

A novel BCI-SSVEP based approach for control of walking in Virtual Environment using a Convolutional Neural Network

Vitoantonio Bevilacqua
Giacomo Tattoli
and Domenico Buongiorno
Department of Electrical
and Information Engineering (DEI)
Polytechnic of Bari, Bari, Italy

Claudio Loconsole
Daniele Leonardis
Michele Barsotti
Antonio Frisoli
and Massimo Bergamasco
PERCRO Laboratory, TECIP Institute
Scuola Superiore Sant'Anna, Pisa, Italy

Abstract—A non-invasive Brain Computer Interface (BCI) based on a Convolutional Neural Network (CNN) is presented as a novel approach for navigation in Virtual Environment (VE). The developed navigation control interface relies on Steady State Visually Evoked Potentials (SSVEP), whose features are discriminated in real time in the electroencephalographic (EEG) data by means of the CNN. The proposed approach has been evaluated through navigation by walking in an immersive and plausible virtual environment (VE), thus enhancing the involvement of the participant and his perception of the VE. Results show that the BCI based on a CNN can be profitably applied for decoding SSVEP features in navigation scenarios, where a reduced number of commands needs to be reliably and rapidly selected. The participant was able to accomplish a waypoint walking task within the VE, by controlling navigation through of the only brain activity.

I. INTRODUCTION

Brain Computer Interface (BCI) represents one of the few Human-Machine Interaction (HMI) techniques usable by people with severe disabilities (e.g., amyotrophic lateral sclerosis or brainstem stroke) [1]. Through the use of non-invasive electroencephalography (EEG) and processing of features extracted from the EEG signal, it is possible to recognize the intention of a user performing a selection task among different presented options. Thus the method can be applied to several specific tasks, such as the control of assistive devices (e.g. speller [2], an electrical prosthesis [3] or even a robot [4]).

The existing EEG-based BCI designs rely on a variety of EEG signal features including slow cortical potentials [5], oscillatory activity [6], P300 potentials [7], motor-related potentials [8] and visually evoked potentials (VEPs) [9]–[11].

One BCI solution, with successful performance in terms of optimizing both speed and accuracy, relies on an involuntary response known as the Steady-State Visual Evoked Potential (SSVEP). This is a periodic response elicited by the repetitive presentation of a visual stimulus, at a rate of 68 Hz or more [12]. The available frequency range for the visual stimuli affects the number of SSVEP features that, in general, can be discriminated by a BCI, due to frequency dispersion and superimposition of harmonics. According to Herrmann [13], the range of frequencies capable of generating a brain response is 6100 Hz. However, other works, such as the interesting

review by Zhu et al. [14], affirm that stronger responses can be achieved for the lower frequencies (from 12 to 25 Hz) even though this range can cause seizures in photosensitive epilepsy subjects [15]. This conclusion is confirmed by the research work conducted in [16] for higher frequency SSVEP experiments. Higher frequencies, over 40 Hz, which are unnoticeable to the human eye, implies weaker brain response of the subject, resulting in a harder detection of the intention.

One of the interesting field where BCIs have been studied is the control in navigation tasks. Such paradigms can be applied for the control of physical devices, such as wheelchairs [17], by people affected by severe disabilities.

Navigation paradigms controlled by BCI have been applied also in Virtual Environments (VEs), with the final aim of providing both a more direct control over navigation and usability by inert participants. An interesting contribution in SSVEP-controlled navigation tasks in VE is that by Legny et al [18], where the authors target navigation in VE using SSVEP and Brain-Computer Interfaces. They studied the navigation task in an outdoor environment evaluating the effect of the real-time visual feedback of the mental activity. The VE scenario was conveyed to the user through a 60 Hz LCD screen and the classification of the signals acquired using SSVEP techniques was performed using three Linear Discriminant Analysis (LDA) classifiers, each for a specific navigation command.

Neural networks are used in many field of research such as on-line torque prediction and control of robot joints using surface electromyography [19], compute kinematic and control of a prosthesis [20], automatic heartbeat classification [21], medical diagnosis [22], modelling injection system [23], breast cancer classification problem [24] and much more.

In this paper, we present a novel BCI-SSVEP approach for walking control in VEs. Our novel paradigm proposes the use of a simplified (with respect to that proposed in [25]), but high performance Convolutional Neural Network for classifying the EEG signals, allowing an enhanced control of navigation. We evaluate the performance and usability of the system in a navigation task within a highly immersive VE, through the use of an enriched virtual scenario representing a realistic virtual city and a high definition stereoscopic 3D head mounted display (HMD) with tracking of the rotation of the head. The

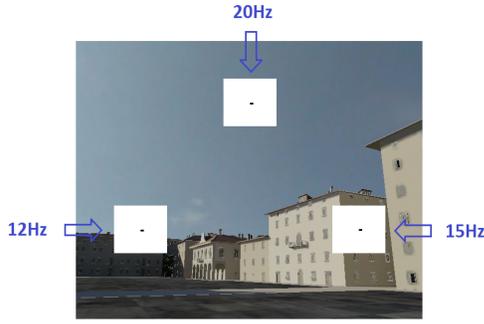


Fig. 1. The Virtual Environment scenario and the flickering symbols (rectangles) used for eliciting the SSVEP response. The flickering frequencies are indicated: 12, 15 and 20 Hz.

level of immersivity and realism of the VE is one of the novelties introduced by the proposed experiment. Immersivity and realism are expected to better exploit the SSVEP approach in navigation in presence of plausible factors related to the participant, in particular the involvement and the focusing of attention, as well as the perception of position and speed of the virtual body within the VE.

