Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **4**, July **2016**

Recognition of Spanish Vowels through Imagined Speech by Using Spectral Analysis and SVM

Diego A. Rojas, Olga L. Ramos

Mechatronics Engineering
Nueva Granada Military University
Carrera 11 101-80, Bogotá-Colombia
u1801620@unimilitar.edu.co; olga.ramos@unimilitar.edu.co

Jorge E. Saby

Department of Engineering Ph.D
Universidad Distrital Francisco José de Caldas
Carrera 8 40-62, Bogotá-Colombia
jesabyb@udistrital.edu.co

Received November, 2015; revised May, 2016

Abstract. *Recent works have been studied the possibility of develop a communication system based on EEG signals as a tool for information transmission in people with disabilities. In this work the results of acquiring and analyzing EEG signals related to the communication process in the human being are presented, all this with the aim to identify two vowels of the Spanish language by using imagined speech. In first place, the acquisition of EEG signals through BCI devices was performed, the next step was the treatment of the signals with DSP techniques such as filtering, and the Blackman-Tukey transform, also a dimensionality reduction method known as Symbolic Aggregate Approximation was used for training the Support Vector Machines. The developed algorithm is able to classify and recognize the signals from the thinking of two vowels with an accuracy of 85.29% using six of the fourteen BCI's sensors that measure the Brodmann areas related with the language process*

Keywords: Imagined Speech; EEG signals; Pattern recognition; Supported vector machines; Symbolic Aggregate Approximation; EMOTIV Epoc ®

1. **Introduction.** Humans, in their daily lives, needs to communicate for interacting with other humans and behave appropriately in a particular field, such as social, family or work. For social field, communication has three main functions, the first is transmission, which refers to share information from one generation to another, the second is the correlation, that is the ability to relate an object with its meaning for use it correctly, and finally the survival, through the disclosure of nearby threats or dangers. Communication is a process that can be divided into two categories, verbal, requiring the use of conventional linguistic signs for a specific language, it can be oral or written [1], and nonverbal, which is to send information by using gestures or signs [2]. Nevertheless, due to the nature of speech, this is a mechanical process in terms of sound production and for this reason part of the population cannot perform properly this process because of disabling diseases, like amyotrophic lateral sclerosis, locked-in syndrome and spinal muscular atrophy, which affect gradually the motor skills, but cognitive skills remain intact [3]. Based on the above, methods for allowing communication among subjects with these kind of diseases

have been developed, and the regular factor in these methods is the acquisition of electrical and physical signals from the brain. Studies that measure blood pressure in the brain, like [4] for the thinking of two commands 'yes' or 'no' [5], have been realized, but most of them uses electrical signal captured by electroencephalography (EEG) or electrocorticography (ECG) at the moment of thinking in movements or representative symbols of any language. One of the concepts that appears as a possible solution to this problem is imagined speech, which says that is possible to record the electrical activity in the brain at the moment of listening our thoughts without the need to produce a sound, to check this some studies have been performed, as presented in [6], where ECG signals were used for achieve the classification and identification of vowel and consonant sounds of English language, with a success rate of 43% and 38% respectively. Nowadays, the acquisition of this type of signals is less complicated due to technological developments like BCI devices [7], for measuring the electrical impulses generated in the brain cortex in a noninvasive way and without the necessity of specialized medical equipment. With the arrival on the market of BCI devices, several works with different techniques have been made, some of which use classification techniques based on artificial intelligence, for example [8] where achieves an average rate of recognition equal to 78%, by using support vector machines and spatial common patterns, and in [9] achieves an average rate of recognition equal to 78% by using support vector machines and spatial common patterns, and in 9, they used neural networks and statistical data of 13 subjects for classifying the signals from the thinking of five English alphabet vowels, reaching a recognition rate of 44%. In [10], SVM were used again, but this time along with adaptive collections for achieving an identification rate of 85% by using 63 electrodes. Finally in [11] were used two BCI devices, one based on EEG signals, and the other with fMRI technology for classifying the response of 20 subjects, when they listened a pre-selected sentence, and as a main result a classification rate of 90% was achieved in both cases. In this work are presented the results of a pattern recognition algorithm with support vector machines and symbolic aggregate approximation capable of classify and recognize the signals from the imagined speech of two Spanish vowels, specifically 'A' and 'E', with an average rate of 85.29%, using 6 of the 14 sensors that has the BCI device EMOTIV Epoc ®, concluding that is possible to identify vowels through the thinking with low cost BCI equipment. The first part of this document is about the imagined speech and the methods used for its quantification, like electroencephalography, electrocorticography and functional magnetic resonance imaging, showing several works that use different artificial intelligence techniques for recognizing the patterns from the thinking of distinctive vowels or signs for a specific language. The second part has the description of the material and methods used for this research. In third part are shown the results obtained with the feature extraction stage and the implementation of support vector machines. The last part contains the conclusions of this work and future perspectives for keeping the researches in this field.

