Towards an API for EEG-Based Imagined Speech classification

Luis Alfredo Moctezuma[1], Marta Molinas[1], A. A. Torres García[2], L. Villaseñor Pineda[2], and Maya Carrillo[3]

[1] Department of Engineering Cybernetics, Norwegian University of Science and Technology. Trondheim, Norway
`luisalfredomoctezuma@gmail.com, marta.molinas@ntnu.no`

[2] Computer Science Department, Instituto Nacional de Astrofísica Optica y Electrónica. Puebla, Mexico
`{alejandro.torres, villasen}@ccc.inaoep.mx`

[3] Faculty of Computer Science, Benemérita Universidad Autónoma de Puebla. Puebla, Mexico
`cmaya@cs.buap.mx`

Abstract. In this paper, imagined speech classification is performed with an implementation in Python and using scikit-learn library, to create a toolbox intended for real-time classification. To this aim, the Discrete Wavelet Transform with the mother function Biorthogonal 2.2 is used to then compute the *instantaneous* and *Teager* energy distribution for feature extraction. Then, *random forest* is implemented as a classifier with 10-folds cross-validation. The set of experiments consists of *imagined speech classification, linguistic activity and inactivity classification* and *subjects identification.* The experiments were performed using a dataset of 27 subjects which imagined 33 repetitions of 5 words in Spanish *up, down, left, right and select.* The accuracy obtained with the models were 0.77, 0.78 and 0.98 respectively for each task. The high accuracy rates obtained as a result attest for the feasibility of the proposed method for subject identification.

Keywords: Imagined Speech, Linguistic activity, Subjects identification, Discrete Wavelet Transform, Brain Computer Interfaces, Electroencephalograms, Application Programming Interface

1 Introduction

In the last years, exploration into the identification of various brain activities has considerably increased, motivated by its potential use as a new way of communication, with applications ranging from games to medicine and in general to augmented human capabilities. A brain computer interface (BCI) is a communication system that monitors the brain activity and translates features, corresponding to the user's intention/thoughts/movements into control commands.

When thoughts features are identified and extracted, this can become a new way of human-machine interaction that allows users to employ their thoughts to control/operate external devices. BCI techniques can be classified into invasive and non-invasive, the first one requiring surgical procedures for which very clear brain signals are obtained because the measurements are not attenuated by the skull and the scalp. Non-invasive BCI, does not require any surgery but signals obtained are weaker. Non-invasive BCI are the most used due to their relatively low cost and easy setup.

Neurophysiological exploration based on the bioelectrical brain activity (Oscillations of brain electric potentials, frequency spectrum in Hz) registered during unconstrained rest, sleep, or with different activation functions are called Electroencephalography (EEG). Electrophysiological sources refer to the neurological mechanisms (also known as neuro-paradigm) used by a BCI user to generate control signals [1]. Wolpaw et al. [2] and Bashati et al. [1] separated these electrophysiological sources into categories based on neuronal mechanisms and the recording technology they use. There are some invasive categories, but for the interest of this work the non-invasive categories are: sensory-motor activity, potential P300, visual evoked potentials (VEP), slow cortical potentials (SCP) and response to mental and multiple tasks neuromechanisms.

Later [3] added the imagined or internal speech; which refers to the internal or imagined pronunciation of words but without uttering sounds and without articulating gestures. Imagined speech as an electrophysiological source has advantages over others, because it needs little training. In this research, imagined speech from EEG signals are employed, the dataset consisting of EEG signals from 27 subjects captured while imagining 33 repetitions of five words in Spanish; *up, down, left, right* and *select*.

The state-of-the-art reports the use of imagined speech as Electrophysiological source, but in the majority of the cases they are limited to reduced vocabularies (yes or no, etc.) or limited to syllables/phonemes. However, the applications can be limited in these cases. If instead of syllables/phonemes are used complete words, the applications can range from medical to biometric systems. The work in this paper is aimed at a new way of communication in real-time between people who cannot produce sounds or with specific diseases like Amyotrophic Lateral Sclerosis [4,5].

Although the reported literature includes research employing imagined speech, the on-line/real-time process is not considered. For a real-time implementation it is first important to identify if the signals correspond to linguistic activity and once the linguistic segment is identified in a signal, another process using a multi-class classifier can determine the imagined word associated to the respective signal.

In this area, there is need for efforts to create a method for *transfer learning* [6], because in real applications people in need of BCI solutions may have some difficulties to train a model anew. The idea of *transfer learning* is to create a classifier with training from a group of subjects and use in a different group of subjects. In [7], the authors reported experiments and propose that a first stage of calibration is needed because the signals are sufficiently different between subjects and between sessions. Other experiments also suggest the use imagined speech from EEG data as a biometric measurement for subject identification [8], and show experiments limited to syllables. Security systems used by organizations to manage access to facilities, equipment or resources and to protect against theft or espionage by denying unauthorized access, is one of the first applications envisioned by this BCI concept. Different types of safety measures have been proposed and used for a long time, ranging from the use of standard systems (security guards, smart cards, etc) to the use of biometric measurements (fingerprint, palm-print, etc.). A biometric recognition system is able to perform automatic recognition of subjects based on their physiological and/or behavioral features[9]. Any human physiological and/or behavioral characteristic can be used as a biometric characteristic as long as it satisfies the following requirements: *universality, permanence, collectability, performance, acceptability and circumvention.*

Biometric systems are advantageous compared to generic system, because they are more difficult to steal, compromise or duplicate, and can be more convenient for the users since a single biometric trait can be used for the access into several accounts. However, current biometric systems are vulnerable to attacks aimed at undermining the integrity of the authentication process [10]. For example, an intruder may fraudulently obtain the latent fingerprints of a user and later use it to construct a digital or physical artifact of the user's finger [11].

Building on the above existing knowledge and on the need for inviolable methods for subject identification, this paper proposes to advance the research by employing imagined speech from EEG data to identify subjects in real-time. For the experiments, an on-line environment (simulated real-time) was used, as it is explained in the following sections.

The paper is organized as follow; first, the proposed method for different tasks are presented and explained in brief. Next, some experiments to show the application of the proposed method in *imagined speech classification, linguistic distinction* and *subjects identification* are described, to finally discuss the feasibility of the method in real application and possible future improvements.

2 Proposed method

In general, the method can be summarized in 3 fundamental steps: Pre-processing, feature extraction and classification, as described in the flowchart in figure 1. To ensure that instances in training data (off-line) are not used in Test data (on-line), 30% was first separated for *Test Set* and 70% for training and to create the model using 10-folds cross-validation.

To implement the proposed method and used it in real-time, it was created an Application Programming Interface (API) using *Django* with Python 2.7 and to store the models per each task and for all experiments a database in MySQL was used.

In this version of the project the API[4] consist of 2 EndPoints; *dwt/training* and *dwt/{model_id}* with the POST method according to HTTP methods [12].

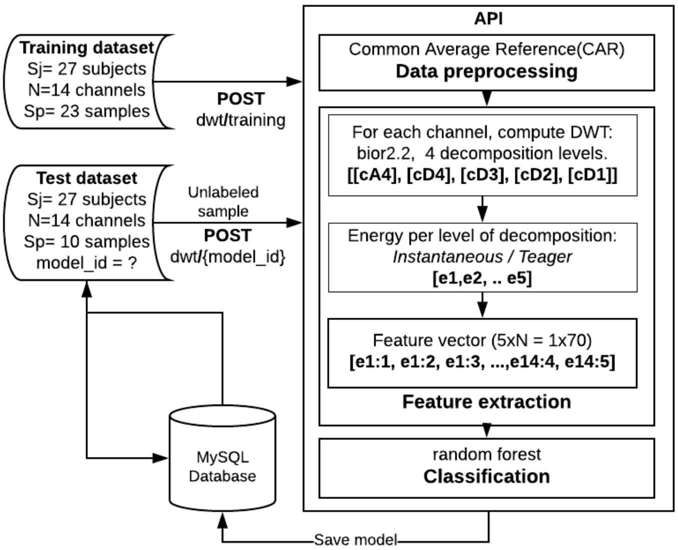

Fig. 1. Flowchart summarizing the steps of the proposed method.

2.1 Dataset

The dataset consist of EEG signals from 27 subjects captured using EMOTIV EPOC while imagining 33 repetitions of five imagined words in Spanish; up, down, left, right and select (Corresponding to *arriba, abajo, izquierda, derecha* and *seleccion*). Each repetition of the imagined words was separated by a state of rest, as shown in figure 2 and described in [13].

EEG signals were recorded from 14 high resolution channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4; see figure 3) with a sample frequency of 128 Hz which were placed according to the 10-20 international system [14].

2.2 Pre-processing

In order to reduce the signal-to-noise ratio, the common average reference (CAR) method was used [15,16]. As we can see in the formula 1, the CAR method remove the common data in all electrodes recorded simultaneously.

[4] For more information about this public project, visit: `https://github.com/wavesresearch/eeg_api_dwt`

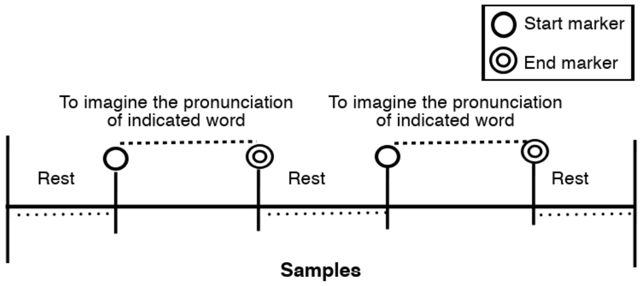

Fig. 2. Protocol designed in [13] for EEG signal acquisition using EMOTIV EPOC.

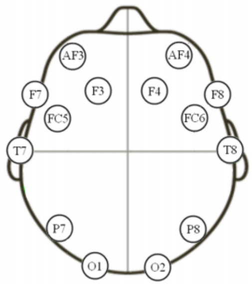

Fig. 3. 10-20 international system for 14 channels [14].

$$V_i^{CAR} = V_i^{ER} - \frac{1}{n}\sum_{j=1}^{n} V_j^{ER} \tag{1}$$

Where V_i^{ER} is the potential between the ith electrode and the reference, and n is the number of electrodes.

2.3 Feature extraction

The EEG signals are usually non-stationary, they change rapidly over time and patterns of brain activity contain information related to specific variations over time. A representation of the signal that considers this behavior is necessary for a proper feature extraction.

When the DWT is applied to a signals S with a decomposition level j=*4*, it will give a structure with vectors of approximation cA_j and detail cD_j coefficients: $[cA_4, cD_4, cD_3, cD_2, cD_1]$, as shown in the figure 4 and the table 1 shows the related frequencies per level of decomposition.

According to the average size of the dataset, 4 decomposition levels were applied. However, the wavelet coefficients for each decomposition level will vary depending on the signal size(the duration of imagined pronunciation of the word,

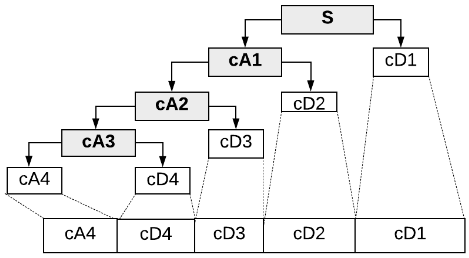

Fig. 4. Coefficient vectors in the 4th decomposition level of DWT for a signal S.

Table 1. DWT with 4 decomposition levels, frequencies ranges and related brain rhythms.

Level	Frequency range	Brain rhythm
cD1	32-64	*Gamma*
cD2	16-32	*Beta (16-30 Hz) and Gamma (30-32 Hz)*
cD3	8-16	*Alpha (8-12 Hz) and Beta (12-16 Hz)*
cD4	4-8	*Theta*
cA4	0-4	*Delta*

between imagined words from the same subject and between imagined words from different subjects).

To deal with this problem the *instantaneous* and *teager* energy distribution was calculated [17]. These energy distributions were used since they have shown the best results related to imagined speech [18,7]. When energy coefficients are calculated, it is possible to have the same number of features for all instances.

In this work, the feature vector for each instance was represented with energy coefficients that were calculated for each decomposition level of the DWT biorthogonal 2.2(bior2.2) and for each channel which were then concatenated in order to have a single feature vector. The expressions for these energy distributions are shown below:

- Instantaneous: gives the energy distribution in each band [17]:

$$f_j = log_{10}\left(\frac{1}{N_j}\sum_{r=1}^{N_j}(w_j(r))^2\right) \tag{2}$$

- Teager: This energy operator reflects variations in both amplitude and frequency of the signal and it is a robust parameter for speech recognition as it attenuates auditory noise [19,17].

$$f_j = log_{10}\left(\frac{1}{N_j}\sum_{r=1}^{N_j-1}|(w_j(r))^2 - w_j(r-1) * w_j(r+1)|\right) \tag{3}$$

At this point, instead of having a features vector for each decomposition level we have a single value for each one, and the process is repeated for each channel. After this process, we have 5 values per channel $(CA_4, CD_4, CD_3, CD_2, CD_1)$ and then all 14 channels are concatenated in order to have a single feature vector with 70 coefficients to represent each instance of the EEG signal.

2.4 Classification

Once feature vectors are compute and obtained for each instance of the EEG signal, *random forest* was used for automatic classification with an implementation in Python 2.7 using the library scikit-learn [20]. For all experiments, the parameters for *random forest* in scikit-learn were: $max\_depth = 5, random\_state = 0, criterion = gini$, that were selected after tested all possibilities. This classifier was selected because of the good results reported in the authors' previous work on imagined speech classification using EEG signals [13,18].

To evaluate the classifier performance with 10-folds cross-validation, an accuracy index was defined and calculated.

3 Experiments and results

In this paper an implementation of the proposed method towards on-line BCI for identification of imagined speech from EEG signals is developed.

The results from several experiments using DWT bior2.2 with Instantaneous and Teager energy distribution are reported here. For all described experiments the models were created off-line and tested on-line, using the API created. The experiments were separated in 3 different groups. First we present the result of using the proposed method for imagined speech classification. Then, we separate the imagined speech into a class named *linguistic activity* and the states of rest into a class named *linguistic inactivity*, in order to check if the proposed method can distinguish the imagined speech from others activities (rest). In addition the signals were separated and tagged with a subject ID (S1, S2, .. S27) to create a *machine learning* model for subjects identification and them use it in real-time. It were made with subject-level analysis and subject-word-level analysis as shown in the next experiments.

3.1 Imagined speech classification

Once the feature extraction was applied to the EEG signals, a classifier for each subject using *random forest* was created. For each subject, the classifier consists of 5 classes corresponding to the 5 imagined words (up, down, left, right and select), and for each imagined word 23 instances were used.

The average accuracy for imagined speech classification task is presented in figure 5, where it can be noted that when the Teager energy distribution was used, the highest accuracy was reached (0.77).

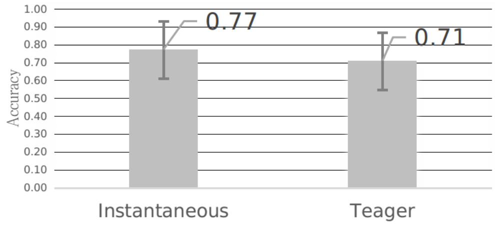

Fig. 5. Average accuracy and standard deviation obtained with 27 subjects for imagined speech classification.

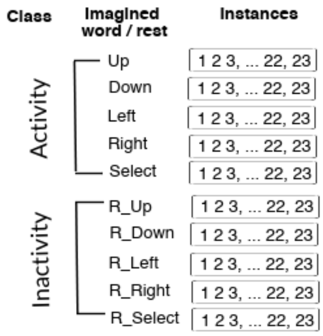

Fig. 6. Procedure to separate instances into linguistic activity and inactivity.