II. SYSTEM DESCRIPTION

A. System overview

The proposed system allows to control the navigation (walking) into an immersive virtual environment (VE) directly by means of brain activity using a non-invasive EEG BCI based on SSVEP elicitation and a CNN for the classification. The virtual scenario, and the generation of visual stimuli for evoking the SSVEP are managed by a dedicated XVR (eXtreme Virtual Reality <http://www.vrmedia.it/en.html> [26], [27]) software module, and provided to the user by an Oculus Rift device (<http://www.oculusvr.com/>), obtaining a highly immersive environment, conveyed to the participant by means of the HMD.

The proposed SSVEP navigation paradigm is based on a four classes selection of three navigation commands plus a rest state, respectively for moving forward, turning left, turning right and stopping the walking. The participant is intended to select one of the navigation commands by focusing attention on one of the three different symbols superposed over the presented virtual scenario (see Fig. 1). The flashing of the symbols at different frequencies evokes SSVEP, that is detectable in the acquired EEG signals. The EEG signals are processed in real time by a BCI classification algorithm. The output of the BCI represents the navigation command currently selected by the participant. The navigation command is received by the virtual environment manager and is used for properly moving the avatar into the virtual environment.

Thus, the proposed system can be divided in three modules:

- the EEG acquisition and pre-processing module;
- the CNN processing module;
- the Virtual environment manager module.

For sake of clarity, the flow diagram of the proposed system is reported in Fig. 2. A detailed description of each module is reported in the following subsections.

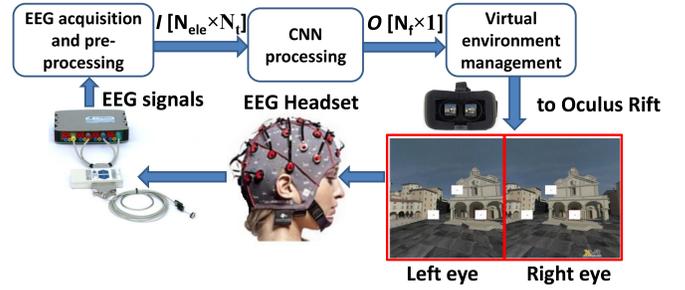


Fig. 2. The block diagram of the proposed BCI-SSVEP based approach.

1) The EEG acquisition and pre-processing module:

The EEG acquisition module is composed of an EEG headset with five active electrodes and a g.tec amplifier and analog to digital converter (g.USBamp, Guger Technologies, Schiedlberg, Austria - <http://www.gtec.at/>). The electrodes are placed over the occipital area in proximity of the visual cortex as shown in Fig. 3, according to the standard 10-20 positioning (Cz, Pz, PO3, PO4, Oz). Channels are referenced to AFz [28] and the amplifier is grounded to the earlobe.

A band-pass filter in the range 2-60 Hz and a notch-filter at 50 Hz were applied internally to the amplifier with the purpose of limiting the presence of artifacts and noise in the EEG signals. Then, signals were sampled and digitally converted at a frequency of $F_s = 256$ Hz.

In order to pre-process the raw EEG signals as input to the CNN classifier, we considered $N_{ele} = 4$ signals given by the Pz, PO3, PO4, Oz electrodes, configured in bipolar mode with respect to Cz. Each signal was normalized for obtaining zero mean and unitary variance on the basis of a moving window (with window duration $T_s = 2$ s and 75 % overlapping).

It results that the input of the subsequent CNN classifier is a matrix I of size $N_{ele} N_t$ (4×512 elements), where N_t is the number of samples in a window that are used for the analysis: $N_t = F_s \times T_s$ ($256 \text{ Hz} \times 2 \text{ s}$).

2) The CNN module:

The classifier proposed in this paper, that allows to classify the acquired SSVEP signals, is a simplification of the Convolutional Neural Network (CNN) classifier presented in [25]. The CNN topology in [25] has demonstrated to represent a suitable solution for the classification of SSVEP responses. In fact, in the same work, the authors reported a high mean recognition rate of 95.61 % about the classification of five different types of SSVEP responses. Besides the classification features of the CNN, the main issue solved in [25] is the ability of the CNN to perform the detection of SSVEP response features directly in the frequency domain without requiring a custom pre-processing of the data acquired in the time domain.

However, the CNN proposed in this paper simplified the network presented in [25] avoiding the use of the time filter hidden layer after layer L1.

Moreover, our FFT hidden layer (L2) analyzes only 12, 15 and 20 Hz frequencies, whereas in [25], the authors considered also two multiple harmonics of the base frequencies, 6.66, 7.5, 8.57, 10 and 12 Hz (different from ours) to obtain the classification.

Before introducing our CNN topology, it is necessary to

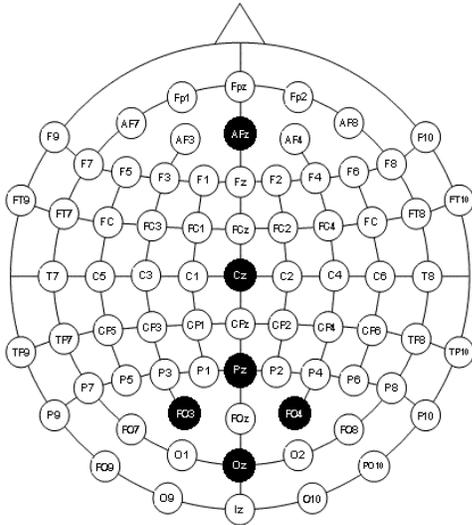


Fig. 3. The pattern of electrodes represented in the standardized International 10-20 system for EEG that describes the location of the scalp positions used for EEG applications.

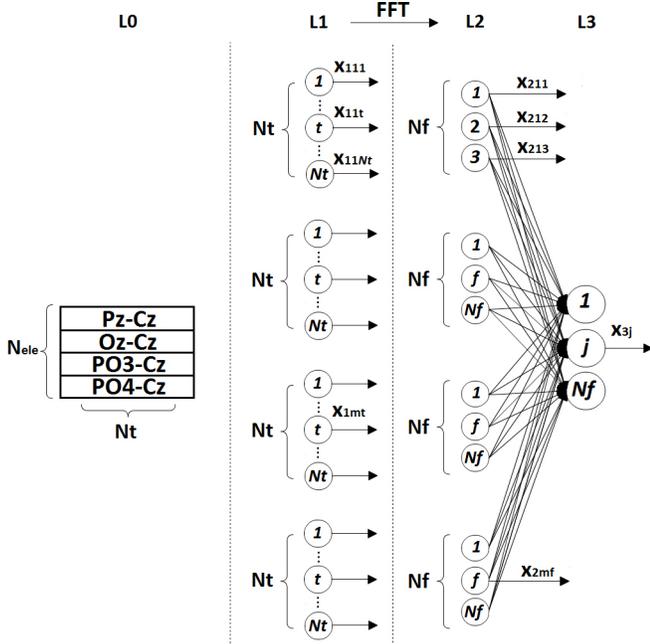


Fig. 4. The proposed Convolutional Neural Network is shown. The network acquires, as input, the EEG signals from the electrodes and computes the best spatial filter to optimize the frequency response. The CNN provides as output the classification of the signal that corresponds to the indication of which symbol the user is focusing on.

point out the concepts of map and virtual electrode. A map is a group of neurons belonging to the same layer and featuring the same weights and biases. A map is defined as a layer entity with a specific semantic. The first hidden layer is composed of a set of maps, where each map represents a virtual electrode.

Therefore, a virtual electrode is a combination of Nele electrodes, resulting in a spatial filtering. The goal of the creation of a virtual electrode is to enhance a particular information that is contained in the signal. Furthermore, some electrodes may contain the same kind of noise. So, through the combined analysis of the electrodes, some noise can be suppressed.

Usually different spatial filters are created to perform the classification, each of them corresponding to a different channel. We distinguish the electrode inputs, which are a specific case of channel and channels that represent a combination of the electrodes, i.e. virtual electrodes.

The CNN developed in this work is composed of four layers (see Fig. 4):

- the input layer, Layer 0 (L0): it takes as input the matrix I of size $Nele \times N_t$;
- two hidden layers:
 - Layer 1 (L1): the first hidden layer is used for creating the different virtual electrodes and is composed of Nele maps. Each map of L1 is N_t sized. This layer corresponds to have Nele virtual electrodes and can suppress the artifacts mainly due to the phase difference between the signals recorded by two relatively close electrodes;
 - Layer 2 (L2): the second hidden layer is composed of Nele maps. Each map of L2 has N_f neurons, where N_f is the number of frequencies to be detected;
- the output layer, Layer 3 (L3): this layer has only a map of N_f neurons, which represents the N_f frequencies to be detected. This map is fully connected to the maps belonging to L2.

It is worth to notice that the weights used in the spatial filter (that is the creation of the virtual electrodes) can be fixed or adaptive. In this work, the weights are fixed during the experiments, but adjusted within the CNN training routine. In this way, it is possible to regulate the weights according to the specific subject for a specific SSVEP frequency [29]. Next sections are dedicated to introduce the formulas needed for the CNN propagation of the input data and for the CNN error due to backpropagation.

a) CNN propagation:

In this section, the formulas that allow to calculate the propagation of the input data to the output of the CNN are introduced.

Generally, in the following we will indicate with σ , $f()$ and $x=f(\sigma)$ the weighted sum of the inputs plus the bias, the activation function and the output of a single neuron, respectively.

Layer 1:

The goal of the first hidden layer (L1) is to create Nele virtual electrodes, so it is composed of $M = Nele$ maps.

Each map contains N_t neurons sharing the same weights and biases.

Considering in the following equations the notation: σ_{1mt} , to indicate the weighted sum plus the bias of the inputs for the neuron t of the map m ; w_{1me} for the weight of the map m for the electrode e ; I_{et} for the sample t of the signal coming from the virtual electrode e ; w_{1m0} for the bias for neurons of the map m , we obtain:

$$\sigma_{1mt} = \sum_{e=1}^{N_{ele}} w_{1me} \cdot I_{et} + w_{1m0}$$

$$x_{1mt} = f_1(\sigma_{1mt}) = \tanh(\sigma_{1mt})$$

Layer 2:

The goal of the second hidden layer (L2) is to transform the domain of the input data: from time domain to frequency domain. This layer is composed of M maps, where each map contains N_f neurons.

We indicate with $X_m(t)$ the M -th spatially filtered signal. Each neuron of the same maps calculates the fast Fourier transform (FFT) $Y_m(u)$ of $X_m(t)$ at a specific frequency. $Y_m(u)$ is based on $X_m(t)$ with N_{FFT} points by using zero padding technique. In our case N_{FFT} is set to 1024. The values $Y_m(S(f))$ are only computed for $S(f)$ with $1 \leq f \leq N_f$. Indeed, the phase of the transformed signal (Θ_{2mf}) is kept fixed to reconstruct the signal in the time domain during the backpropagation, obtaining:

$$Y_m(u) = \frac{1}{N_{FFT}} \sum_{t=1}^{N_t} x_{1mt} \cdot e^{-\frac{i2\pi}{N_{FFT}} \cdot u(t-1)}$$

$$x_{2mf} = |Y_m(S(f))| \quad \Theta_{2mf} = \angle Y_m(S(f))$$

where X_{2mf} is the output value of the neuron f in the map m .

Neurons in the same position in different maps calculate the FFT for a given frequency (one among the selected frequencies; in our case they are three: 12, 15 and 20 Hz).

Layer 3:

The output layer (L3) has N_f neurons corresponding to the number of frequencies to be detected (three in the proposed case). It is fully connected with the M maps of the second hidden layer L2. The output value of each neurons is bounded in the range $[0, 1]$. Hence, the goal of this layer is to recognize which navigation commands the user is focusing on (or if the user is in the rest state). For this aim, the output of this layer is a vector O of size N_f . Each element of the vector, corresponding to a particular navigation command (indicated with C_i), contributes to the final frequency selection of the CNN according to this function:

$$O = \begin{cases} C_i & \text{if } \begin{cases} \max(O_k) = O_i & 1 \leq k \leq N_f \\ O_i > th_{CNN} \end{cases} \\ \text{rest state} & \text{if } \max O_i \leq th_{CNN} \end{cases}$$

where $1 \leq i \leq N_f$, and th_{CNN} represents an experimentally evaluated threshold used to classify stimulation classes from the rest.

Considering the following equations, the notation σ_{3j} us indicates the weighted sum plus the bias of the inputs for

the neuron j ; w_{3mfj} to indicate the weight between the neuron f of the map m (in L2) and the neuron j of the layer L3 (output layer); w_{3j0} is the bias of the neuron j of the output layer. It is possible to obtain:

$$\sigma_{3j} = \sum_{m=1}^M \sum_{f=1}^{N_f} w_{3mfj} \cdot x_{2mf} + w_{3j0}$$

$$x_{3j} = f_3(\sigma_{3j}) = \frac{1}{1+e^{-\sigma_{3j}}}$$

b) CNN backpropagation:

In this section, the CNN backpropagation error routine is described. The rules for weight updating of each layer are shown. The weights are corrected through a gradient descend by minimizing the least mean square error. The training routine is stopped in case the limit of 1000 epochs is reached or in case the error is increased over a span of 50 epochs.

Layer 3:

The updating of the weights (belonging to the neurons of the output layer) follows the classical backpropagation rule. For each neuron j in the output layer (L3) using the following notation: δ_{3j} is the error of the neuron j , $f'_3(\sigma)$ is the first derivative of the activation function $f_3(\sigma)$ with respect to σ evaluated in σ_{3j} and γ_3 is the learning rate, it is possible to write:

$$\delta_{3j} = e_j \cdot f'_3(\sigma_{3j}) = (x_{3j} - x_{target,j}) \cdot f'_3(\sigma_{3j})$$

where $X_{target,j}$ is the expected value for the neuron j , whereas the other terms are equal to:

$$f_3(\sigma_{3j}) = \frac{1}{1+e^{-\sigma_{3j}}} = x_{3j}$$

$$f'_3(\sigma_{3j}) = x_{3j} \cdot (1 - x_{3j})$$

$$\Delta w_{3mfj} = -\gamma_3 \cdot \delta_{3j} \cdot x_{2mf}$$

$$w_{3mfj}(p+1) = w_{3mfj}(p) + \Delta w_{3mfj}$$

where Δw_{3mfj} is the variation of the weight w_{3mfj} and p indicates the step number of the backpropagation routine.

Layer 2:

Indicating with δ_{2mf} the error of the neuron f of the map m , it is possible to write:

$$\delta_{2mf} = \sum_{j=1}^{N_3} w_{3mfj} \cdot \delta_{3j}$$

Layer 1:

The error calculated for the layer 2 has to be transformed back in the time domain (from the frequency domain). For this aim, we use the Inverse Fast Fourier Transform (IFFT) in order to calculate the weights of the first hidden layer.

