trial eeg classification study. *Frontiers in neuroscience*, *14*, 66.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical neurophysiology*, *113*, 767–791. doi: 10.1016/S1388-2457(02)00057-3.

Xiao, X., Xu, M., Wang, Y., Jung, T.-P., & Ming, D. (2019). A comparison of classification methods for recognizing single-trial p300 in brain-computer interfaces. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3032–3035). IEEE. doi: 10.1109/EMBC.2019.8857521.

Yu, X., Aziz, M. Z., Sadiq, M. T., Fan, Z., & Xiao, G. (2021). A new framework for automatic detection of motor and mental imagery eeg signals for robust bci systems. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–12.

Zani, A., & Proverbio, A. M. (2003). Cognitive electrophysiology of mind and brain. In *The cognitive electrophysiology of mind and brain* (pp. 3–12). Elsevier.

Zhu, D., Bieger, J., Garcia Molina, G., & Aarts, R. M. (2010). A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, *2010*. doi: 10.1155/2010/702357.

Journal Pre-proof

Journal Pre-proof

HIGHLIGHTS

- Identify ERP deflection in single trial EEG reduces BCI response time;
- Matricial EEG representation simultaneously express spatio-temporal information;
- A CNN-LSTM network detect the ERP and map the stimuli to their sensorial domain;
- Auditory and visual ERPs are distinguished, increasing BCI potentials.

Journal Pre-proof

Jessica Leoni: Conceptualization, Methodology, Software, Formal Analysis, Writing – Original Draft Preparation, Visualization.

Mara Tanelli Conceptualization, Methodology, Resources, Writing – Review & Editing, Visualization, Supervision, Project Administration.

Silvia Carla Strada: Conceptualization, Resources, Writing – Review & Editing.

Alessandra Brusa: Data Curation, Investigation, Validation.

Alice Mado Proverbio: Conceptualization, Resources, Writing-Reviewing and Editing, Supervision.

Journal Pre-proof

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof