International Journal of Bio-Science and Bio-Technology
Vol.7, No.5 (2015), pp.43-54
http://dx.doi.org/10.14257/ijbsbt.2015.7.5.05

Comparative Analysis of Signals Acquired Before and After the Place of Articulation

Germán Darío Buitrago Salazar[1], Darío Amaya Hurtado[2], Olga Lucía Ramos Sandoval[3] and Jorge Enrique Saby Beltrán[4]

[1, 2, 3]Research Group GAV, Faculty of Mechatronic Engineering, Nueva Granada Military University, Bogotá, Colombia
[4]Research Group of Magnetic and Resonance, Francisco José de Caldas Distrital University, Bogotá, Colombia
gedabusa@gmail.com[1], dario.amaya@unimilitar.edu.co[2], olga.ramos@unimilitar.edu.co[3], jesabyb@udistrital.edu.co[4]

Abstract

Communication is a key factor in a developing society. It is usually carried out by using signals that come from the brain, and then produced as sounds in the Place of Articulation. An alternative to this type of communication is subvocal speech or "subvocalization", which consists in the acquisition of the signals produced in the cerebral cortex and their analysis before the voice production process in the Place of Articulation is made. This paper proposes a prototype system for acquisition of the signals produced in the Place of Articulation; by using a Data Acquisition System (DAS) and MATLAB® as a tool for comparing the obtained results.

Keywords: *Subvocal Speech, Signal acquisition, Embedded Systems, Place of Articulation*

1. Introduction

Dialogue is important on our daily lives, since it allows interaction between the individuals of a society: expressing their ideas, giving instructions, transmitting knowledge, and communicating. One way of communication among human beings is through speech.

The voice production associated to speech, is performed by the Place of Articulation where sounds that can be interpreted by other people are produced [19]. However, problems related to communication can be presented in people who have lost partially or totally the ability to talk, not being able to communicate using speech due to losses of certain properties of the Place of Articulation. Also, noisy environments affect the transmission of sound when the speech task is performed, this represents a constraint during communication where the speech message cannot be transmitted appropriately, and therefore the information can be deformed or misinterpreted.

In these environments an alternative of communication is required, for instance, a recent area of interest named "subvocalization" or silent speech gives a solution to these problems by taking a different approach. A definition of subvocalization is given in [21], where the author defines it as the process in which a person thinks in phrases or words, but there are no signal orders from the brain that would allow facial movements or excitation of the vocal folds.

Many techniques have been used in order to capture and identify subvocalization signals. In [1] a paper where "Electromagnetic Articulography" (EMA) is used for monitoring the co-articulation movements in the mouth is presented. When a person pronounces a word, a sensor detects the variation of the magnetic field, which allows a

ISSN: 2233-7849 IJBSBT

recreation of the trajectory of the tongue and mouth. So, by using several algorithms, recognition of words can be achieved. One of the main problems of using this technique, is the sensibility, which allows interference of alternating magnetic fields of other equipment.

Another method of performing subvocal speech recognition is presented in [2], where speech recognition is carried out by using "surface Electromyography" (sEMG), which allows data collection of the electric signals of the muscles involved in the process of voice production. In the methodology presented in [2], 11 sensors were placed around the face and neck of a subject, in order to distinguish three different ways of speaking: vocalizing, mouthing and the process of thinking the word during the speech task. The resulting recognition accuracy for vocalized speech, was of 92.1% and 86.7% for mouthed speech, these results show that the method is suitable for recognition tasks.

The paper presented in [9], authors performed tests of a subvocal recognition system, using sEMG as the signal acquisition method, complemented with "Hidden Markov Models" (HMM) to achieve the recognition process. Based on them, a training method model that takes into account the maximum likelihood criteria was proposed. The results of this study, using 60 Mandarin words, show an average recognition of 87.07%, which indicates high recognition reliability, when signals are taken from the muscles of the face.

NASA has also implemented EMG technology to control the interface of a web browser [3]. Detection of the words was performed by collecting signals from larynx, through an EMG acquisition process. 17 vowels, 23 consonants and 10 digits (in English) were identified. An artificial neural network was used as a classification strategy. This application, is important in the field of aerospace; because it allows speech task under circumstances where it would not be possible to pronounce words, for example, in underwater operations, in space or in noisy environments where speech messages could be distorted. NASA also wants to broaden the uses of this technology for processes that need a higher level of privacy, by sending messages that cannot be heard by nearby people; or to avoid the need of building an extra human-machine interface.