Indicating with Z_{2mf} the IFFT of the error committed for the neuron f in map m and with γ_3 the learning rate, we obtain:

$$Z_{2mf} = \delta_{2mf} \cdot e^{i\theta_{2mf}}$$

$$\delta_{1mt} = \left\{ \Re \sum_{f=1}^{N_f} Z_{2mf} \cdot e^{-\frac{i2\pi}{N_{FFT}} \cdot S(f) \cdot (t-1)} \right\}$$

$$\Delta w_{1met} = -\gamma_1 \cdot \delta_{1mt} \cdot I_{et}$$

where w_{1met} is the variation of the weight w_{1met} . The weight w_{1met} is a candidate weight that links neuron t of the map m to the sample t of the signal from the electrode e , considering that the weights of a spatial filter are time independent. Hence, we have to calculate the average Δw_{1met} in the time domain:

$$\Delta w_{1me} = \frac{1}{N_t} \sum_{t=1}^{N_t} \Delta w_{1met}$$

$$w_{1me}(p+1) = w_{1met}(p) + \Delta w_{1me}$$

where w_{1me} is the variation of weight w_{1me} and p is the step number of the back propagation routine.

3) The virtual environment manager:

The virtual environment manager has two objectives:

- 1) providing the navigation interface, including the generation of visual stimuli for evoking the SSVEP;
- 2) applying the navigation commands selected by the participant into the rendered VE.

The navigation interface is composed by three symbols, each one related to one of the following navigation commands “walking forward”, “turn left” and “turn right”. Each symbol was represented by a 2D rectangle over impressed to the visual representation of the 3D VE.

The flickering frequency for each symbol is selected considering the constraint of the 60 Hz refresh rate of the head mounted display. The frame frequency has to be an integer multiple of each stimulation frequency, as shown in the following equation:

$$\frac{ScreenFreq}{StimFrames} = StimFreq$$

where *ScreenFreq* is the frame rate of the screen, and *StimFrame* is the number of frames required for the repetition of a single flash. The resulting available frequencies are shown in Table I. The selected stimulation frequencies for the three symbols are respectively: 12, 15 and 20 Hz, and the flickering is between solid white and solid black with a duty cycle of 50% rounded to the nearest integer number of frames. A stationary black dot was added in the middle of each symbol with the purpose of easing the fixation by the participant.

In order to apply the navigation control in the VE, the VE manager receives the current BCI output, corresponding to one of the following classes: “walk forward”, “turn left”, “turn right” and “rest”.

TABLE I.

StimFrames	StimFreq
3	20 Hz
4	15 Hz
5	12 Hz
6	10 Hz
7	8.5714 Hz
8	7.5 Hz
9	6.6667 Hz

Available stimulus frequencies according to the constraint imposed by the screen frame rate of 60Hz.

The avatar body in the VE is moved forward with a linear velocity of 1.6 m/s (related to objects in the virtual environments) as long as the walk forward command is selected. Similarly, it is rotated with an angular velocity of

0.5 rad/s as long as the turn left or turn right command are selected respectively. When the stop command is selected (corresponding to no symbol focused by the participant), the linear and angular velocities of the avatar body are set to zero. In order to avoid abrupt movement and stopping of the avatar body, a low pass filter (time constant 0.1 s) is applied to the imposed linear and angular velocities.

Once the movement of the avatar body is computed on the basis of the BCI output, the VE scenario is rendered from the point of view of the virtual avatar. With the aim of reproducing a plausible walking task, an immersive and realistic virtual scenario was provided to the participant. The implemented 3D model was a virtual reconstruction of a real city area (the main square of Livorno, Italy) allowing for extended navigation tasks in presence of detailed objects such as buildings, roads and other urban components.

To convey immersion, the VE was provided by means of a head mounted display, performing stereo vision, and tracking of the orientation of the head (Oculus Rift, 640×800 resolution for each eye, 90° horizontal FOV, frame rate 60Hz).

The whole virtual environment manager including the VE rendering was developed using XVR software.

III. EXPERIMENTAL DESCRIPTION AND RESULTS

In order to validate the proposed approach, we conducted two experiments: the first aimed to evaluate the offline performance of the CNN, fed with EEG signals acquired during a training and test session; the second aimed to evaluate the whole proposed method applied to a navigation task in the VE.

Both experiments were performed by four healthy male participants aged between 20 and 28, with no previous experience in BCI systems. The two experimental sessions took place in the same day.

A. CNN performance validation

The training and test experimental session for evaluating the CNN performance was conducted as follows. The participant sat on a comfortable chair, wearing the HMD (Oculus rift) and the cap with electrodes (see Fig. 5). The flashing SSVEP

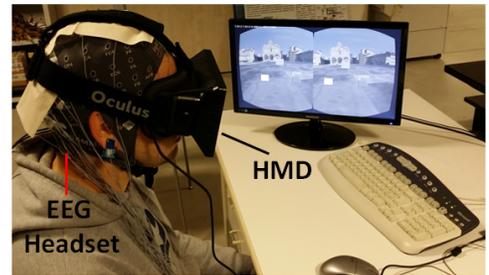


Fig. 5. The experimental setup used to validate the proposed SSVEP-BCI approach based on CNN classification.

navigation interface was presented, as well as the VE, in order to keep the training and test sessions as close as possible to the final experiment involving the whole navigation task. During the training session, all the symbols were flashing, in order to take in account any possible visual interference occurring during the final system operation. However the participant was



Fig. 6. The logical scheme of the training phase. Each trial is composed by three phase of rest (in violet) and visual stimulation (in green).

asked to select only a particular symbol at time according to a fixed sequence shown in Fig. 6.

