16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

TIME DELAY NEURAL NETWORK WITH FOURIER TRANSFORM FOR MULTIPLE CHANNEL DETECTION OF STEADY-STATE VISUAL EVOKED POTENTIALS FOR BRAIN-COMPUTER INTERFACES

Hubert Cecotti and Axel Gräser

Institut für Automatisierungstechnik (IAT)
University of Bremen
Otto Hahn Allee NW1 28359 Bremen, Germany
{cecotti;ag}@iat.uni-bremen.de
http://www3.iat.uni-bremen.de/

ABSTRACT

In BCI (Brain-Computer Interface) systems, brain signals must be processed to identify distinct activities that convey different mental states. For example, in BCIs that rely on Steady-State Visual Evoked Potentials (SSVEP) recognize EEG activity that reflects attention to a particular stimulus. In this paper, we propose a new technique for the classification of SSVEP-activity for non-invasive BCI. The proposed method is based on a Time Delay Neural Network that includes a Fourier transform in order to switch between the analysis of layers in the time domain to layers in the frequency domain. The first step allows the creation of different channels. The second step is dedicated to signal processing of these channels in the frequency domain. The last step is the classification. The presented results on offline processing correspond for electroencephalographic (EEG) signals obtained with 6 electrodes on 3 subjects. We compare our method on two time segments with the average combination method. These results outperform the average combination method and are promising for online processing.

1. INTRODUCTION

Brain-computer interface (BCI) systems allow people to communicate through direct measures of brain activity [1]. Unlike all other means of communication, BCIs require no movement [9]. Therefore they have been used primarily to enable communication for persons with severe disabilities who are unable to communicate through any classical ways. Although BCIs are mostly dedicated to persons with disabilities, recent works have shown that BCIs could be efficiently used by healthy persons on different applications like as a combination and complement with other interfaces [7].

To classify different brain signals, the knowledge of the stimuli drives the solution to some specific signal processing analysis. Usually, a BCI is decomposed into four parts: the signal acquisition, the signal processing, the output device components and the operating protocol that links the three previous components [1]. We focus here on the signal processing component. This part includes two main steps: the extraction of brain signal features and the translation of these signals into device commands.

The actual two challenges in BCIs are:

- to reduce the latency between the user's command and its activation. Typically, reducing BCI latency impairs accuracy. It's useless to deliver a fast latency if the order is not well interpreted, i.e., if the signals are not well analyzed. This issue depends on the application: for disabled people, the quality of the controls must be very high for keeping an optimal security of the subjects. The system architect has to find the best compromise between speed and accuracy.
- to avoid the need of an expert to set the system. The system must be set automatically to the user. Furthermore, it is not possible to generate a single set of parameters for several users. The system must be set manually by an expert or automatically online by solving local problems or offline by using some training sessions and their feedback. In this last case, the time dedicated for the experiments and the choice of representative experiments is crucial.

The experiments in this work are made on a particular case of BCI:

- The system is non-invasive. It does not require surgery to implant electrodes.
- It only uses sensors with contact on the surface of the scalp via standard electroencephalographic (EEG) electrodes.
- The stimuli are only flickering lights and their responses should correspond to Steady-State Visual Evoked Potentials (SSVEP). The system must reflect the user attention to a fast oscillating stimulus. The best response for these signals are obtained for stimulation frequencies between 5 and 20Hz [14].

In the following sections, we propose a model combining signal processing techniques that incorporate a priori knowledge of the problem and machine learning methods to adapt to each subject [3]. The first section presents the EEG signals from SSVEP. The main system is described in the second section. The experiments are detailed in the third section and the results are discussed in the last section.

2. EEG SIGNALS

The classification of EEG signals is a major challenge for real applications [4]. Different types of classifiers have been used for EEG problems like neural networks [2, 8] and Hidden Markov Model [15]. For the EEG signal acquisition, we consider $N_{elec}+1$ electrodes, where one has only a reference and ground purpose. Let SF be the sampling frequency and TS the time segment in second attributed to the analysis of the signal. For SSVEP stimuli, we consider visual stimulation with a flicker frequency of $f\,Hz$.

We estimate the EEG signal measured as the voltage between a reference electrode and the electrode number i as:

$$X_{i,j} = \sum_{k=1}^{k=SF} (a_{i,k} sin(2\pi * k * f * j + \Phi_{i,k})) + B$$

where $0 \leq j < SF * TS$, B corresponds to the noise. Each sinusoid on each electrode has its own amplitude ($a_{i,k}$) and phase ($\Phi_{i,k}$).

The nuisance signals B can have several origins:

- The environment and its effect on the subject.
- Natural physical disturbance like other brain processes, breathing artifacts.
- The noise of each electrode on the cap.

A channel is used for a linear combination of the signals measured by the N_{elec} electrodes. The values of a channel c at a time j are defined by:

$$c_j = \sum_{i=1}^{i=N_{elec}} w_i * X_{i,j}$$

The information from the electrodes is contained in one scalar at a time j. For the EEG signal processing, one first goal is to find an optimal set $w_{i,k}$, $0 \leq i < N_{elec}$ and k is a finite number, as small as possible.

There exist different methods for the creation of one or several channels:

- Average combination. For this approach, the sinusoids signals should have equal phases and the noise should be equally distributed across the electrodes for a low dependence between electrodes. The average method fuses each electrodes in one channel with equal weight for each electrode.
- Native combination. Neuroscience works provide the information that the phases of the SSVEP sinusoids vary with the location of the electrodes on the scalp [5]. In this case, each channel represents the signal of one electrode.
- Bipolar combination. The noise is present in each electrode signal and the average combination may amplify the noise instead of reducing it. The goal of the bipolar approach is to obtain a better signal by canceling the common nuisance signals [12].
- Laplacian combination. This combination is an alternative to the bipolar solution. One electrode has a high weight and its neighboring electrodes has negative weights.
- Minimum energy combination. This method allows the combination of a fixed number of electrodes that cancel the noise as much as possible.
- Maximum contrast combination. This method is a variation of the previous method. The SSVEP frequencies are maximized whereas the energy in the noise is minimized simultaneously [6].

In this paper, we propose to determine the weights of each channel function to their discriminant power when the channels are combined. The goal is to determine the optimal set of weights for a finite number of channels that can help to solve the main problem. In the following section, we propose a solution for creating such channels.

3. SYSTEM OVERVIEW

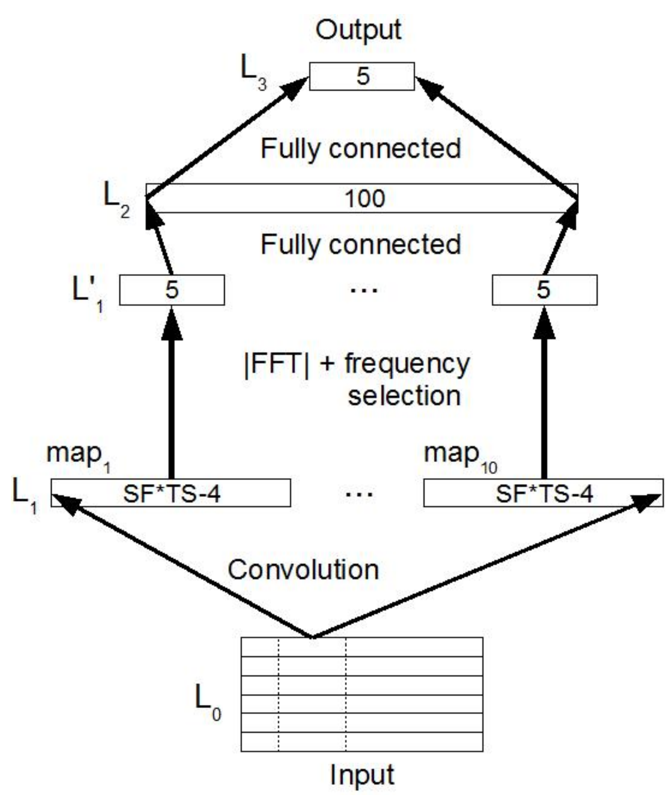
We propose a model for the processing and classification of EEG signals that correspond to a finite number of SSVEP signals. The model is based on a time delay neural network (TDNN), which is a special case feedforward neural network [11]. The goal is to directly classify the raw signal and to integrate the signal processing functions within the discriminant steps. For different aspects, it is not always possible to clearly separate the pre-processing, the feature extraction and the classification steps for a classification task. The pre-processing and feature extraction usually require a lot of signal processing methods in order to transform the raw data in some possible discriminant data for a classifier. That's why we propose to combine signal processing methods and the feedback of its effect on the final output. The framework of the system is described in the figure 1. The inputs are the EEG signal values from the electrodes during a time segment, $X_{i,j}$, $0 \leq i < N_{elec}$, $0 \leq j < SF * TS$. The output is a vector of size N_{sig} (the SSVEP frequencies). Thus, for the classification task, there are N_{sig} classes.

The process within the neural network can be described as follows:

1. The signal normalization, denoising and combination.
2. The selection of the frequencies and their harmonic.
3. The classification of the signals.

Before processing the signal, data is normalized: $X_{i,j} \leftarrow (X_{i,j} - \bar{X_{i,j}})/\sigma_{i,j}$ where $\bar{X_{i,j}}$ and $\sigma_{i,j}$ are respectively the average and the standard deviation values of the electrode i at the time j in a time segment TS. These data are extracted from the whole training database.

Figure 1: System overview.



3.1 Neural network topology

The network is composed of 4 layers. Each layer is composed of one or several maps. We define a map as an layer entity that has a specific semantic: each map of the first hidden layer is a channel. The first hidden layer is dedicated to the denoising of the input data and the creation of the different channels.

The layers:

- Layer 0 (L_0): the input layer. $X_{i,j}$ with $0 \leq i < N_{elec}$ and $0 \leq j < SF*TS$. $SF*TS$ corresponds to the number of samples in TS seconds.
- Layer 1 (L_1): the first hidden layer. L_1 is composed of 10 maps. We define L_1M_m, the map number is m. Each map of L_1 has the size 1*$SF*TS-4$. The size of the maps in the time dimension are inferior to the input layer because of the border effect of the filtering in time.
- Layer 1' L'_1): L'_1 is the result of L_1 after some signal processing. Each map of L'_1 has the size 1*5.
- Layer 2 (L_2): the second hidden layer. L_2 is composed of 1 map of 100 neurons. This map is fully connected to its corresponding map on L'_1.
- Layer 3 (L_3): the output layer. This layer has only one map of 5 neurons, which represents the 5 classes. This layer is fully connected to each map of L_2.

The learning τ rate is defined by:

$$\tau = \frac{0.3}{\sqrt{L_lM_n(n)_{N_{input}}}}$$

where $L_lM_n(n)_{N_{input}}$ is the number of input of the neuron n in the map m of the layer l.

For the first layer, we have a different learning rate τ':

$$\tau' = \tau * \frac{2}{\sqrt{L_1M_m(0)_{N_{shared}}}}$$

where $L_1M_m(0)_{N_{shared}}$ is the number of neurons that share the same set of input weights.

The weights of the network are initialized with a standard distribution around $\pm 1/L_lM_n(n)_{N_{input}}$.

3.2 Propagation

The first hidden layer is dedicated to 2 tasks:

- The automatic creation of the channels (their weights), one for each map.
- The automatic linear filtering of the signal in time. This step may be useful to cancel some artifacts due to difference of the phase in the signal between electrodes.