In [10] other results of the work performed by NASA are presented; they obtained recognition of 15 English words, under high level of noise in the communication. The main purpose of this application, was to move a robotic platform by recognizing basic commands from the user. Results from this work show recognition reliability between 71% and 77% with a 95% confidence interval.

One of the last techniques that has been developed, is recognition of subvocal speech of Non Audible Murmur (NAM). As it is shown in [4], acquisition of the signals can be done by using a microphone placed over the bell of a stethoscope, acquiring the articulated sounds produced in respiration. Those signals are produced by the interaction and motion of the vocal tract organs and transmitted through soft tissue of the head [5]. The acquisition and pattern recognition technique was proposed by Yoshikata Nakajima in 2003.

In 2010, a paper about implementing a technique for recognition of words and phonemes of the Japanese language, was presented in [12]. The paper shows that due to the small spectral distance in the signals of speech detected by NAM, if HMM is used to perform the recognition of words, the spectral distance between words compared to vocalized speech would also be lower; this means that detection and identification of words would be poor. In order to improve this method of subvocal recognition, the author added recognition of the facial movements of the speakers. This allowed the efficiency of the detection process to be improved; initially it was of 81.5% and after including facial movement recognition was improved in 10%.

There are other techniques for subvocal recognition, whose methodologies include several and different steps as: implementing the tracking of tongue and lips using image processing, complemented with a synchronization of the vocal tract variation measurement using Ultra Sound (US) [13]; using a physiological sensor (PMIC), which

captures the vibrations produced by the speech, due to skin contact at the cricoid and thyroid cartilages at the larynx [14]; the use of an electroencephalogram (EEG), using 16 different channels with the purpose of recognizing words by simply thinking them [15]; and interpretation of the brain signals, using an interface between the brain and a computer (BCI), where the information of the process of speech, is predicted due to the activity of the neurons involved in the speech task [16]. The results obtained by each of these techniques are presented in [17], where the advantages and disadvantages of each method are described.

This paper performs recognition of speech by using a NAM microphone. The paper is divided in three parts: in the first one, the methodology and necessary tools to perform the procedure are specified; then, a comparison and analysis of the signals acquired before and after the place of articulation is accomplished; and last, the conclusions and possible future investigations are presented.

2. Tools and Methodology

Many techniques can be used in order to detect the signals needed to carry out subvocal speech. NAM technique was chosen to acquire the input signals from a Data Acquisition System (DAS); where, from the articulated respiratory sounds that are produced, the signals can be detected in the soft tissue of the head [4]. The architecture of the project is essentially divided in three parts, as shown in Figure 1.

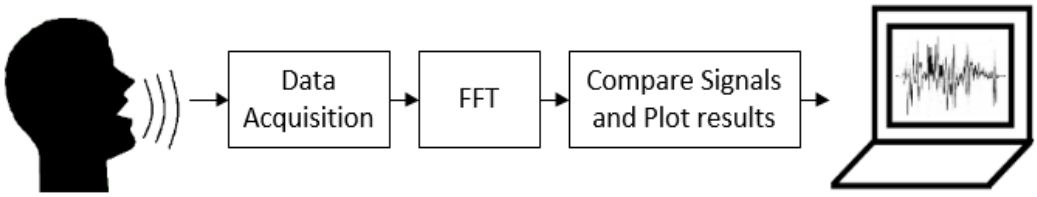

Figure 1. Project Stages

2.1. Data Acquisition

In order to obtain the NAM signals from the procedure described in [5], filling the cavity of the bell and membrane of a stethoscope with a polymer substance is proposed. In this case, silicon rubber was used, which allows improving the acquisition process of the signals. The purpose of this, is to achieve an insulation to enhance the acquisition of the signal, avoiding noise from the environment and from the human body.

An "Electret" Microphone was used as the sensor to acquire the input signals. With the microphone and the modification made to the bell of the stethoscope, certain characteristics of detection are improved, such as the sensibility to detect signals in the frequency bandwidth of subvocal speech [4, 5], thus signals detection error is minimized. Figure 2 shows the arrangement of the developed device for detection of the signals.

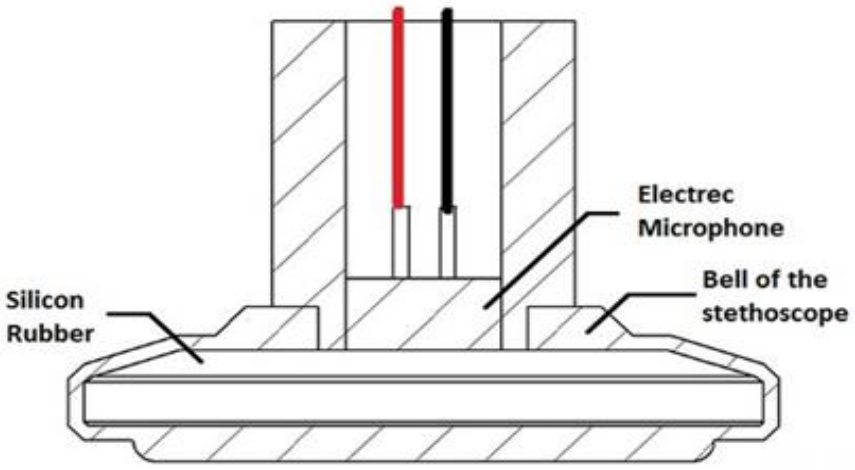

Figure 2. Cross Section and Parts to Make the NAM Microphone Signals

The NAM microphone was placed in the back-bottom part of the ear, as shown in Figure 3. The output of the microphone was connected to the DAS along with stages of amplification and coupling of impedances to make the signal conditioning, to finally send the signals to the computer.

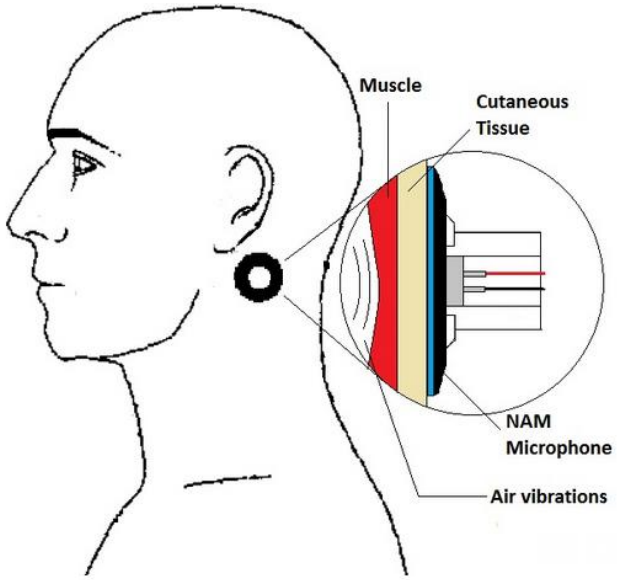

Figure 3. Location of Sensor Signals to Make the NAM

For the digitalization process of the subvocal signals, an embedded system with a built-in input filter was used. This allows an initial filtering of the signal of noise from the environment. In order to choose the sampling frequency, Nyquist-Shannon theorem was implemented. Nyquist-Shannon theorem states that the minimum sample frequency for digitalization of a signal, has to be the double of the maximum frequency of the signal [6].

The human voice frequency ranges are from 50 Hz to 3800Hz [7], so the chosen sample frequency was of 8000 Hz. For the filter design, it was taken advantage of Butterworth's filter characteristics of ripple elimination in the passband, and constant gain at the cutoff frequency [8]. The equations shown in [18], were used for the design of the embedded filters, with a cutoff frequency of 3800Hz, for a bandwidth of -3 dB.

