Electro-myographic patterns of sub-vocal Speech: Records and classification

Patrones Electromiográficos del Habla Subvocal: Registros y Clasificación

Luis Enrique Mendoza, Jesús Peña Rodríguez, Jairo Lenin Ramón Valencia

Abstract

This paper describes the results obtained from recording, processing and classification of words in spoken Spanish by means of analysis of subvocal speech signals. The processed database has six words (forward, backward, right, left, start and stop), In this article, the signals are sensed with surface electrodes (placed on the surface of the throat) and acquired at a sampling frequency of 50 kHz. The signal conditioning consists of a couple of steps, namely the location of area of interest, using energy analysis; and a filtering stage, using Discrete Wavelet Transform. Finally, feature extraction is achieved in the time-frequency domain using Wavelet Packet and statistical techniques for windowing. Classification is carried out with a back propagation neural network whose training is performed with 70% of the database obtained. The correct classification rate was 75%±2.

Keywords: Electromyography, subvocal speech, Wavelet, neuronal network.

Resumen

En este trabajo, se describen los resultados obtenidos a partir de los registros, procesamiento y clasificación de las palabras en el idioma español a través del análisis de las señales de voz subvocal. La base de datos procesada tiene seis palabras (adelante, atrás, derecha, izquierda, arranque y parar), en este artículo, las señales fueron detectadas con electrodos superficiales colocados sobre la superficie de la garganta y adquiridas con una frecuencia de muestreo de 50 kHz. El acondicionamiento de la señal consiste en: la ubicación del área de interés mediante el análisis de la energía, y la filtración usando Transformada Discreta Wavelet. Por último, la extracción de características se realiza en el dominio de tiempo-frecuencia utilizando Wavelet Packet y técnicas estadísticas para ventanas. La clasificación se llevó a cabo con una red neuronal de retropropagación cuya formación se llevó a cabo con 70% de la base de datos obtenida. La tasa de clasificación correcta fue de 75% ± 2.

Palabras clave: Electromiografía, habla subvocal, Wavelet, red neuronal.

Introduction

Sub-vocal speech is the recording and interpretation of bioelectric signals that control the vocal cords and tongue during the verbal communication process; these signals provide information and are matched to the words that we perceive aurally. Furthermore, sub-vocal speech signals are present regardless of the physical production of sound on the part of the individual [2, 3].

At present, noisy environments and physiologic pathologies are two of the big problems that affect the oral communication among individuals in such a way that information is distorted or simply never generated. Communication that does not depend on the physical production of audible signals would offer an effective solution to problems that arise from an increase in the vulnerability of verbal communication.

Sub-vocal speech presents a tentative answer to these limitations because it does not depend on the emission of sound; instead, this type of speech uses electrical signals transmitted to the brain, to the vocal cords and tongue [2, 3]. Various applications emerging form the concept of "silent communication" include control of virtual interfaces, interpersonal communication, robotics control, communication in industrial environments, underwater communication, exchange of highly confidential military information, mobile communication, transportation systems control (for people with motor problems), aids for people with speech pathologies, and rescue operations.

Different works on applications of the sub-vocal speech have been carried out, some of the most important applications include [2], [3] and [4], (studies carried out by the Ames Research Center NASA). However, the existence of sub-vocal speech has been studied for decades; in 1969 Curtis & Lewis analyzed different patterns of sub-vocalization, recording EMG signals using surface electrodes placed on opposite sides of the thyroid cartilage [6]. In the 80s, with the arrival of digital electronics and electronic devices, more and more small, different research groups focused on speech recognition projects using facial and sublingual electromyography. Most studies on the behavior of vocal muscles have been carried out mainly in an attempt to control electronic prosthetics

[5] depending on characteristics such as amplitude and frequency of the electro-myographic signals. The analysis of sub-vocal signals not only deals with processing characteristics like amplitude and frequency, but also with the use of detailed signal information.