2. Material and Methods.

2.1. **Brain-Computer Interface.** Brain-computer interface are devices that have the required hardware, firmware and sometimes software for allowing the communication and sending orders to a computer by using the signals from brain activity. Typically, these systems have three stages, the first is the acquisition and conditioning of bio electrical signals generated in the brain cortex, the second is related to the processing and feature extraction of acquired signals, and the last one is about the commands interpretation by the computers. For this work were used 6 of the 14 sensors that have the headset EMOTIV

Epoc ®, which are distributed according to the 10-20 EEG System as is depicted in figure 1.

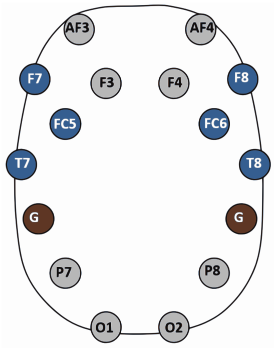

FIGURE 1. Distribution of EMOTIV Epoc ® sensors

2.2. **Acquisition and Signal Processing.** Samples of the electrical stimulus produced in the brain cortex at the moment to think in any of the chosen vowels, a or e, were taken in quiet environment, without distraction and supported visually with the use of white screens for neutral pattern and the image of the letter for the vowels. In the figure 1 can be appreciated the chosen electrodes highlighted with blue.

TABLE 1. Chosen electrodes and the Brodmann Areas that each one measures

Electrodes		Broadmann Areas	Main Function
Left	Right		
F7	**F8**	45	Semantic generation [12]
FC5	**FC6**	44	Phonological and syntactic process [13]
T7	**T8**	21	Speech production [14]

Table 1 summarizes the nomenclature of the sensors according to the 10-20 EEG system, and also has a brief description of each Brodmann area that is measured. EEG signals were acquired using a sample frequency of 128, because these signals have a frequency range between 0.5 Hz to 64 Hz as is observed in the table 2.

TABLE 2. Brain waves and its frequency bands.

Wave	Frequency (Hz)
Delta	0.5 to 3.5
Theta	4 to 7
Alpha	8 to 12
Beta	13 to 30
Gamma	31 to 100

Since speech is considered a conscious process, beta waves were selected for the study. An IIR Butterworth pass band filter, as the shown in equation 1, with pass bands located in 13 and 30 Hz was used to obtain information from these frequency bands. The figure 2 shows an example of the acquired signals and its behavior after the filtering process.

$$G(\omega) = \sqrt{\frac{1}{(1 + (\omega)^{2n})}} \quad (1)$$

Here, w is the angular frequency and n is the filter order.