Digital filters have a transfer function as the one shown in equation (1), which relates the numerator coefficients $B(z)$ and the denominator coefficients $A(z)$. From this

equation, the magnitude of the frequency response is determined, as shown in equation (2).

$$H(Z)=\frac{B_1+B_2z^{-1}+\ldots+B_{n+1}z^{-n}}{1+A_2z^{-1}+\ldots+A_{n+1}z^{-n}} \quad (1)$$

$$\left|H\left(e^{j\omega}\right)\right|^2=\frac{P(x)}{Q(x)}=\frac{P\left(\frac{1}{2}-\frac{1}{2}\cos\omega\right)}{Q\left(\frac{1}{2}-\frac{1}{2}\cos\omega\right)} \quad (2)$$

From the classic theory of Butterworth filter design and by using equations (1) and (2), in [20] is proposed a general equation as the one shown in equation (3), where L is the number of zeros when z=1, M is the number of zeros that affect the passband, N is the total number of poles, ω_0 is the frequency with maximum magnitude, x_0 is the simplification of the expression: $1/2\,(1-\cos\omega_0)$, τ_N is the notation for the polynomial truncation, R(x) and T(x) are two auxiliary polynomials contemplated in [20] and c is a free parameter to set with accuracy the transition band. The value of c is given by equation (4) and it depends on the type of filter to be used.

$$F(x)=\frac{(1-x)^L\left(R(x)+cT(x)\right)}{\tau_N\left\{(1-x)^L\left(R(x)+cT(x)\right)\right\}} \quad (3)$$

$$c=\frac{4(1-x_0)^LR(x_0)-\tau_N\{(1-x)^LR(x)\}(x_0)}{\tau_N\{(1-x)^LT(x)\}\left(x_0-4(1-x_0)^LT(x_0)\right)} \quad (4)$$

Once the filter was designed, the DAS, samples the signal that is within the voltage specifications, given by the default resolution of the convertors of the board; then the signal passes through the previously designed filter; and finally, the signal digitalization data are sent to the computer, through a RS232 communication protocol, using de serial module of the board. The scheme of the process is described in Figure 4.

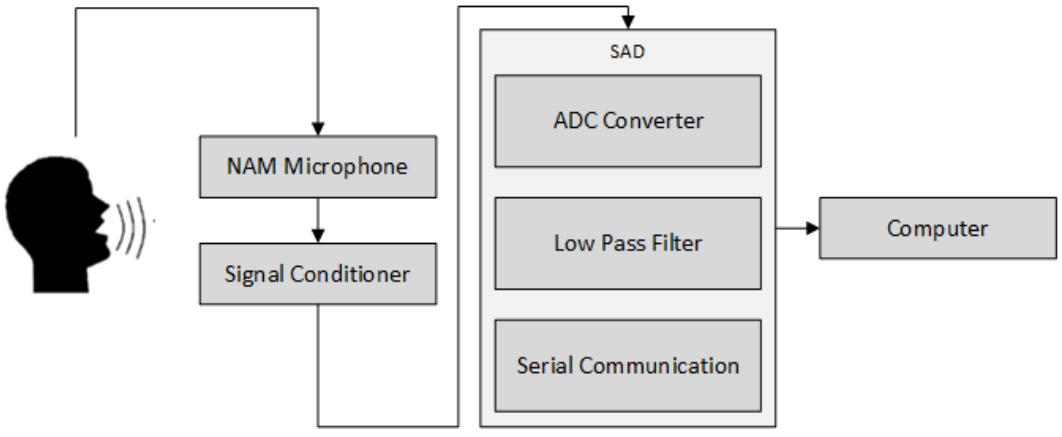

Figure 4. Data Acquisition Stage for Subvocal Speech

2.2. Obtaining Fast Fourier Transform (FFT) of Signal

In this stage of the process, the sampled and filtered signal are captured by a MATLAB interface, as the one shown in Figure 5. The interface controls all the timing and stages of the process, such as the command to start detecting the speech during a given time,

programmed by the user. The interface, also controls the reception and transmission syncing timings of the data towards the computer.