An important point in this work is the recording of the electro-myographic signals (EMG) that are acquired using surface electrodes placed on the throat in differential mode. In the implementation phase, electrode placement was based on the proposed rule by SENIAM (Surface Electromyography for Noninvasive Assessment of Muscles) [1].

Provided that EMG signals are non stationary (their frequency varies in time), feature extraction was carried out using time-frequency representations by means of tools such as the Discrete Wavelet Transform and Wavelet Packet Transform [7], obtaining a classification effectiveness of about 75%.

This paper describes the methods and steps employed in the acquisition, analysis, processing and classification of sub-vocal speech signals as well as some relevant results in terms of signal classification.

Methodology

a. Data Acquisition

The EMG signals of sub-vocal speech were acquired using surface electrodes of Silver Chloride (AgCl) in bipolar configuration [2, 3], these were located on the right inferior part and left superior part under the throat. After cleaning the skin with isopropyl alcohol at 70 % to reduce the impedance of the skin (to compensate for the low amplitude of sub-vocal signals), the EMG signals were amplified by a factor of 18700.

In the acquisition system to reduce noise and to attenuate frequencies which are not part of the EMG signal, we used a band-pass Butterworth filter of order 8, formed by a high-pass filter of 30 Hz and a low pass filter of 450 Hz; the information was recorded at a sampling frequency of 50 kHz in vectors of 2 seconds and 50 signals per word. The implemented hardware was specifically designed

for this application by GIBUP (Biomedical Engineering Group at the University of Pamplona). The final implementation is shown in Fig. 1.

Figure 1. *EMG card implemented in the acquisition and conditioning of subvocal signals.*
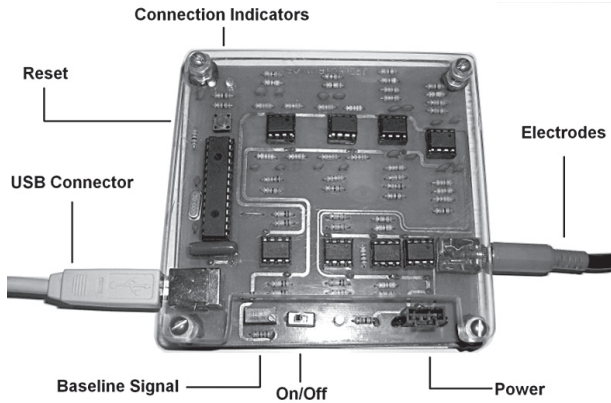
b. Signal Conditioning

Once the signals were recorded and stored, we proceeded to prepare and extract relevant features or patterns. This process was performed in order to obtain important parameters in each group of signals analyzed, and finally accomplish an effective classification process. The conditioning process is described below:

Initially, a process of location of the active area was carried out. This process was performed by means of energy analysis of the signal using windows that were sensitive to sudden changes in EMG-signal amplitude (windows of $400\mu S$ in this particular case). The signal energy for windowing is defined as:

$$E_n = \sum_{i=1}^{W}\left(X_{(n-1)}W + i\right) \quad (1)$$

Where U is the threshold and E_{max} is the maximum peak of the EMG-signal energy vector. Fig. 2 shows an example of energy vector for the intended signal as well as its detection threshold.

$$U = (0.15)E_{max} \quad (2)$$

Figure 2. *Location of the beginning of the active zone by means of thresholding of the EMG-signal energy vector.*
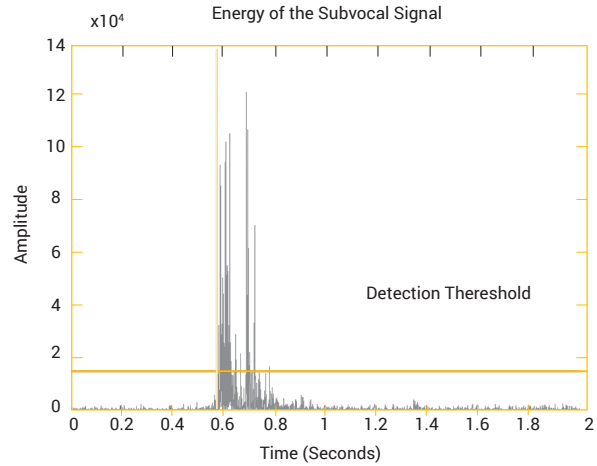
Then, a window of 0.8 seconds is applied from the beginning of the active zone; in this part of the signal is where all the information of the subvocal activity lies. Fig 3 shows (in green) the active area of the sub-vocal speech signal.