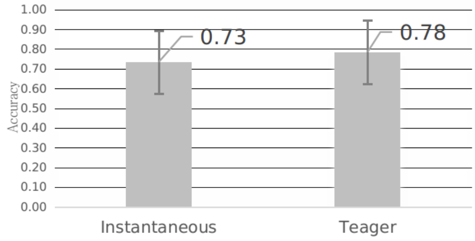

Fig. 7. Average accuracy and standard deviation obtained with 27 subjects for Linguistic activity and inactivity distinction.

In this case, the model saved in the database to use it in real-time was with instantaneous energy values. Using the trained model in on-line environment with the 30% of the instances, the average accuracy obtained was **0.85**, it can be noted that the accuracy is highly related to the accuracy of the model.

3.2 Linguistic activity and linguistic inactivity

In a complete real-time application (In a continuous recording without restrictions) it is necessary to distinguish the brain activity generated by the subject when imagining a word (linguistic activity) from any other brain activity (non-linguistic activity or inactivity). In this part the "complete real-time application" refer to identify the linguistic activity segment and then use another classifier to detect the specific imagined word.

In this experiment the EEG signals were separated into 2 classes, a set of imagined words considering the class of linguistic activity; and states of rest (unconstrained rest) as examples of other brain activity. In this work, the latter are called *linguistic inactivity*. The process to separate the instances into 2 classes is shown in the figure 6, where *R_up, R_Down, R_Left, R_Rigth, R_Select* correspond to rest states between each repetitions.

The model saved in the database was with the instantaneous energy distribution, and in the test stage the average accuracy obtained was **0.78**.

3.3 Subjects identification

This experiment was carried out to check if there are sufficient information in the EEG data for this task and to create a model for subject identification. For this, the 115 instances of imagined words per subject (corresponding to 23 repetitions per each of 5 imagined words) were considered as a single class, tagged it with a subject ID (S1, S2, ..., S27).

This experiment was performed with 27 subjects using Instantaneous and Teager energy distribution based on the DWT bior2.2. The results obtained in the classification step with 10-fold cross-validation with *random forest* are shown in the table 8.

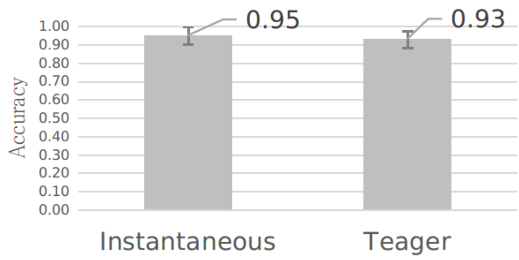

Fig. 8. Accuracy obtained when a classifier with all imagined words as a single class was created with 27 subjects using 10-fold cross-validation.

In figure 8 it can be observed that when using instantaneous energy distribution the best accuracy obtained is 0.95. However using the Teager energy the accuracy is similar and we can use them both. However, the saved model was with instantaneous energy, and after the use of *Test Set* the average accuracy obtained was 0.92.

This result suggests that it is possible to identify subjects regardless of the word that the subject is imagining. This means that a subject can be identified using different words. This brings up the question whether there a specific word that is best suited for subjects identification.

To answer this question, the following experiment was carried out, which consists of testing the classification with *random forest* with all 27 subjects but with a classifier per imagined word. For this, the 23 repetitions of each imagined word were used separately and the experiment was repeated for the 5 words from the dataset of EEG signals. The classification for each of the imagined words was done with the two feature extraction ways used in the previous experiment (Instantaneous and Teager) in order to compare their strengths. The results obtained in the classification step with 10-fold cross-validation using *random forest* are shown in figure 9.