For each map of L_1, we apply the following algorithm:

- Convolution: The value of a neuron n of L_1M_m is defined by:

$$L_1M_m(n) = f\left(\sum_{i,j=0,0}^{i,j=N_{elec},5} (X_{SF*TS*i,j+n} * W_{m,i,j}) + W_{m,threshold}\right)$$

where $0 \leq m < 10$, $0 \leq i < N_{elec}$, $0 \leq j < SF*TS$, $0 \leq n < SF*TS-4$ and f is defined by:

$$f(\sigma) = 1.7159 * tanh((2.0/3.0)*\sigma)$$

where $f(1) = 1$ and $f(-1) = -1$ [10]. Notice that each neuron of the map shares the same set of weights and it is only connected to a window of size $5*N_{elec}$. This window allows filtering the signal in the space and time domain at the same time. Instead of learning one set of weights for each neuron, dependent on the neuron's position, the weights are learned independent of their corresponding output neuron. It is this particular feature that qualifies the network as time delay neural network, which is widely used in handwriting recognition [11].

- Once the channel is created, the Fourier Transform is applied on the neuron's value to change to the frequency domain.

$$L''_1M_m(n) = \frac{1}{SF*TS} \sum_{k=0}^{k=SF*TS-1} L_1M_m(n) * e^{\frac{-2\pi i}{SF*TS}*n*k}$$

- We select the frequencies between 13Hz and 17Hz, the fundamental expected frequencies. For example, if $TS = T$ seconds, $T > 1$, the neuron of $L'_1M_m(n)$ gets the following new value:

$$L'_1M_m(n) = \frac{1}{T} \sum_{t=T*(n+13)}^{t<T*(n+13+1)} |L''_1M_m(t)|$$

- Now we can consider the layer L'_1 and its map $L'_0M'_m$, which have all 5 neurons.
The neuron $L'_1M'_m(n)$ corresponds to the frequency $n+13$, $0 \leq n < 5$.

The first hidden layer can be seen as a feature extraction level. Although it could be possible to choose directly the maximum value for each map of L'_1M_m as a valid output answer, it misses the fact that the channels have been tailored for performing together.

Between L'_1 and L_3, it corresponds to a classical multilayer Perceptron. L_2 and L_3 are fully connected, respectively to L'_1 and L_2. The activation function each neuron of L_2 and L_3 is:

$$f(\sigma) = 1/(1+e^{-\sigma})$$

A signal X is attributed to the class C_i, $i \leq 0 < N_{sig}$ when $i = argmax_n L_3(n)$. We note $E(X)$, the recognized class.

3.3 Backpropagation

The backpropagation for the L_3 and L_2 is achieved by using a gradient descent by minimizing the least mean square error. However we have to transform the error back in the time domain by using the Inverse Fourier Transform for correcting the weights in the first hidden layer. As the errors in this layer are complex numbers, only the real part is used for updating the weights:

- First the frequencies between 13Hz and 17Hz are transfered back to the frequencies function to the time segment analysis. The error is 0 on all the frequencies except between 13Hz and 17Hz.
- The error δ is then transfered back into a complex number as only the magnitude was propagated.

$$\delta''_1M_m(n) = \delta'_1M_m(n/TS)e^{i*\theta(m,n)}$$

where $\theta(m,n)$ is the angle of $L''_1M_m(n)$

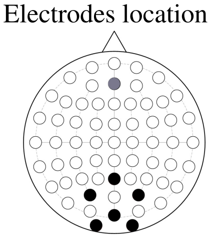
- The Inverse Fourier Transform is applied on $\delta"_1 M_m(n)$ and we keep only the real value for the error.

$$\delta_1 M_m(n) = Re\left(\frac{1}{SF*TS}\sum_{k=0}^{k=SF*TS-1} \delta"_1 M_m(n) * e^{\frac{2\pi i}{SF*TS}*n*k}\right)$$

4. EXPERIMENTS

EEG signals were recorded on 3 subjects. For each subject, we have 5 different trials for each of the 5 stimuli at 13, 14, 15, 16 and 17Hz. The subjects had to focus for about 20s at a light flickering at one of the 5 frequencies. The subjects sat on a chair, wearing a cap with electrodes placed on the surface of the scalp (the contact between the skin under the hair and the electrodes on the cap is possible thanks to some gel). The figure 2 represents the location of the electrodes (in black) and the reference electrode (in grey) on the cap.