During each “stimulus” period, the participant had to fix a particular symbol, indicated to him by a visual cue (a red frame appearing around the symbol to select). The “stimulus” periods lasted 8 s each and were spaced by a “rest” period lasting 4 s. During the rest period, no visual cues were presented and the participant was asked to look at the center of the screen.

The training session was composed by six repetitions of the described sequence. Recorded EEG data were triggered by the presented visual cues and used as a training and test dataset for the CNN learning phase.

1) EEG data pre-processing:

After EEG data registration, the EEG signals were pre-processed for CNN learning phase. We eliminated the first 2 s of each stimuli phase and the first 0.5 s of each rest phase, in order to overcome possible delays due to the attention of the participant and initial instability of the VEP: in general, at the beginning of the visual stimulation at a constant frequency firstly an unstable VEP is measured; then, a steady state potential is elicited. This time depends by brain and nerve bundles of the eyes and are different between subjects. Moreover, the initial 2 s cut suppressed possible artifacts generated by muscle activity due to eye movements (saccades) required for focusing a different symbol.

EEG data corresponding to the presentation of the same symbol (or during the rest period) are concatenated, obtaining a time sequence of 36 s of stimulation for each symbol and 63 s for the rest state. Each sequence is, then, subdivided in 2 s time frames constituting an input element for the CNN.

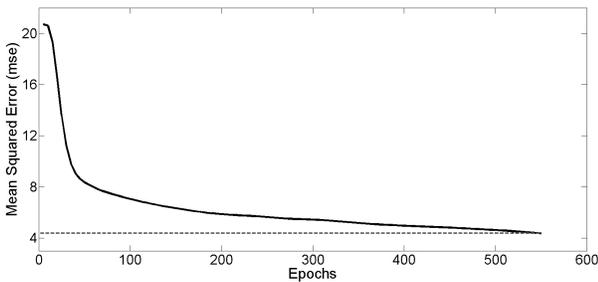


Fig. 7. Trend of the Mean Squared Error (MSE) with respect to the number of epochs for the training dataset of Subject 2.

2) NN learning:

For the CNN learning phase we used 10 and 8 frames of the dataset respectively as training and validation sets. The two used learning rates γ_1 and γ_3 are set to 0.7 chosen heuristically with experimental methodology. The weights and the thresholds of each neuron are initialized randomly a standard distribution around a mean of 0.5. In Fig. 7, the trend of the

Mean Squared Error (MSE) is shown. A new learning phase is initiated if the error does not converge.

The training time was about 20 min using a computer equipped with a AMD A6-3400M processor (1.4 GHz), with 4GB RAM.

3) NN testing:

In the testing phase, we tested 8 frames for each of the 4 classes of the dataset, corresponding to the selection of the symbol flashing at 12 Hz, 15 Hz, 20 Hz and to the “rest” state respectively. The classification time (for each frame) is around 20 ms allowing for an on-line processing of the BCI algorithm.

The graph in Fig. 8 shows the correct rate for each class and for each subject, obtained by averaging the binary classification result of the 8 processed frames.

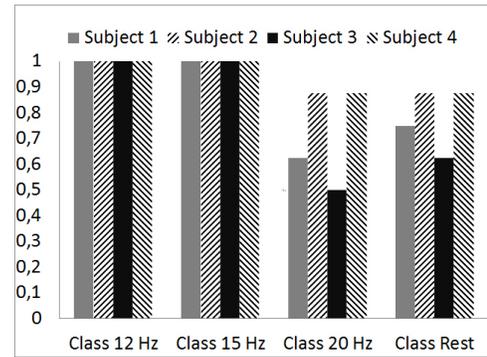


Fig. 8. Correct classification rate for each subject and each class, performed by the CNN on the test dataset.

The classification results obtained on the test dataset show that the CNN is able to classify the measured SSVEP with relatively high accuracy: the total classification accuracy, averaged over the subjects, is 0.875 ± 0.076 . Classification accuracy appears lower for the stimulus with the highest frequency, 20 Hz, and for the “rest” class, that is actually related to the absence of SSVEP in the EEG data.

The proposed method was compared to a classical method based signal to noise ratio (SNR). The SNR method extracts features as in [30] and compares them to a baseline feature’s value extracted from the rest classes. As shown in Fig. 9, the convolutional neural network outperform the classical SNR based method. The CNN average classification accuracy over all subject is 0.875 ± 0.076 while for SNR method is 0.695 ± 0.140 .

B. The BCI-SSVEP approach for navigation task

The second experiment evaluates the usability of the system within a navigation task in VE. The experiment involved three of the participants already enrolled in the first experimental session (S1, S3 and S4). The weights of the CNN were thus the same obtained from the training datasets recorded by each participant in the previous experimental session.