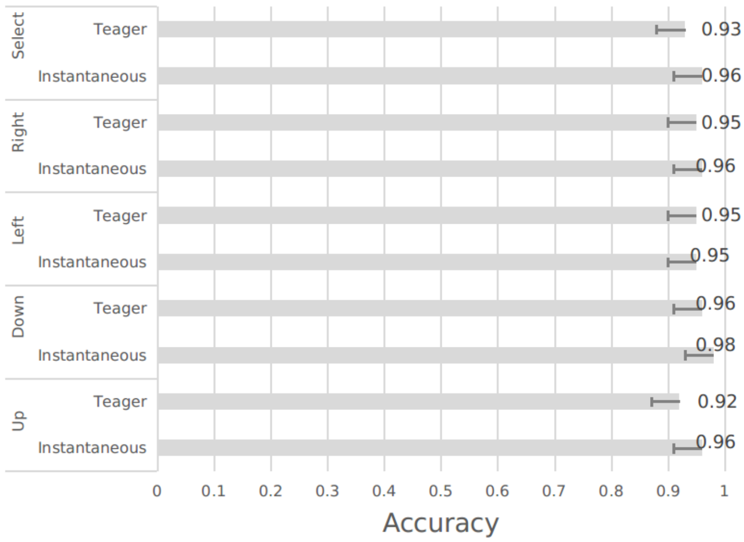

Fig. 9. Accuracies obtained when classifiers for each word separately were created with 27 subjects using 10-fold cross-validation.

Figure 9 again shows that when using instantaneous energy, the accuracy is higher. The highest accuracy is obtained when using the imagined word *Down*. From this, it can preliminarily be asserted that the most suitable word for the task of subjects identification is the imagined word *Down*. However, when using the other imagined words, the results are not much different and in all cases they are above 0.92 of accuracy.

For subjects identification task, the models saved for each imagined word was when the instantaneous energy distribution was used in all cases. As in the proposed was described, for the test stage was used the 30% of the instances (In this case 270 instances per experiment; corresponding to 10 imagined words per each of 27 subjects) for on-line environment. The average accuracy per imagined word was **0.92, 0.91, 0.92, 0.95** and **0.90** respectively.

4 Discussion and Conclusions

I this work, EEG-recorded dataset using the neurological source *imagined speech* were used to create machine learning-based models. To use the method in real-time, an API using Django in Python was designed and created.

The accuracy obtained for on-line environment, is highly related to the model created. In the best-case the average accuracy obtained with the *Test Set* was 0.95. This accuracy level was obtained in the subjects identification task using the imagined word *Down*.

The time for training and saving a machine learning model is not important since it is an off-line process. Using the API for real-time classification, the time to process a new unlabeled entry depends on the size of the signal and the DWT computational complexity ($\mathcal{O}(N \log_2 N)$) [21], however for fast responses it is necessary to implement techniques to distribute the memory/work.

In general, as the experiments in this paper report, EEG signal as a new way of communication can be possible and the work to process new signals can be separated using an API (useful for several predictions at the same time). In addition, EEG signals can be used as a password or as a measure for a biometric security system for several environments. The first experiment for subjects identification shows that subjects can be identified independent of the imagined word. This suggests that it is possible to use a classifier to detect subjects and then use the classified imagined word as a control command (using the API only with other model_id corresponding to the application) in a real application. For example, it can be used as a 2-step verification or to send 2 or more commands at the same time (i.e Give access and call the police, Give access and turn on the lights, etc), in summary, for domestic security applications.

Distributing the work in an API has benefits and will contribute to the use of a single machine learning model for several applications, for example, for subjects identification, the same model could be used to manage the access in 2 or more places and for several users and different tasks. In addition, it could accelerate the process of classification because there are lower restrictions when using computers (I.e High Performance Servers and Supercomputers).

Future research efforts will be dedicated to explore the extent to which specific channels can provide more information for these task in order to reduce the number of channels for real-time applications and decrease the time for a new unlabeled entry. In addition, implementation in a real environment (with additional noise) and with new feature extraction techniques will be tested.

Acknowledgments. This work was done under partial support of CONACYT (scholarship #591475), and the project "David versus Goliath: single-channel EEG unravels its power through adaptive signal analysis - FlexEEG" which is supported by Enabling Technologies - NTNU.