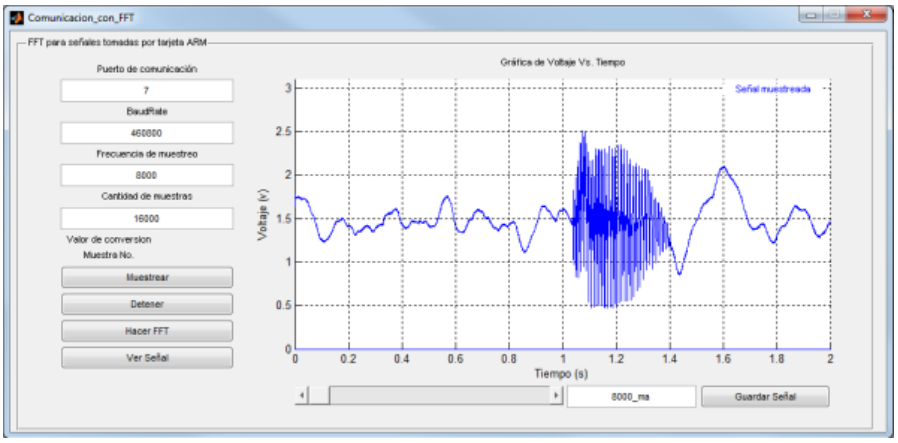

Figure 5. Interface for Signal Acquisition

Having the signal digitalized and saved in the computer, the signal data are put through another band-pass filter, to eliminate frequencies lower than 70 Hz and higher than 3700 Hz. The filter also allows eliminating the offset level of the signal and obtaining the necessary information needed for the range of study. A Butterworth filter was chosen, given that using window-based filters, the results were not as optimal for the application compared to a Butterworth filter.

This second filter was designed using equations (3) and (4), and varying L, M and N parameters. After designing the filter in discrete time and associating it by a difference equation, the polynomial form of the filter, was programed in a second interface, as shown in Figure 6.

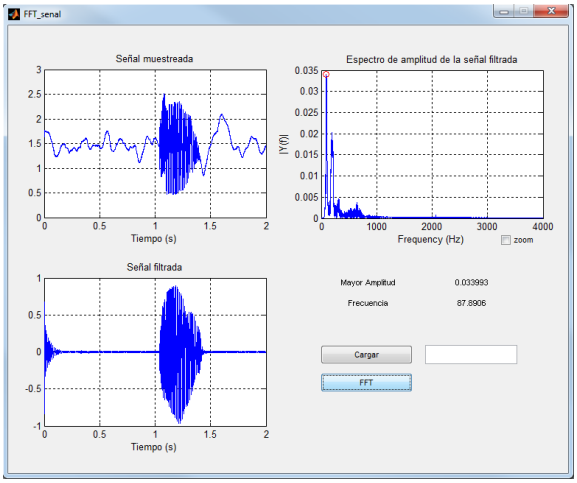

Figure 6. FFT Interface for the Signal

The subvocal speech signal, passes through the filter and both signals are plotted (filtered and unfiltered signal) as shown in Figure 7.

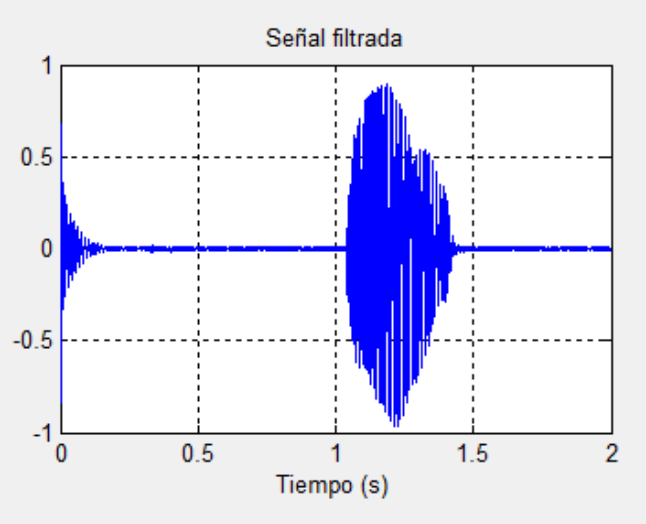

Figure 7. Subvocal Speech Signal after the Second Filter

In order to differ and compare the sounds, the Discrete Fourier Transform (DFT) was used, so that the signal is passed to the frequency domain. The DFT is defined by equations (5) and (6), where X(k) corresponds to the sampling of the signal in the time domain, and $x(n)$ corresponds to the samples of the signal in the frequency domain. W_N is the unit of the n^{th} root or twiddle factor, and N is the number of samples of the signal.

$$X(k)=\sum_{n=0}^{N-1} x(n)W_N^{nk} \quad , k=0,1,\ldots,N-1 \tag{5}$$

$$W_N=e^{\frac{-2\pi}{N}} \tag{6}$$

From the model presented in [11] an eight-point FFT with butterfly module is proposed, as it is shown in figure 8. This model was chosen since it is a simple way to perform the DFT and the execution time is shorter.