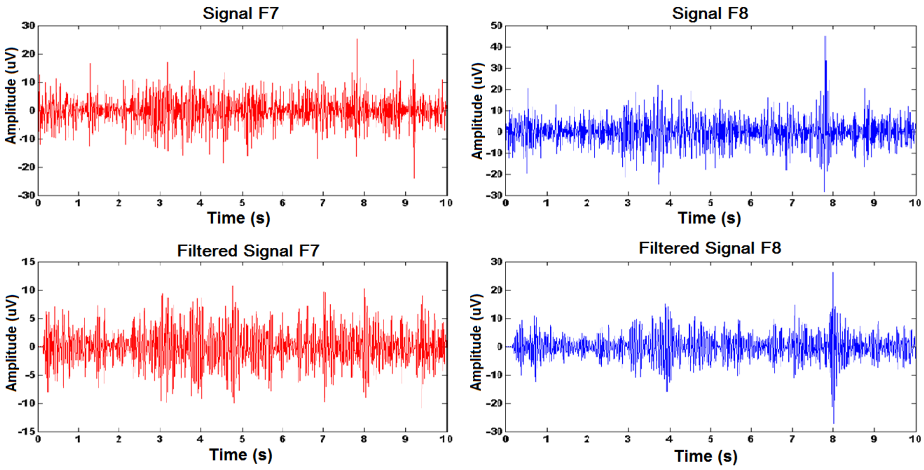

FIGURE 2. Signals acquired with F7 and F8 electrodes and its filtered version

After the filtering step, was used the Blackman-Tukey transform [15], as is presented in equation 2 ,for analyzing the signals in the frequency domain and obtain data of each one, and eliminate problems related to data length, because some of the samples were captured with different time intervals.

$$P_x(f) = \sum_{k=0}^{M-1} \omega_k r_k e^{-i\omega\tau} \tag{2}$$

Here, r is the auto correlation of the signal, w is a Hann window, f is the frequency, M is the maximum lag to be considered and k is the increase of the lag.

2.3. **Feature Extraction.** The method used for feature extraction and dimensionality reduction was Symbolic aggregate approximation [16]. The method used for feature extraction and dimensionality reduction was Symbolic aggregate approximation. This method consist on to divide the vertical axis in many parts as be required with the aim of giving a symbolic value of each section or division as can be seen in figure 3.

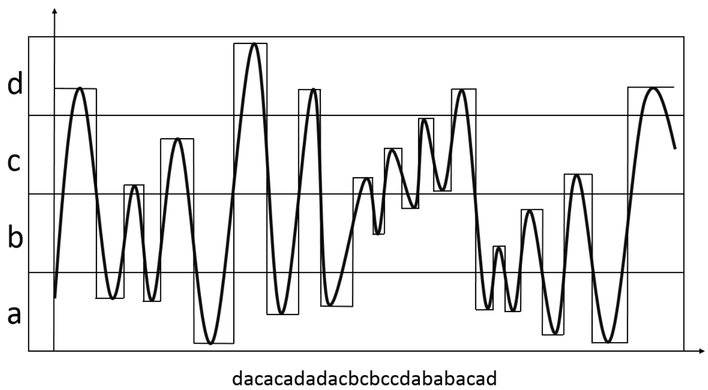

FIGURE 3. Signal partition by using SAX method

Once the signal has been segmented and the string of symbols that represents the signal has been obtained, we should compute the matrix with the frequency of each symbol. As a visual resource, the mapping of the data matrix can be performed using a color scale in order to observe the behavior of the data, an example of the above is shown in figure 4 where a gray-scale map was used.

aa (0)	ab (2)	ac (4)	ad (3)
ba (2)	bb (0)	bc (2)	bd (0)
ca (3)	cb (2)	cc (1)	cd (0)
da (4)	db (0)	dc (0)	dd (0)

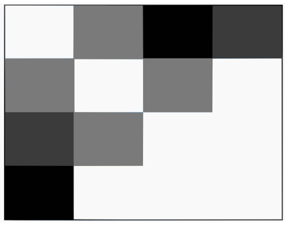

FIGURE 4. Frequency matrix and its gray-scale mapping.