In the experiment, the participant is asked to control the walking movement and direction in the VE using the proposed SSVEP based navigation interface, in order to walk through a given path. The path is composed by 8 waypoints, each one marked by a flag in the VE. Only one flag is visible at time, corresponding to the active waypoint, while the others are

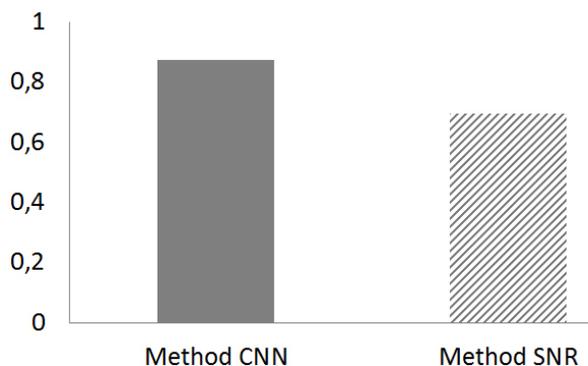


Fig. 9. Mean of the correct classification rate for all subject for each method proposed.

TABLE II.

Subjects	Time [s]	CNN Classification Accuracy
S1	470 s	0.84
S3	483 s	0.78
S4	260 s	0.93
mean	333 s	0.875
std	178 s	0.076
JoyCtrl	120 s	X

The first three rows indicate for each subject the elapsed time for completing the task and the classification accuracy of the CNN obtained in the first experimental session. Then we report the mean and the standard deviation for the three subject. In the last row, "JoyCtrl" represents the elapsed time for completing the task controlling the navigation with a joystick instead of the BCI.

hidden. When the subject is in close proximity to a waypoint, the flag disappears and the next target flag appears. The position of the waypoints is defined such to increase variability in the path directions.

For comparison purposes, after the experimental session one subject was asked to accomplish the same navigation task using a joystick instead of the BCI for the navigation control.

All the participants were able to complete the given sequence of waypoints using the BCI. Table II shows the time required by each subject to complete the task, while Fig. 10 shows the comparison between the trajectory of walking of Subject 4, and the optimal one, obtained using a physical interface. It appears that the BCI allowed a reasonable precision in controlling the navigation trajectory. However, a considerable difference in the time required for completing the task was registered: 120 s for the physical interface, and 333 s in average for the navigation with BCI. As expected, the time required for completing the path seems related to the classification accuracy of the CNN obtained during the training phase by each participant.

IV. CONCLUSION AND FUTURE WORKS

A novel BCI system for navigation in VE has been proposed and evaluated. The approach proposes the use of a simplified convolutional neural network for detection of SSVEP during a navigation task, and the implementation of the system in a highly immersive VE for evaluating its usability in a plausible navigation task.

The peculiarity of a navigation task in VE is that, unlike

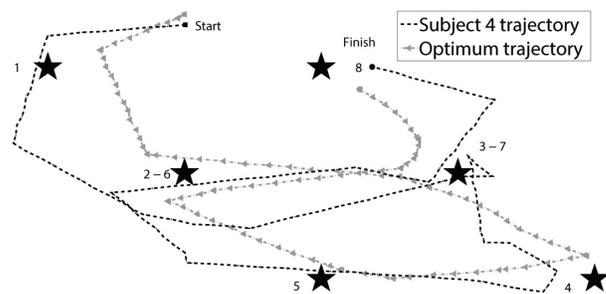


Fig. 10. The given waypoints path, marked by stars, and the actual trajectory (Black dotted line) travelled by one of the participant using the SSVEP-BCI for control of navigation. Grey dotted line represents the trajectory travelled using a physical keyboard interface in place of the BCI.

other BCI applications (such as spellers and other communication interfaces), they rely on a reduced number of control commands, though, the selection requires to be reliable and rapid in order to achieve usability of the interface itself.

Results of the first experiment on offline data show that a CNN can be applied for identification of SSVEP features in the EEG with considerably high accuracy and reliability. The main advantage of an NN is given by its adaptability to features that are subject and montage-dependent, such as conformation of the skull, conductivity of the skin-electrode interface, interference and artifacts due to brain and muscular activity, and exact positioning of the electrodes. Moreover a CNN allows to directly address the identification of features that are defined in frequency, such as the SSVEP, though obtained by combination of different channels in the space domain.

The second experiment evaluates the performance of the system through a walking navigation task within a VE scenario. The high immersivity and the realism of the VE allow to evaluate the overall usability of the proposed approach in presence of plausible factors occurring during real navigation, such as the involvement of the participant attention within the navigation task, the perception of the space in terms of position and relative velocity of obstacles and waypoints with respect to the avatar body, as well as possible interferences of the changing background scenario with VEP elicitation [31].

Taking into account the presence of those factors, results of the final experiment are promising, as the participant was able to successfully accomplish a waypoint walking task within the VE, though with a considerable delay if compared to a physical keyboard interface. It suggests that the proposed approach can be profitably applied to navigation scenarios, allowing the control of navigation task even to inert people by the use of the only brain activity.