2.3. Comparison of Signals

On the final stage of this work, a comparison between the signals of subvocal speech and words after the place of articulation was performed. This was done by pronouncing the sounds */ma/*, */me/*, */mi/*, */mo/* and */mu/*.

An Electrotec microphone placed at the height of the mouth, was used to acquire the signals after the place of articulation. The same procedure as the one described for subvocal speech signals, was used to acquire and filter the signals. The FFT was performed to both the voice signal and subvocalization signal, obtaining the maximum power level of the signals, as shown in Figure 9. It can be noticed that the maximum value of the subvocal speech signal is displaced in frequency compared to the voice signal, so there is a difference for the same sound comparing the two methods.

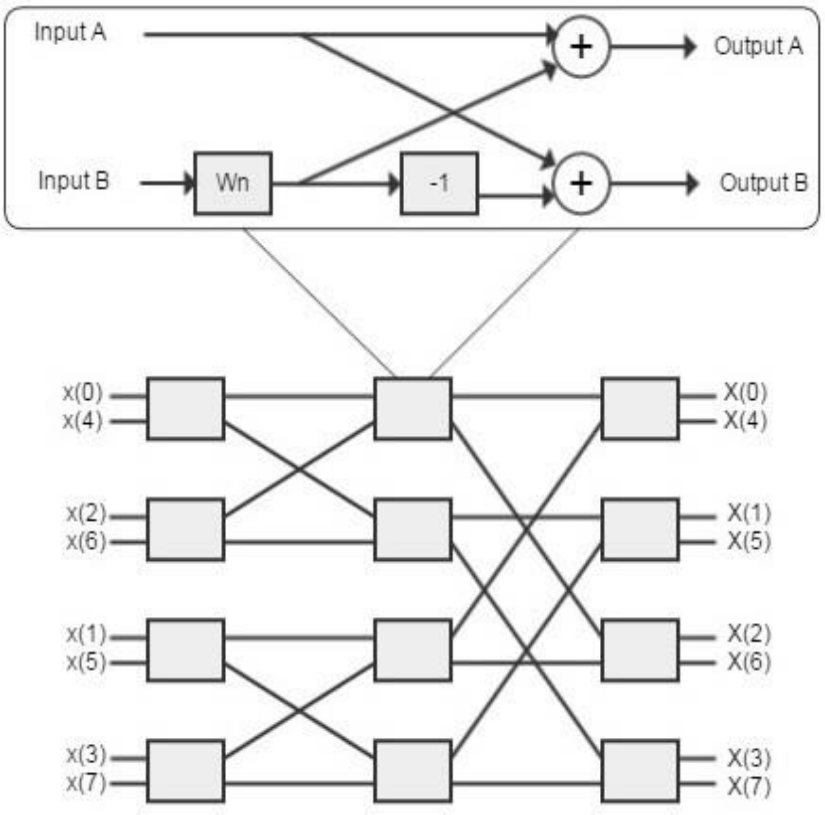

Figure 8. 8-point FFT Butterfly Network

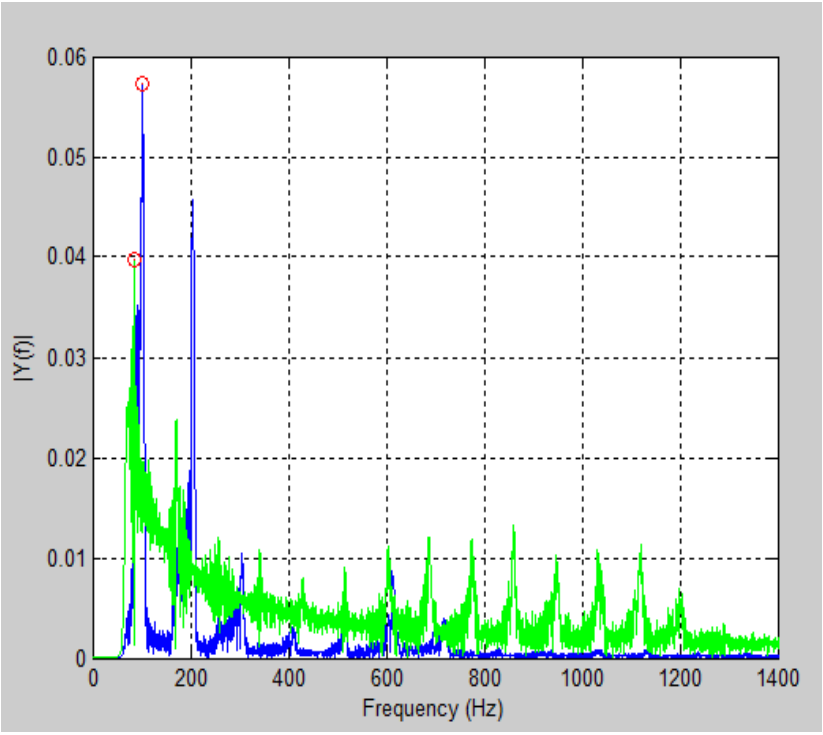

Figure 9. Frequency Spectrum of the Speech Signal Subvocal (blue) and Voice (green)

Table 1, shows the value of the harmonics with the greatest power in the pronunciation of each of the described sounds. Table 2, shows the values for each of the pronounced sounds and sensed after the place of articulation.

Table 1. Frequencies with Maximum Power in Subvocal Speech

High Frequency Values for Subvocal (Hz)					
Sample	**Word**				
	Ma	**Me**	**Mi**	**Mo**	**Mu**
1	87,89	89,36	130,86	89,36	104,98
2	203,61	104,00	214,84	95,21	101,07
3	94,24	92,29	214,84	188,96	99,61
4	101,56	97,17	212,40	182,62	102,05
5	102,05	99,61	209,96	94,73	102,54
6	106,93	95,21	106,93	93,26	106,93
7	100,59	101,56	197,27	91,80	203,13
8	98,63	106,93	97,17	92,29	102,05
9	192,87	102,05	103,52	93,75	103,03
10	92,29	108,40	102,54	98,63	96,19

Table 2. Frequencies with Maximum Powers after the Place of Articulation

High Frequency Values after the place of articulation (Hz)					
Sample	**Word**				
	Ma	**Me**	**Mi**	**Mo**	**Mu**
1	83,98	96,68	98,63	76,66	90,82
2	91,80	98,63	214,84	92,29	91,80
3	84,96	97,66	202,64	98,14	94,73
4	90,33	92,29	103,03	99,12	94,73