Figure 3. *Active zone of the sub-vocal signal for the word "Forward".*
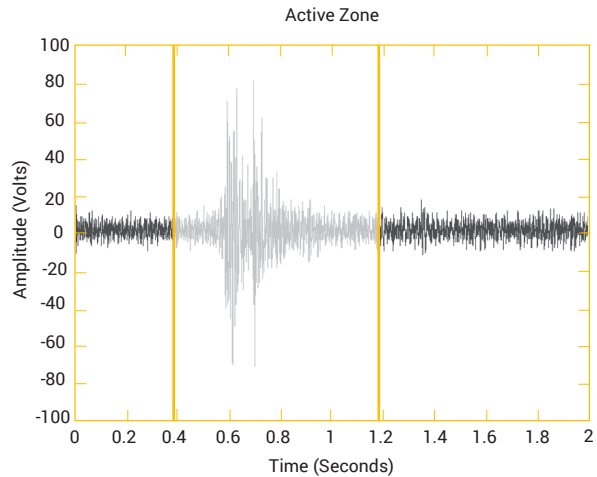
After the location process of the active zone is completed, the signal is filtered using Discrete Wavelet Transform (DWT), which is expressed mathematically, as [8]:

$$\psi_{jk}[n] = 2^{\frac{j}{2}}\psi\left(2^j n - k\right) j,k \in \mathbb{Z} \quad (3)$$

Where j is the scaling parameter, k is the translation parameter and Ψ is the mother wavelet. In this work, the sub-vocal signal was decomposed in four levels using mother wavelet Daubechies 5. The threshold for the filter is defined as:

$$Uf = \sqrt{2Log(n)} \quad (4)$$

Where n is the number of samples of the signal and Uf is the threshold. Once the threshold Uf is found, it is applied to all decomposition coefficients of the signal by using the Hard method. The Hard thresholding method is defined as, [9]:

$$f(x) = \begin{cases} 0, & |x| > U_f \\ x, & |x| > U_f \end{cases} \quad (5)$$

Where x is the decomposition coefficient, Uf is the threshold value and $f\,x$ is the thresholding signal. Then the subvocal signal is reconstructed by means of thresholding coefficients, using the Inverse Wavelet Transform, which is expressed as, [8]:

$$x[n] = \sum_{j \in Z}\sum_{k \in Z} C[j,k] \cdot \psi_{j,k}[n] \quad (6)$$

Where $[n]$ is the reconstructed signal, C are the thresholding coefficients and Ψ is the wavelet basis.

The reconstruction of the sub-vocal signal is shown in Fig. 4. The signal in blue represents the original signal and the signal in red is the filtered signal.

Figure 4.Comparison between the recorded sub-vocal signal and the (DWT) filtered sub-vocal signal.
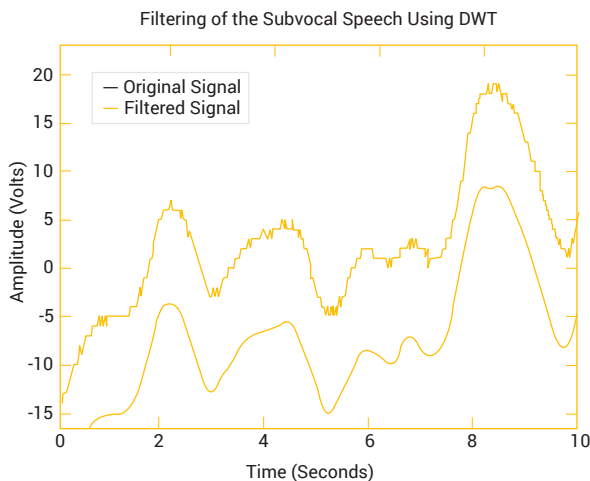

