Revista digital de Semilleros de Investigación REDSI
ISSN: 2711-3892

TRANSFORMATION OF ELECTROMIOGRAPHIC SIGNALS OF SUBVOCAL SPEECH USING COMPRESSIVE SENSING AND ARTIFICIAL INTELLIGENCE

J. Daniel Ramírez *[a], Luis E. Mendoza[b], Leonardo A. Carrascal[a]
[a] University of Pamplona (Pamplona-Norte de Santander, Colombia). daniel.ramirez@unipamplona.edu.co.; [b] University of Pamplona (Pamplona-Norte de Santander, Colombia). luis.mendoza@unipamplona.edu; [a] University of Pamplona (Pamplona-Norte de Santander, Colombia). ingleonardocarrascal@gmail.com;

ABSTRACT

This article shows the acquisition of electrical signals from the nerves of the throat and oral cords, transforming them into voice signals using advanced techniques in digital signal processing and artificial intelligence, based on the extraction of patterns based on compressive sensing, Entropy, discrete Wavelet transform and a classifier based on vector support machines of least squares, once the system has been calibrated, the results showed that 95% + - 0.34 of data were correctly classified. The developed system was used in 500 signals and is based on an Open Source programming language implemented in an embedded system. Finally, it was shown that it is possible to use compressive sensing to extract subvocal speech patterns.

Keywords: Sparse, Python, electromyography multi - resolution analysis, Raspberry Pi.

1. INTRODUCTION

Subvocal speech is the recording and interpretation of the bioelectric signals that control the vocal cords and the tongue during the process of verbal communication, these signals contain important information that is related to words that we perceive auditively. In addition, subvocal speech signals are registered without the need for the physical production of sound by the individual.

At present, noisy environments and physiological pathologies are problems that cause, communications are affected or information is distorted or simply never generated. That is why the design of silent or subvocal speech interfaces [1], starting from the acquisition process for the recognition of phonemes, syllables, words or complete sentences in order to transmit a message that one intends to say without actually arriving to be expressed as an acoustic signal [2], poses an effective solution to these problems and could also be immersed in areas such as interpersonal communication, robot control, communication in industrial environments, underwater communication, highly confidential military information exchange, Mobile communication, control of transportation systems for people with motor problems, help people with speech pathologies and rescue operations [3] [4] [5].

Different studies have been carried out on the applications of subvocal speech, among of the most important are: those made by NASA's Ames Research Center. However, the existence of subvocal speech has been studied for decades, in 1969 Curtis & Lewis; analyzed various sub vocalization patterns, recording the signals by surface electromyography using electrodes located opposite the thyroid cartilage. Also, there are currently methods to extract subvocal speech signal patterns [6] [7]. This work, as in previous studies, signals from the electrodes were acquired on the thyroid cartilage by non-invasive electromyography with surface electrodes arranged in the throat in differential mode the extraction of characteristics was performed using low - density or sparse representations of signals based on the time - frequency plane using tools such as the discrete wavelet TWD and compressive sensing - CS a classifier based on vector support machines

ISSN: 2711-3892

of minimum squares - SVM implemented in the Raspberry Pi - RPi system [8] and implemented under the Python programming environment .

METODOLOGY

The words reproduction from subvocal speech signals is summarized in non-invasive electromyographic acquisition, processing and pattern extraction, and the training of an intelligent classifier through supervised training that is able to translate these patterns into previously recorded sounds in the system.

Subvocal speech acquisition

A differential configuration of detection electrodes (silver / silver chloride - Ag / AgCl) arranged on the surface of the neck. In addition to a reference electrode located in the bone Mastoid behind the ear Figure. 1. The signals from the electrodes were acquired through a previously designed system that debugged non-relevant information and sent them to an embedded system for further processing.

Figure 1. Positioning of the electrodes for subvocal speech detection. Source: Authors.

The acquisition system

The acquisition system consists basically of three fundamental stages, in the first stage the instrumentation amplifier INA 128 amplifies the difference of the two signals from the electrodes located in the neck to a gain of 1000 Figure1. In the second stage of the system an active analog filter is implemented Butterworth passes band of eighth order to a set bandwidth for electromyographic signals - EMG of 30-500Hz. Finally, the purified signal is encoded using the Nyquist sampling theorem. Sampling the signal at a frequency of 2kHz by implementing the digital analog converter ADS7813.