References

1. Bashashati, Ali, Mehrdad Fatourechi, Rabab K. Ward, and Gary E. Birch: A survey of signal processing algorithms in braincomputer interfaces based on electrical brain signals. Journal of Neural engineering **4**, no. 2 (2007): R32.
2. Wolpaw, Jonathan R., Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan: Brain-computer interfaces for communication and control. Clinical neurophysiology **113**, no. 6 (2002): 767-791.
3. Desain, Peter, Jason Farquhar, Pim Haselager, Christian Hesse, and R. S. Schaefer: What BCI research needs. In Proc. ACM CHI 2008 Conf. on Human Factors in Computing Systems (Venice, Italy) (2008):.
4. Elman, Lauren B., and L. McCluskey: Clinical features of amyotrophic lateral sclerosis and other forms of motor neuron disease. Up-to-date. Waltham: Wolters Kluwer Health (2012): 23.

5. Feller, T. G., R. E. Jones, and M. G. Netsky: Amyotrophic lateral sclerosis and sensory changes. Virginia medical monthly **93** , no. 6 (1966): 328.
6. Jayaram, Vinay, Morteza Alamgir, Yasemin Altun, Bernhard Scholkopf, and Moritz Grosse-Wentrup: Transfer learning in brain-computer interfaces. IEEE Computational Intelligence Magazine **11**, no. 1 (2016): 20-31
7. Moctezuma, Luis Alfredo: Distinción de estados de actividad e inactividad lingüística para interfaces cerebro computadora. Thesis project of master degree (2017):.
8. Brigham, Katharine, and BVK Vijaya Kumar: Subject identification from electroencephalogram (EEG) signals during imagined speech. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on (2010): 1-8
9. Jain, Anil K., Arun Ross, and Salil Prabhakar: An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology **14**, no. 1 (2004): 4-20.
10. Jain, Anil K., Arun Ross, and Umut Uludag: Biometric template security: Challenges and solutions. Signal Processing Conference 13th European IEEE (2005): 1-4.
11. Uludag, Umut, and Anil K. Jain: Attacks on biometric systems: a case study in fingerprints. In Security, Steganography, and Watermarking of Multimedia Contents VI, vol. 5306, International Society for Optics and Photonics (2004): 622-634
12. Fielding, Roy, Jim Gettys, Jeffrey Mogul, Henrik Frystyk, Larry Masinter, Paul Leach, and Tim Berners-Lee: Hypertext transfer protocol–HTTP/1.1. No. RFC 2616. (1999):.
13. Torres-García, A. A., C. A. Reyes-García, L. Villaseñor-Pineda, and J. M. Ramírez-Cortís: Análisis de señales electroencefalográficas para la clasificacin de habla imaginada. Revista mexicana de ingeniería biomédica **34**, no. 1 (2013): 23-39.
14. Jasper, Herbert: Report of the committee on methods of clinical examination in electroencephalography. Electroencephalogr Clin Neurophysiol **10** (1958): 370-375.
15. Bertrand, O., F. Perrin, and J. Pernier: A theoretical justification of the average reference in topographic evoked potential studies. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 62, no. 6 (1985): 462-464.
16. Alhaddad, Mohammed J: Common average reference (car) improves p300 speller. International Journal of Engineering and Technology **2**, no. 3 (2012): 21.
17. Didiot, Emmanuel, Irina Illina, Dominique Fohr, and Odile Mella: A wavelet-based parameterization for speechmusic discrimination. Computer Speech & Language **24**, no. 2 (2010): 341-357.
18. Moctezuma, Luis Alfredo, Maya Carrillo, Luis Villaseñor Pineda, and Alejandro A. Torres García: Hacia la clasificación de actividad e inactividad lingüistica a partir de senales de electroencefalogramas (EEG). Research in Computing Science **140** (2017): 135-149.
19. Jabloun, Firas, and A. Enis Cetin: The Teager energy based feature parameters for robust speech recognition in car noise. In Acoustics, Speech, and Signal Processing. 1999 IEEE International Conference on, vol. 1. (1999): 273-276
20. Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al.: Scikit-learn: Machine learning in Python. Journal of machine learning research **12**, no. Oct (2011): 2825-2830.
21. Averbuch, Amir Z., and Valery A. Zheludev: Construction of biorthogonal discrete wavelet transforms using interpolatory splines. Applied and Computational Harmonic Analysis **12**, no. 1 (2002): 25-56.

View publication stats