Figure 2: Electrodes location on the cap.



4.1 Database description

The first 2s of each trial are not considered as they contain too much artifacts. Within the 18s remaining of the signal, we cut the signal into different windows of size $SF*TS$ every 100ms. For each signal, the three first trials are dedicated to the training of the system, the fourth trial is used for the validation and the fifth is for the test. Two time segments have been considered: 1s and 2s. The number of signals to process for each database are presented in the table 1.

Table 1: Number of patterns for each database.

	Learning	Validation	Test
1s	1260	420	420
2s	1185	395	395

5. RESULTS

We define the accuracy of the system by its recognition rate R. The recognition rate is defined by:

$$R = \frac{\sum_{X\in DB}((E(X) = C_i) \text{ and } (X \in C_i))}{\sum_{X\in DB} X \in C_i}$$

where DB is the considered database that contains the EEG signals of length TS.

The recognition rates obtained thanks to the average method are given in the table 3. The table 2 presents the results obtained with the TDNN. For each subject and for the 2 time segments, the recognition rate in the learning, validation and test is given for the epoch that gives the best recognition

Table 2: Results of the TDNN.

Subject	1	2	3
Reco. Learning			
1s	90.95	92.62	93.02
2s	95.36	99.32	97.64
Reco. Validation			
1s	70.71	59.29	62.86
2s	78.48	60.51	74.18
Reco. Test			
1s	82.14	53.10	48.33
2s	93.67	61.77	58.99

Table 3: Results of the Average combination.

Subject	1	2	3
Reco. Learning			
1s	62.53	50.47	35.95
2s	70.21	62.02	38.14
Reco. Validation			
1s	56.19	28.09	40.00
2s	60.25	28.60	46.32
Reco. Test			
1s	68.33	25.95	46.90
2s	76.70	31.89	47.08

Table 4: Confusion matrix for the subject 1 (test) - 1s.

	13Hz	14Hz	15Hz	16Hz	17Hz
13Hz	77	0	0	0	7
14Hz	5	65	4	3	7
15Hz	8	4	70	0	2
16Hz	0	0	0	84	0
17Hz	24	2	3	6	49

Table 5: Confusion matrix for the subject 1 (test) - 2s.

	13Hz	14Hz	15Hz	16Hz	17Hz
13Hz	79	0	0	0	0
14Hz	1	74	1	2	1
15Hz	7	4	68	0	0
16Hz	0	0	0	79	0
17Hz	3	1	5	0	70

rate on the validation base. Thus the test base is like a blind test. The test and validation database have the same behavior during the learning for their evaluations.

Concerning the strategy of the channel creation, our approach is good as it allows the creation of channels that combine well. It is possible to classify these noisy signals thanks to our strategy. However, for the validation and the test set, accuracy is poor but always better than the average combination. These results validate the idea behind the method, but also the need to adapt to the user. We can also notice that it is hard for a subject to focus for even 20s at the same signal. The subject can shift his gaze and produce unwanted signals. There is thus an inevitable risk that the data used for training may have some parts that do not correspond to a SSVEP response. Furthermore, the electrodes only represent a small part of all physical possible features.

Subject 1 is the best and gives very good SSVEP response. Subject 2 seems to respond well on the trials of the learning database but fails for the validation and test base. For subject 3, the contrary pattern is apparent. He exhibits bad results on the learning database but succeeds better on the two others.

The confusion matrix in tables 4 and 5 for the best subject lets us think that it is difficult to diagnose the errors from the ranking. A hypothesis could have been made that the errors would be in the neighborhood frequencies of the expected frequency. However, we do not find any evidence of this behavior on the confusion matrix. Finally, the recognition decrease is not linear function to the time and it is already possible to achieve good results with a time segment of 1s; it suggests that it is still possible to decrease the time segment for some subjects.