Raspberry Pi System

Signals were recorded in real time with the implementation of the Serial Peripheral Interface (SPI) between the developed acquisition system and the RPi system, where in the latter they were digitally stored and conditioned by mathematical techniques implemented under the Open Source platform of Python [8].

Pattern conditioning and extraction

The EMG signals from the acquisition system were stored in vectors of length 20,000 points in the RPi Figure. 2. (A) A methodology was established to detect the area of interest (the area where the most relevant information of the signal is

Revista digital de Semilleros de Investigación REDSI
ISSN: 2711-3892

located), The entropy value (1) was calculated in fixed-length windows of 100 points Si, (2) of the signal, the entropy values of each of the windows were stored in a vector of length of 200 points Figure 2. (B).

$$E = \sum_{i=1}^{n} {S_i}^2 \times \mathrm{LOG}_2 {S_i}^2 \quad (1)$$

and

$$S_i = X_I \times \varphi(n) \quad (2)$$

Where:

- $\boldsymbol{\varphi(n) = A(t - k_i)}$ Amplitude one unit step function.
- X_i= Electromyographic signal
- S_i= window

Duration $\boldsymbol{\varphi(n)}$ is defined as:

$$\boldsymbol{{k_1}^i \leq t \leq {k_2}^i}$$

Where:

- $\boldsymbol{{k_1}^i}$ = {0, 100, 200, 300… , L-100}
- $\boldsymbol{{k_2}^i}$ = {100, 200, 300, 400…,L}
- i= 1, 2, 3, 4…, L/n, L (number of signal points)

With the trace of a threshold (3), we identified the starting point where the relevant energy Figure. 2. (B) is concentrated. From the start point of the entropy vector we multiply by the value of the window Si, for Identify the starting point of the zone of interest from the original signal Figure. 2 (A), by cutting a window of 8000 points length, thus limiting the greater information of the signal Figure 2 (A).

Revista digital de Semilleros de Investigación REDSI
ISSN: 2711-3892

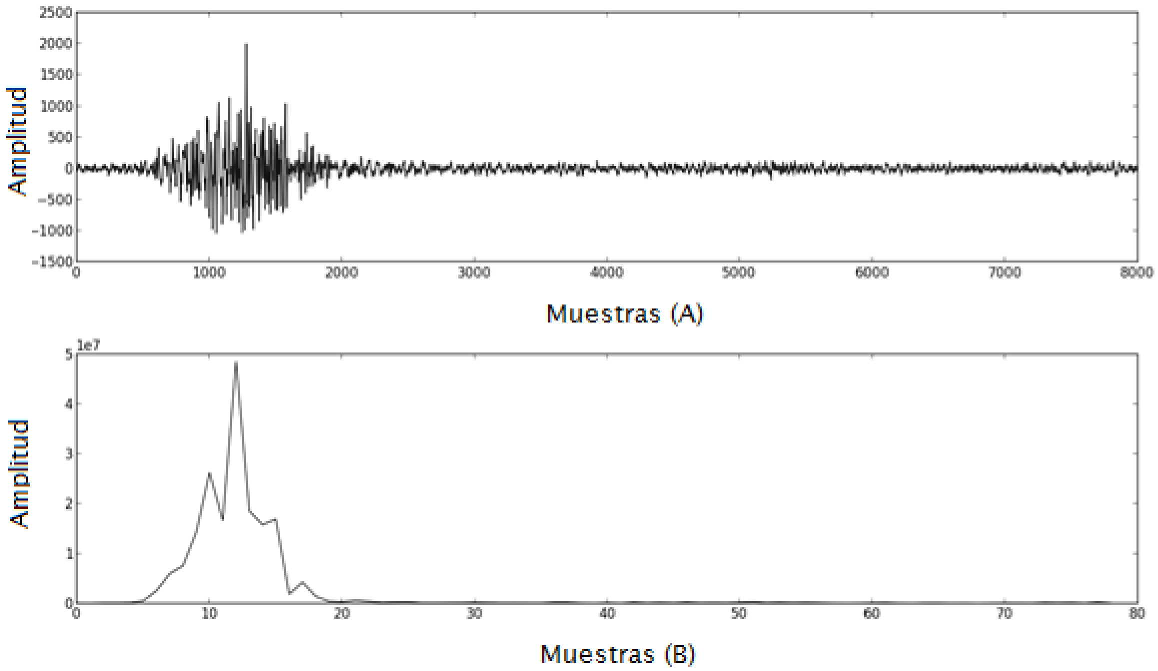

Figure. 2. (A) Electromyographic signal. (B) Entropy signal. Source: Author.

$$U = (0.5) \times A_{max} \tag{3}$$

Where:

- U= Detection threshold.
- A_{max} = Maximum peak of the entropy vector of the EMG signal.