ACKNOWLEDGMENT

This work has been partially funded from the EU FP7 project n. 257695 VERE and from the Italian POR 2007-2013 Apulian ICT Living Labs project IHCS.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

- [2] I. Sugiarto, B. Allison, and A. Graser, "Optimization strategy for ssvep-based bci in spelling program application," in *Computer Engineering and Technology, 2009. ICCET'09. International Conference on*, vol. 1. IEEE, 2009, pp. 223–226.
- [3] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an ssvep-based bci," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 1, pp. 361–364, 2008.
- [4] A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 1169–1179, 2012.
- [5] N. Birbaumer, T. Hinterberger, A. Kubler, and N. Neumann, "The thought-translation device (ttc): neurobehavioral mechanisms and clinical outcome," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 120–123, 2003.
- [6] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
- [7] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.
- [8] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 13, no. 1, pp. 12–17, 2005.
- [9] E. E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses," *Journal of Microcomputer Applications*, vol. 15, no. 1, pp. 31–45, 1992.
- [10] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 10, pp. 1181–1186, 2002.
- [11] E. Lalor, S. Kelly, C. Finucane, R. Burke, R. Reilly, and G. McDarby, "Brain computer interface based on the steady-state vep for immersive gaming control," *Biomedizinische Technik*, vol. 49, pp. 63–64, 2004.
- [12] N. Galloway, "Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine," *The British journal of ophthalmology*, vol. 74, no. 4, p. 255, 1990.
- [13] C. S. Herrmann, "Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experimental brain research*, vol. 137, no. 3-4, pp. 346–353, 2001.
- [14] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in ssvep-based bcis," *Computational intelligence and neuroscience*, vol. 2010, p. 1, 2010.
- [15] R. S. Fisher, G. Harding, G. Erba, G. L. Barkley, and A. Wilkins, "Photic-and pattern-induced seizures: A review for the epilepsy foundation of america working group," *Epilepsia*, vol. 46, no. 9, pp. 1426–1441, 2005.
- [16] G. Garcia, "High frequency ssveps for bci applications," in *Computer-Human Interaction, 2008*.
- [17] S. Muller, T. F. Bastos-Filho, and M. Sarcinelli-Filho, "Using a ssvep-bci to command a robotic wheelchair," in *Industrial Electronics (ISIE), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 957–962.
- [18] J. Legény, R. V. Abad, and A. Lécuyer, "Navigating in virtual worlds using a self-paced ssvep-based brain-computer interface with integrated stimulation and real-time feedback," *Presence: Teleoperators and Virtual Environments*, vol. 20, no. 6, pp. 529–544, 2011.
- [19] C. Loconsole, S. Dettori, A. Frisoli, C. A. Avizzano, and M. Bergamasco, "An emg-based approach for on-line predicted torque control in robotic-assisted rehabilitation," in *Haptics Symposium (HAPTICS), 2014 IEEE*. IEEE, 2014, pp. 181–186.
- [20] V. Bevilacqua, M. Dotoli, M. M. Foglia, F. Acciani, G. Tattoli, and M. Valori, "Artificial neural networks for feedback control of a human elbow hydraulic prosthesis," *Neurocomputing*, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.neucom.2013.05.066>
- [21] M. Laurino, A. Piarulli, R. Bedini, A. Gemignani, A. Pingitore, A. L'Abbate, A. Landi, P. Piaggi, and D. Menicucci, "Comparative study of morphological eeg features classifiers: An application on athletes undergone to acute physical stress," in *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*. IEEE, 2011, pp. 242–246.
- [22] V. Bevilacqua, "Three-dimensional virtual colonoscopy for automatic polyps detection by artificial neural network approach: New tests on an enlarged cohort of polyps," *Neurocomputing*, vol. 116, pp. 62–75, 2013.
- [23] G. Binetti, P. Lino, and B. Maione, "Neural network modelling of a new injection system for compressed natural gas engines," in *IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2011, pp. 628–633.
- [24] V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, and A. Pedone, "A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: the breast cancer classification problem," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*. IEEE, 2006, pp. 1958–1965.
- [25] H. Cecotti, "A time-frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses," *Pattern Recognition Letters*, vol. 32, no. 8, pp. 1145–1153, 2011.
- [26] M. Carrozzino, F. Tecchia, S. Bacinelli, C. Cappelletti, and M. Bergamasco, "Lowering the development time of multimodal interactive application: the real-life experience of the xvr project," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM, 2005, pp. 270–273.
- [27] F. Tecchia, M. Carrozzino, S. Bacinelli, F. Rossi, D. Vercelli, G. Marino, P. Gasparello, and M. Bergamasco, "A flexible framework for wide-spectrum vr development," *Presence: Teleoperators and Virtual Environments*, vol. 19, no. 4, pp. 302–312, 2010.
- [28] G. Chatrian, "Ten percent electrode system for topographic studies of spontaneous and evoked eeg activity," *Am J Electroencephalogr Technol*, vol. 25, pp. 83–92, 1985.
- [29] M. A. Pastor, J. Artieda, J. Arbizu, M. Valencia, and J. C. Masdeu, "Human cerebral activation during steady-state visual-evoked responses," *The journal of neuroscience*, vol. 23, no. 37, pp. 11 621–11 627, 2003.
- [30] L. Maggi, S. Parini, L. Piccini, G. Panfilì, and G. Andreoni, "A four command bci system based on the ssvep protocol," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 1264–1267.
- [31] C. van Hemert, "The impact of visual distractions in ssvep-based bci," in *11th Twente Student Conference on IT, Enschede 29th June, 2009*.