2.4. **Support Vector Machines (SVM).** SVM is a supervised learning algorithm commonly used for pattern recognition tasks. These algorithms use a set of patterns labeled according to the group or class to which they belong in order to train classification models with the aim of predicting the category of new patterns. The training of the model seeks to find the hyperplane or group of hyperplanes that separates the classes with the widest margin, resulting in a better generalization of patterns. For developing this work were used non-linear SVM with radial function, as presented in equation 3, for the machine's kernel.

$$K(x, x') = e^{\frac{|x-x'|}{2\sigma^2}} \tag{3}$$

However, an SVM is designed to separate correctly a maximum of two classes, but on a practical level it is more common to find multiple classification problems that involve more than two classes. To solve these problems there are different methodologies, one of these is called *1vsRest* [17], that consist in to program the same amount of SVM's as classes, when the *k-th* model *Yk(x)* is trained using the data of class *Cn* as positive cases and the remaining data as negative cases. Other option is to use the *1vs1* method, in which are programmed *K(K-1)/2* SVM and then the task of identification is performed depending on the number of "votes" that the pattern obtains from the SVM's, however this method increases the computational cost.

3. **Results.** As the main result a classification algorithm based on support vector machines with the ability to identify a pattern signal and two signals from the thinking of vowels of Spanish language was obtained. The SAX method was used for reducing the size of data and decrease the cost of training for SVM, obtaining the results shown in Figures 5 and 6.

In figure 5 was used the SAX technique with a partition of four symbols, showing the difference between the recorded control patterns and signals generated by the thinking of a vowel, specifically /a/

Moreover, in figure 6 six symbols were used, and can be seen how the frequency matrix changes if the partition is reduced or increased, this is helpful if we want to describe with more detail each pattern, but for our work was enough with a SAX based on six divisions. To implement the identification task with the support vector machine algorithm was selected the 1vsRest method explained above, and due to we have three classes, three SVM's were programmed as it can be seen in table 3, one SVM for the neutral pattern and the others two for identifying the signals generated by the thinking of vowels /a/ and /e/ respectively.

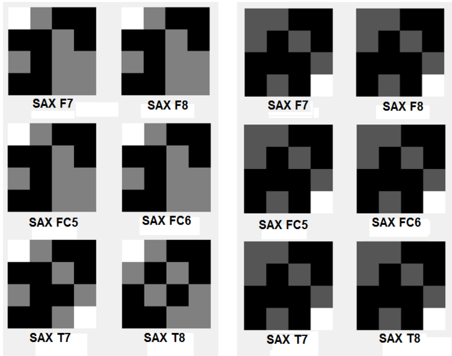

FIGURE 5. SAX with 4 symbols for neutral pattern (left) and vowel A pattern (right).

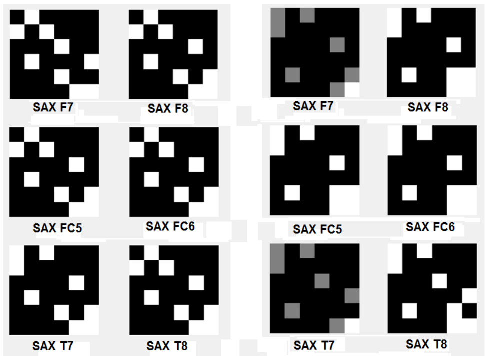

FIGURE 6. SAX with 6 symbols for neutral pattern (left) and vowel A pattern (right).

TABLE 3. Percentage of class separation according to the amount of samples acquired.

#Samples	**Class 'N'**	**Class 'A'**	**Class 'E'**
3	56.1%	47.2%	52.6%
6	65.3%	58.4%	58.4%
9	78.1%	73.7%	70.6%
11	80.1%	84.5%	82.2%
12	85.6%	81.2%	89.2%
15	85.6%	85.8%	84.4%

In the table 3 are summarized the classification values in terms of the samples per class that were used, and can be observed that the separation was more accurate in the SVM trained with 12 or more samples. The recognition phase was divided into five stages, the first one is related to the acquisition of new samples from the thinking of any of the vowel that we want to identify, the second part consists of perform the same feature extraction techniques that were used with the training samples, in third stage we have to test or evaluate the support vector vector machines with the features extracted recently, the fourth is about the computation of the SVM's responses and the last stage is the interpretation of the outputs in order to achieve a correct classification.