6. CONCLUSION

A new model based on a Time Delay Neural Network for the classification of SSVEP signals on BCI systems has been presented. This method allows the automatic creation of channels and linear time filter functions to the user and their discriminant powers for the signal classification. The network integrates the Fourier transform between 2 hidden layers, which changes the layer semantic from time to frequency domain analysis. The obtained results are promising, as the experimental condition and test do not exploit all the information that may be available during the BCI application. Further works will deal with the problem of transition between EEG signals and how to integrate in an optimal way temporal information about the previously recognized signals to improve the recognition.

Acknowledgment

This research was supported by a Marie Curie European Transfer of Knowledge grant Brainrobot, MTKD-CT-2004-014211, within the 6th European Community Framework Program.

REFERENCES

[1] B. Z. Allison, E. W. Wolpaw, and A.R. Wolpaw, "Brain-computer interface systems: progress and prospects", Expert Review of Medical Devices, vol. 4, no. 4, 2007, pp. 463–474.

[2] C.W. Anderson, S.V. Devulapalli, and E.A. Stolz, "Determining Mental State from EEG Signals Using Parallel Implementations of Neural Networks", IEEE Workshop on Neural Networks for Signal in Processing, Cambridge, MA, USA, 1995, pp. 475-483.

[3] N. Birbaumer, and L. G. Cohen, "Brain-computer interfaces: communication and restoration of movement in paralysis", Journal of Physiology-London, 579(3), 2007, pp. 621–636.

[4] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, K.R. Muller, "The Berlin brain-computer interface: EEG-based communication without subject training", IEEE Trans. on Neural Systems and Rehabilitation Engineering, vol. 14, no. 2, 2006, pp. 147–152.

[5] G. Burkitt, R. Silberstein, P. Cadush, and A. Wood, "Steady-state visual evoked potentials and travelling waves", Clin. Neurophysiol, vol. 111, no. 2, 2000, pp. 246–258, 2000.

[6] O. Friman, I. Volosyak, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces". IEEE Trans. on Biomedical Engineering, vol. 54, no. 4, 2007, pp. 742–750.

[7] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain-computer interface (BCI)?" IEEE Trans. Neural Syst Rehabil Eng, vol. 11, no. 2, 2003, pp. 145–147.

[8] E. Haselsteiner and G. Pfurtscheller, "Using time dependent neural networks for EEG classification", IEEE Trans. on Rehabilitationi Engineering, vol. 8, no. 4, 2000, pp. 457–463.

[9] A. Kostov, and M. Polak, "Parallel man-machine training in development of EEG-based cursor control", IEEE Trans Rehabil Eng, vol. 8, no. 2, 2000, pp. 203–205.

[10] Y. LeCun, L. Bottou, G. Orr, and K. Muller, "Efficient BackProp, in Neural Networks: Tricks of the trade", G. Orr and Muller K., eds., 1998.

[11] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, 1998, pp. 2278–2324.

[12] G. Muller-Putz, R. Scherer, C. Brauneis and G. Pfurtscheller, "Steady state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components", J. Neural Eng., vol. 2, no. 4, pp. 123–130, 2005.

[13] D.J. McFarland, D.J. Krusienski, and J.R. Wolpaw, "Brain-computer interface signal processing at the Wadsworth Center: mu and sensorimotor beta rhythms, Progress in Brain Research", vol. 159, chapt. 26, 2006, pp. 411–419.

[14] M. Pastor, J. Artieda, J. Arbizu, M. Valencia, and J. Masdeu, "Human cerebral activation during steady-state visual evoked responss", J. Neuroscience. vol. 23, no. 37, pp. 11 621-11 627, 2003.

[15] S. Zhong and G. Joydeep, "HMMs and coupled HMMs for multi-channel EEG classification", In Proc. IEEE Int. Joint. Conf. on Neural Networks, vol. 2, 2002, pp. 1154–1159.