5	94,24	73,73	93,26	94,73	210,94
6	91,80	89,84	196,78	99,12	110,35

3. Analysis of Results

After obtaining the samples and organizing them, it can be noticed that for the sound */ma/*, the frequency value bandwidth, varies from 95 Hz to 105 Hz. There is a great difference between this ranges, compared to the pronounced voice signals data, being that the frequency values for the second range tend to 89.5 Hz.

For the sound */me/*, the frequency values have a mean of 99.48 Hz, while the pronounced sound is at 91.47 Hz. For the sound */mi/*, other two harmonics appear close to 200 Hz, being these the values of higher power. The frequency of two harmonics are shifted, when these are compared with the frequency of the spoken sound. This phenomenon is caused by the power of the signal.

For the sounds */mo/* and */mu/*, the frequency of the subvocal signal ranges from 89 Hz to 94 Hz and 96 to 106 Hz, respectively. Leaving aside the harmonics greater than 200 for

the spoken signal, and taking into account the harmonics around 100Hz, the value for the spoken signals for the */mo/* and */mu/* sounds range from 92 Hz to 99 Hz, and 90 Hz to 110 Hz, respectively.

4. Conclusions and Future Works

From the obtained results with each of the sounds, acquired by the NAM microphone and by regular vocalization; it can be observed that the values of subvocal speech are in a specific frequency range compared to the voice signals, being the latter signals frequencies displaced. This phenomenon might occur, because of the effect of the amplification of the signals using the bell of the stethoscope, or due to the location of where the NAM signals were taken from.

The technique for subvocal recognition using NAM, has several advantages. For example, cost reduction, the use of a detecting and acquisition system that is easy to build and use, and that it can be used in spaces where noise reduction is required. This silent speech recognition interface could be used in the future for people with speech problems involving voice production, as well as becoming a communication system between people.