Sparse

Identified the area of interest became the EMG signal, in a scattered signal or sparse where it only presents a few nonzero values in the time domain or in some other domain. The number of non-zero values of a scattered signal is known as its scatter level or sparcity. To verify that so sparse is a signal, the norm norm calls L1 (4) is the one used.

$$\|X\| = \sum_{i=1}^{n} |X_i| \tag{4}$$

where:

- X_i = Points that make up the signal.

The DWT (5) allowed to make a multiresolution analysis at 12 levels of decomposition of the coefficients of detail of the EMG signal. With a Wavelet mother Daubechies db4 signals of the area of interest were plotted in sparse signals Figure. 3.

$$C_{a,b} = \sum_{a,b} X(n)\, \varphi\left[\frac{n-a}{b}\right] \tag{5}$$

Where:

- $X(n)$ = Original signal.
- $\varphi\left[\frac{n-a}{b}\right]$ = Mother Wavelet.
- a = Translation coefficient.
- b = Scaling coefficient.

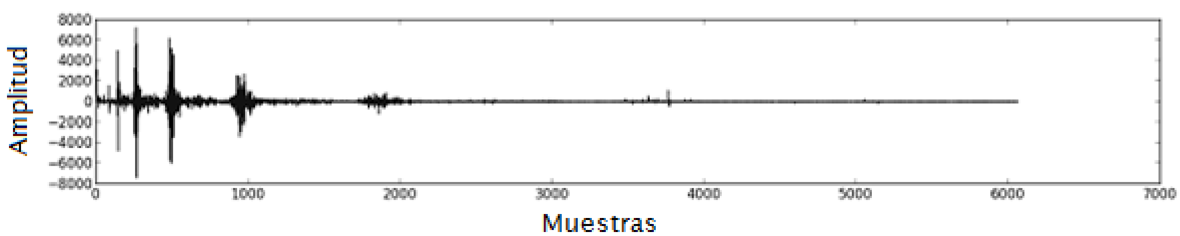

Figure. 3 Conversion to signal type sparse, using the discrete wavelet transform.

Compressive sensing

CS can sample a signal (sparse or a compressible representation of a signal) with fewer samples than those required by the Shannon-Nyquist theorem, mathematically CS is defined as:

$$\mathbf{Y} = \mathbf{\Phi X} \tag{6}$$

Where:

- Φ = matrix encoding.
- X = sparse signal (k values differents than zero)
- Y = Result of the product of the K-Sparse values of the signal X

CS Graphically Figure. 4. Modeling (6), defines that a coding matrix Φ multiplied by a sparse column vector X, we obtain the vector column Y which is the compressed representation of the sparse values of the signal X [9]

.

Revista digital de Semilleros de Investigación REDSI
ISSN: 2711-3892

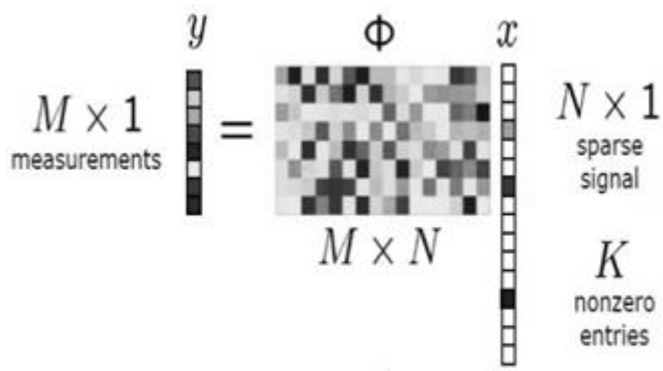

Figure. 4. Sample base Φ, the original signal X and the compressed signal Y.