c. Feature Extraction

The relevant feature extraction was achieved by means of DiscreteWavelet Packet Transform (DWPT) ,[10].

DWPT decomposes the approximation coefficients and detail coefficients, forming a tree structure. In this case, the sub-vocal signal was decomposed in three levels, generating 14 coefficients, as shown in Fig. 5.

Figure 5 Selecting the Best DWPT Basis
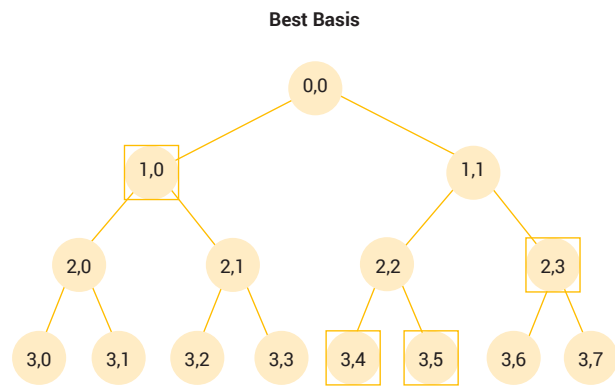
To select the optimal basis or coefficients that best represent the sub-vocal signal, a cost function that measures the level of information is used. In this case, Shannon's entropy was used as the cost function, namely [10]:

$$H(q) = \sum_k |q(k)| Log \frac{1}{|q(k)|} \quad (7)$$

Where H is the Shannon entropy and $q\,(k)$ is the normalized energy of the wavelet coefficients. Once the cost function of each coefficient is obtained, we proceed to choose the optimal basis by means of the parameters suggested in [10]. The result was that the EMG signal of sub-vocal speech is better represented by coefficients (1.0), (2.3), (3.4) and (3.5), shown in Fig. 5.

Taking the results of the DWPT, we proceeded to extract the patterns of the optimal bases using statistical methods such as Root Mean Square (RMS), Absolute Mean Value (AMV), Differential Absolute Mean Value (DAMV) and variance ($\sigma 2$), using windows of 16 ms to reduce the size of the vector characterization of sub-vocal signals. The definition of these methods is illustrated in (8), (9), (10) and (11), respectively:

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k)^2} \qquad (8)$$

$$AMV = \frac{1}{N}\sum_{k=1}^{N}|x_k| \qquad (9)$$

$$DAMV = \frac{1}{N}\sum_{k=1}^{N}|x_{k+1}-x_k| \qquad (10)$$

$$\sigma^2 = \frac{\sum_{k=1}^{N}(x_i-\bar{x})^2}{N-1} \qquad (11)$$

Where N is the window length and x is the optimal coefficient. Once the feature vectors are obtained using statistical methods, we applied Principal Component Analysis (PCA) for dimensionality reduction of the data, [11]. By means of PCA, it was found that 98% of important information is conserved in the first four principal components.

d. Classification

Finally, and with the conditioning process as well as the pattern extraction, we proceeded to perform classification.

Classification was realized through a multilayer-perceptron neural network (RN) with supervised back-propagation learning [12]. The RN has 6 input neurons P, 25 hidden neurons b and one output neuron y Figura 6. Shows the corresponding structure of RN in question.