Taking into account that SVM's are supervised learning algorithms, we must have knowl- edge about the class of each machine represents, i.e. if the pattern belongs to the

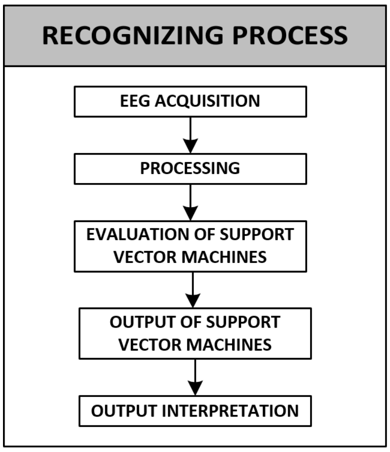

FIGURE 7. Example of figure

neutral category, the SVM located in first place have to respond positively, and the others must give negative values. An example of the SVM response to new pattern that was not taken into account for training is presented in Table 4.

TABLE 4. My caption

Signal	**Electrodes**						**Labels**
	F7	**F8**	**FC5**	**FC6**	**T7**	**T8**	
Neutral	1	1	1	1	1	1	**Neutral**
	-1	-1	-1	**1**	-1	-1	**A**
	-1	-1	**1**	-1	-1	-1	**E**
A	-1	-1	-1	**1**	-1	-1	**Neutral**
	1	1	1	**-1**	**-1**	1	**A**
	-1	-1	-1	**1**	-1	-1	**E**
E	-1	-1	**-1**	-1	-1	-1	**Neutral**
	-1	-1	-1	1	-1	-1	**A**
	1	1	-1	1	1	1	**E**

In table 4 are shown the responses obtained with the tests realized, and are highlighted the wrong answers. As identification parameter, was used the mode of the answers obtained with the testing stage for the SVMs, e.g. in the case of the signal produced by the thinking of the vowel "e", we could say that the signal belongs to 'e' label due to 5 of 6 signals evaluated were positive.

4. **Conclusions.** SVM are classifier algorithms that perform a class separation with equitable margins for each one, resulting in improved patterns generalization, compared to others artificial intelligence techniques for the same task, like neural networks, but has an important limitation in the amount of classes that each SVM can separate.

Although there is a measurable change between the two vocal signals, this is very small, which causes confusion in the decision obtained from support vector machines in the recognizing of patterns.

One of the greatest difficulties to identify a linguistic sign through EEG signals is that the subject who is providing signals must remain concentrated and focused on the sign language that he wants to send, for this it is necessary the use of visual stimuli, and even hearing stimuli to keep the thought while data acquisition is performed.

The sensor with more incorrect measurements was FC6 and the next in order of errors was FC5, that could be due to the brain activity in these areas is related to the syntax, i.e. the elaboration and coherence of utterances, which is one of the most complex processes involved in speech production.

SVM react positively to the patterns extracted from the signals F7 and F8, these sensors measured Brodmann areas related to semantics, its mean the explicit meaning that has a linguistic sign.

As a future perspective is proposed the use of different signal processing techniques to achieve feature extraction, distinctive enough to classify a large number of linguistic signs with a high recognition rate.

Acknowledgment. Special thanks to the Research Vice-rectory of the "Universidad Militar Nueva Granada, for financing the project ING/INV 1762 titled Dispositivo reproductor de voz del lenguaje español a través de habla subvocal e interfaz cerebro-computador"project, 2015 year. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers,which have improved the presentation.