The minimum value of M of the compression matrix Φ is twice the K - sparse values that exist in the signal [10], where:

$$\boldsymbol{M > K << N} \tag{7}$$

The value M is the number of rows of the compression matrix, testing on each type of signal allows finding the measurements that express the perfect balance between high compression and good reconstruction. For the matrix to be able to compress the vectors X properly, the compression matrix Φ must have 2 very important properties orthogonality (8) and orthonormality, that is to say that each set of its vectors is unitary or each of its vectors their magnitude is equal to one.

$$\boldsymbol{M \times M^T = M^I} \tag{8}$$

where:

- M = Matrix.
- M^T= transposed matrix.
- M_I = Identity matrix

The isometric restriction property - RIP (9) characterizes matrices that are almost orthonormal, operated with sparse type vectors.

$$\boldsymbol{(1 - \delta_S)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_S)\|x\|_2^2} \tag{9}$$

where:

- Constant $\delta_S \in (0,1)$
- A=Φ= Random matrix of compression
- X= Signal k – Sparse.
- $\|x\|_2^2$ =It denotes the L2 norm for vectors that is defined as

$$\|x\|_2^2 = \sqrt{\sum_{i=1}^{N} |X_i|^2} \tag{10}$$

The value that defines the maximum compression of the sparse signal (11) defines the length of the training pattern for the classifier

$$M > con \times \log\left(\frac{L}{con}\right) \tag{11}$$

where:

- L = Signal length sparse EMG.
- con = It is the counting variable of values very different from zero,
- M = It is the value of the maximum compression of the signal.

Having the value of M we can know the minimum value of the columns that we can use for the CS conversion dictionary.

Support Vectorial Machines – SVM

The classifier allows to predict to which class a new signal belongs to the system, training a SVM in the system generated a set of training samples by labeling each one of them by classes to train an SVM to construct a model that predicts the class of a new sample. In training a SVM (12) maps the entry points into a feature space of a larger dimension by finding an optimal Hyperplane using the dot product with functions in the feature space called Kernels. The optimum Hyperplane solution can be written as the combination of a few entry points that are called support vectors. A good separation between the classes will allow a correct classification [7].

$$f(x) = sign\left(\sum_{i=1}^{m} y_i \alpha_i (x.x_i) + b\right) \tag{12}$$

where:

- $(x.x_i)$ = optimal kernel hyperlane
- y_i = Group tags.
- α_i=Lagrange coefficient
- B = Coefficient of normalization.

The training matrix is composed of the patterns obtained after applying CS to standardized sparse signals, the sub-vocals were words like "Hello How are you?, Hello, Intruder and I am Cold," The database consisted of 5 signals for Each pattern composed of M samples, the training matrix was structured from 20 rows (signals) by M columns (samples of each signal). For each training class their respective labeling was assigned generating a vector 20 labels. Stored the training matrix and the label vector in the RPi, the SVM is trained with 100% of the data generated. To finally enter new signals to the system in real time and determine the percentage of correct classification.

Revista digital de Semilleros de Investigación REDSI
ISSN: 2711-3892

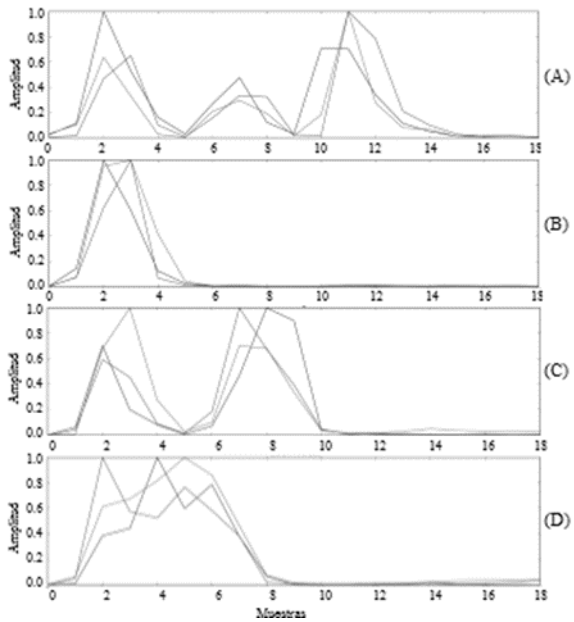

Figure. 5. SVM training patterns for hello, how are you? (A), hello (B), I'm cold (C) and intruder (D). Source: Authors.