Figure 6. *RN architecture implemented in the classification of sub-vocal signals*
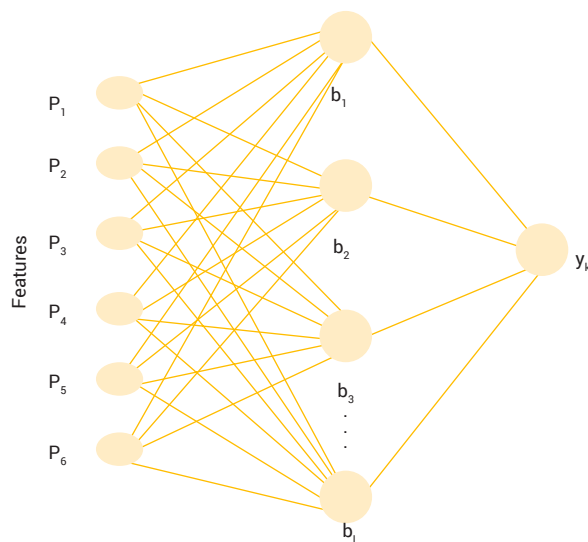
Training of the neural network consisted of 300 cycles and was performed using 70% of the database.

Results

In this section, we show the results of the recognition system in real-time.

a. Classification Effectiveness

The effectiveness test of the classification system for sub-vocal speech was performed using 30% of the database obtained. The database was constructed by capturing signals from a 22-year-old male person in various recording sessions, with conventional noise conditions.

Table 1 shows the results of the classification phase, it can be seen that effectiveness of the algorithm varies depending on the sub-vocal signal being classified, with an average accuracy of 75%.

The effectiveness rate may vary according to various aspects, namely correct placement of the electrodes, sensitivity of system acquisition, signal conditioning, and feature extraction.

Table 1. *Effectiveness of the classification process*

Classification Results						
Word	Right	Left	Forward	Backward	Star	Stop
Efectivity (%)	75.5	78	74.4	77.5	82.8	61.22

b. Results in Real Time

Conditioning, feature extraction, and classification algorithms were implemented in a recognition system (in real time) using MatLab-based software. This system consists of a monitoring interface, which indicates the time of acquisition as well as the connection status of the EMG card and also the classification results. Fig. 7 shows a GUI for the recording of a sub-vocal signal of 2 seconds (time span) and the recognition of the word "Forward."

Figure 7. *GUI for acquisition and recognition interface of sub-vocal speech signals.*
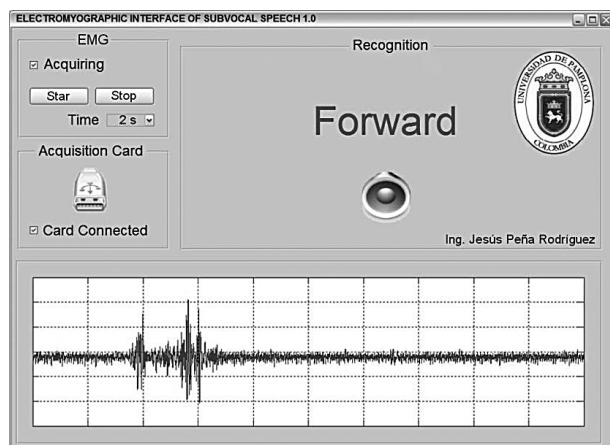
The real time system classifies 3 sub-vocal signals per minute, having an effectiveness of 75% in a test environment, exposed to 70 dB of acoustic noise.

Discussion and conclusions

The results of this study are comparable to studies in [2], [3] and [4], where similar accuracy averages were obtained. The sub-vocal system effectiveness is low, relative to voice recognition systems, but unlike these, the sub-vocal system is not affected by auditory contamination or pronunciation problems.

Other speech recognition systems use brain signals (EEG) and EMG signals, recording multiple signals at once, with increased complexity due to much higher processing and also a lower recognition rate.

The sub-vocal communication system uses only one EMG channel, which reduces the computational cost in terms of processing and also improves accuracy rates.