REFERENCES

[1] J. Knox, Visual-verbal communication on online newspaper home pages, *Vis. Commun.*, vol. 6, no.1, pp. 19-53, 2007.

[2] M. Knapp,J. Hall, and T. Horgan, Nonverbal Communication in Human Interaction, *Eighth Edi. Wadsworth*, 2013.

[3] C. Lomen-Hoerth, J. Murphy, S. Langmore, J. H. Kramer, R. K. Olney, and B. Miller, Are amyotrophic lateral sclerosis patients cognitively normal, *Neurology*, vol. 60, no. 7, pp. 1094-1097, 2003.

[4] J. Niessing, B. Ebisch, K. E. Schmidt, M. Niessing, W. Singer, and R. A. W. Galuske, Hemodynamic signals correlate tightly with synchronized gamma oscillations, *Science*, vol. 309, no. 5736, pp. 948-951, 2005.

[5] M. Naito, Y. Michioka, K. Ozawa, Y. Ito, M. Kiguchi, And T. Kanazawa, A Communication Means for Totally Locked-in ALS Patients Based on Changes in Cerebral Blood Volume Measured with Near-Infrared Light, *IEICE Trans. Inf. Syst*, vol. E90-D, no. 7, pp. 1028-1037, 2007.

[6] X. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans, *J. Neural Eng*, vol. 8, no. 4, p.28-46, 2011.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. furtscheller, T. M. Vaughan,Brain-computer interfaces for communication and control, in *Clin. Neurophysiol*, vol. 113, no. 6, pp. 767-91, 2002.

[8] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, Analysis and classification of speech imagery EEG for BCI, *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 901-908, 2013.

[9] R. Kamalakkannan, R. Rajkumar, M. M. Raj, S. S. Devi, Imagined Speech Classification using EEG, *Adv. Biomed. Sci. Eng*, vol. 1, no. 2, pp. 20-32, 2014.

[10] M. Matsumoto and J. Hori, Classification of silent speech using adaptive collection, *2013 IEEE Symposium on Computational Intelligence in Rehabilitation and Assistive Technologies (CIRAT)*, vol. 1, pp. 5-12, 2013,

[11] Y. He, H. Gebhardt, M. Steines, G. Sammer, T. Kircher, A. Nagels, and B. Straube, The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech, *Neuropsychologia*, vol. 72, pp. 27-42, 2015.

[12] B. T. Gold, D. A. Balota, B. A. Kirchhoff, and R. L. Buckner, Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from FMRI adaptation, *Cereb. Cortex*, vol. 15, no. 9, pp. 1438-50, 2005.

[13] S. Heim, S. B. Eickhoff, and K. Amunts, Specialization in Brocas region for semantic, phonological,and syntactic fluency?, *Neuroimage*, vol. 40, no. 3, pp. 1362-8, 2008.

[14] S. Brown, M. J. Martinez, and L. M. Parsons, Music and language side by side in the brain: a PET study of the generation of melodies and sentences, *Eur. J. Neurosci*, vol. 23, no. 10, pp. 791-803, 2006.

[15] A. Ahmad, F. S. Schlindwein, Comparison of computation time for estimation of dominant frequency of atrial electrograms: Fast fourier transform, blackman tukey, autoregressive and multiple signal classification, *J. Biomed. Sci. Eng*, vol. 03, no. 09, pp. 843-847, 2010.
[16] H. Tayebi, S. Krishnaswamy, A. B. Waluyo, A. Sinha, and M. M. Gaber, RA-SAX: Resource aware symbolic aggregate approximation for mobile ECG analysis, *Proc.- IEEE Int. Conference on Mobile Data Management*, vol. 1, pp. 289-290, 2011.
[17] C. Weinrich, C. Vollmer, and H. M. Gross, Estimation of human upper body orientation for mobile robotics using an SVM decision tree on monocular images, *IEEE -RSJ Int. Conference on Intelligent Robots and Systems*, pp. 2147-2152, 2012.