RESULTS AND DISCUSSION

The effectiveness test was performed by entering in real time 50 EMG signals of subvocal speech for each of the selected patterns, that is to say that the system was entered 200 signals for the process of validation of the classification in table 1. The Summary of the result obtained in each of the classification applying the techniques mentioned above.

Table 1. Summary percentage of success of each processing technique.

	Technique	Signal	% of success	Quantity of data
1	SC	Hello	0%	
		Hello how are you?	0%	
		I am cold	0%	
		Intruder	100%	
2	SC + N	Hello	100%	
		Hello how are you?	0%	
		I am cold	0%	
		Intruder	0%	
3	SC + ZI + N	Hello	100%	
		Hello how are you?	0%	**50**
		I am cold	0%	
		Intruder	4%	
4	SC + ZI + W1 + N	Hello	100%	
		Hello how are you?	0%	
		I am cold	6%	

		Intruder	5%	
5	SC + ZI + W1 + E400 + N	Hello	100%	
		Hello how are you?	70%	
		I am cold	65%	
		Intruder	80%	

It can be observed in Table 1. That the best classification percentages were obtained by applying techniques to the area of interest of the subvocal signals such as Wavelet to the first level of decomposition using a Wavelet db3 mother, selecting the coefficient of detail, entropy analysis and standardization. The effectiveness of the classification algorithm varies depending on factors such as; Disposition of the electrodes, relative humidity, quality of the sensors, preparation of the skin at the registration site and the person tested.

CONCLUSIONS

The sub-vocal communication system a single EMG channel was able to reduce the computational cost of processing the Raspberry pi, demonstrating that it is possible to communicate two people using subvocal speech signals with a low-cost embedded system and achieving a ranking percentage up to the moment of 66% and 72.5%. In addition, we worked under the free and multi-platform Python programming environment which allows it to migrate to other embedded systems with the same characteristics. Finally, new results were presented and will be used for further research.

BIBLIOGRAPHY

[1] G. E. A. D. y. A. O. Gutierrez J, "Interface developed for the detection of sub-vocal speech," Revista Tecnura, vol. Volumen 17, no. Numero 37, p. paginas 138 – 152, Julio - Septiembre de 2013 .
[2] C. Jorgensen, "Sub auditory speech recognition based on EMG signals," Neural Networks, 2003. Proceedings of the International Joint Conference on, 20-24 July 2003.
[3] O. L. Ramos Sandoval, "Arquitectura Algorítmica para el Reconocimiento de Patrones Fonéticos del Habla Sub-Vocal en el Español," Universidad Distrital Francisco José de Caldas, 28 10 2016. [Online]. Available: http://hdl.handle.net/11349/4473. [Accessed 25 06 2017].
[4] "Web Browser Control Using EMG Based Sub Vocal Speech Recognition," System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on, 6-6 Jan. 2005 Big Island, HI, USA, USA.
[5] P. J. M. L. y. V. H. Mendoza L, "Speech Subvocal Signal Processing using Packet Wavelet and Neuronal Network.," TecnoLógicas, Vols. ISSN 0123-7799 Edición EspeciaL, pp. pp. 655-667, octubre de 2013.
[6] C. D. Hardyck and L. F. Petrinovich, "Treatment of Subvocal Speech During Reading," Journal of Reading, vol. 12, no. 5, pp. pp. 361-368., 02 - 1969.
[7] "One Channel Subvocal Speech Phrases Recognition Using Cumulative Residual Entropy and Support Vector Machines," IEEE Latin America Transactions, vol. Volume: 13, no. Issue: 7, pp. 2135 - 2143, July 2015.
[8] J. D. &. M. L. E. Ramírez-Corzo, "Dual silent communication system development based on subvocal speech and Raspberry Pi.," Revista Facultad de Ingeniería,, vol. 25 (43), pp. 111-121, 2016.
[9] L. Mendoza and L. M. Meriño, "COMPRESIÓN ROBUSTA USANDO COMPRESSIVE SENSING (CS)," Revista Colombiana de Tecnologías de Avanzada, pp. ISSN: 1692 - 7257 , 2009.}
[10] Sergio S Pinto, Luis E Mendoza, Hernando J Velandia, Valentin Molina, Leonor J Cervelon. Compressive sensing hardware in 1-d signals. Revista Tecciencia. 2015.