Future work

This work presents the first set of results of a silent communication system with a recognition vocabulary of six words and recognition effectiveness of 75%. Some of the tasks to be carried out in the development of future studies leading to improvement of the whole sub-vocal communication system are the following:

The registration sensors will be implanted by means of an adjustable structure which can be comfortably attached to the neck. The size of the acquisition and conditioning hardware will be reduced for its implementation in real environments.

The sub-vocal signals will be transmitted wirelessly to the processing and real-time-recognition system. The feature-extraction (and classification) algorithm will be improved to obtain a higher effectiveness rate.

Acknowledgments

References

[1] The SENIAM website.Available: http://www.seniam.org/, 2010.

[2] Chuck Jorgensen, Diana D. Lee and Shane Agabon, "Sub Auditory Speech Recognition Based on EMG Signals,"Proceedingsof the International Joint Conference on Neural Networks(IJCNN), IEEE, vol. 4, 2003, pp. 3128–3133.

[3] Chuck Jorgensen and Kim Binsted, "Web Browser Control Using EMG Based Subvocal Speech Recognition,"Proceedings ofthe 38th Annual Hawaii International Conference onSystem Sciences (HICSS), IEEE, 2005, pp. 294c.1–294c.8.

[4] Bradley J. Betts and Charles Jorgensen, "Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment,"Neuro-Engineering Laboratory, NASA Ames Research Center, Moffett Field, California, EEUU, 2005.

[5] Szu Chen S.Jou, "Automatic Speech Recognition on Vibrocervigraphic and Electromyographic Signals", Language Technologies Institute, Carnegie Mellon University, Pittsburgh PA 15213, EEUU, Oct. 2008.

[6] Curtis D. Hardyck and Lewis F. Petrinovich, "Treatment of Subvocal Speech During Reading", en Journal of Reading, pp. 361-368, Febrero 1969.

[7] K. Englehart, B. Hudgins, P.A. Parker, and M. Stevenson, "Classification of the Myoelectric Signal using Time-Frequency Based Representations," Special Issue Medical Engineering and Physics on Intelligent Data Analysis in Electromyography and Electroneurography, Summer 1999.

[8] Marcos C. Goñi y Alejandro P. de la Hoz, "Análisis de Señales Biomédicas Mediante Transformada Wavelet", Concurso de Trabajos Estudiantiles EST, Universidad Nacional de San Martín, Argentina, 2005.

[9] Dora María Ballesteros Larrotta, "Aplicación de la Transformada Wavelet Discreta en el Filtrado de Señales Bioeléctricas,"Umbral Científico, Fundación Universitaria Manuela Beltrán, Bogotá, Colombia, pp. 92-98, Dic. 2004.

[10] F. A. Sepúlveda, "Extracción de Parámetros de Señales de Voz usando Técnicas de Análisis en Tiempo-Frecuencia," Universidad Nacional de Colombia, Manizales, Colombia, 2004.

[11] Lindsay I. Smith, "A Tutorial on Principal Components Analysis," Feb. 2002.

[12] Bonifacio M. del Brío y Alfreda S. Molina, Redes Neuronales y Sistemas Borrosos, 2nd ed., Alfao-mega S.A. de C.V., Ed., México D.F., 2006.

The Authors

MSc. Ing. Luis Enrique Mendoza

Research Group of GIBUP University The Pamplona. Associate Professor Program of Telecommunication engineering, Faculty of Engineering.

Ing. Jesus Peña R.

Research Group of GIBUP University The Pamplona. Electronic Engineering of University of Pamplona, Colombia.

Ph.D, Ing. Jairo Lenin Ramón Valencia

Research Group of BIOAXIS University The Bosque. Associate Professor Program of Bioengineering, Editor of the Journal of Technology, Faculty of Engineering, PhD Industrial Technology. University Polytechnic of Cartagena, Spain. Biomedical Engineer Manuela Beltran University, Bucaramanga, Engineer Industrial Tec Specialty in Electronics, UPCT, Spain.