If the FFT is done with a greater number of points, the results would be better than the samples taken in this paper, but it would increase the number of operations and samples in order to carry it out. This would require a larger memory, and a greater number of resources and would as well lower the speed of the process.

The interfaces previously mentioned and the mechanism to detect subvocal speech by NAM, could be implemented to detect words in Spanish, based on subvocal recognition of sounds. There currently exist other methods for detection of speech (Hybrid Markov Models and voice correlation), which combined with NAM, could be used to make a more reliable, robust and efficient system to detect subvocalization signals.

References

[1] P. Hoole and N. Nguyen, "Electromagnetic articulography in coarticulation research", Instituts für Phonetik und Sprachliche Kommunikation det Universitat München, vol. 35, **(1999)**, pp. 177-184.
[2] G. Meltzner, J. Sroka, J. T. Heaton, L. D. Gilmore, G. Colby, S. Roy and N. Chen, "Speech Recognition for Vocalized and Subvocal Modes of Production Using Surface EMG Signals from the Neck and Face", 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, **(2008)** September 22-26, pp. 2667-2670.
[3] C. Jorgensen and K. Binsend. "Web Browser Control Using EMG Based Sub-vocal Speech Recognition", Proceedings of the 38th Annual Hawaii International Conference on System Sciences HICSS '05, United States, **(2005)** January 3-6, pp. 1-7.
[4] Y. Nakajima, H. Kashioka, K. Shikano and N. Campbell, "Non-audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03, Hong-Kong, vol. 5, **(2003)** April 6-10, pp. 708-711.
[5] T. Toda, K. Nakamura, T. Nagai, T. Kaino, Y. Nakajima and K. Shikano, "Technologies for Processing Body-Conducted Speech Detected with Non-Audible Murmur Microphone", Interspeech 2009, Brighton, United Kingdom, **(2009)** September 6-10, pp. 632-635.
[6] B. Gold, N. Morgan and D. Ellis, "Digital Signal processing", Speech and audio Signal Processing, 2nd Ed., Wiley Publisher, **(2011)**, pp. 73-86.
[7] B. Gold, N. Morgan and D. Ellis, "Speech analysis and synthesis overview", Speech and audio signal processing, 2nd Ed., Wiley Publisher, **(2011)**, pp. 21-39.
[8] S. K. Jagtap and M. D. Uplane, "The Impact of Digital Filtering to ECG Analyisis: Butterworth Filter Application", IEEE International Conference Communication, Information & Computing Technology, Mumbai, India, **(2012)** October 19-20, pp. 1-6.
[9] K. S. Lee, "EMG-Based Speech Recognition Using Hidden Markov Model with Global Control Variables", IEEE Transactions on Biomedical Engineering, vol. 55, no. 3, **(2008)** March, pp. 930-940.
[10] B. J. Betts, K. Binsted and C. Jorgensen, "Small Vocabulary Speech Recognition Using Surface Electromyography", Interacting with computers, vol. 18, **(2006)** December, pp. 1242-1259.
[11] S. K. Lu and C. H. Yeh, "Easily Testable and Fault-Tolerant Design of FFT Butterfly Networks", Proceedings of the 11th Asian Test Symposium, Guam, USA, **(2002)**, pp. 230-235.

[12] P. Heracleous, V. Tran, T. Nagai and K. Shikano, "Analysis and Recognition of NAM Speech Using HMM Distances and Visual Information", IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 6, **(2010)**, pp. 1528-1538.
[13] B. Denby and M. Stone, "Speech Synthesis from Real Time Ultrasound Images of the Tongue", IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada, vol. 1, **(2004)**, pp. 1685-1688.
[14] S. A. Patil and J. Hansen, "The physiological microphone (PMIC): A competitive alternative for speaker assessment in stress detection and speaker verification", Speech Communication, vol. 52, **(2010)**, pp. 327-340.
[15] A. Porbadnigk, M. Wester, J. Callies and T. Schultz, "EEG-based speech recognition – impact of temporal effects", 2nd International Conference on Bio-inspired Systems and Signal Processing (Biosignals), Porto, Portugal, **(2009)**, pp. 376-381.
[16] J. Brumberg, A. Nieto-Castanon, P. R. Kennedy and F. H. Guenther, "Brain-computer interfaces for speech communication", Speech Communication, vol. 52, **(2010)**, pp. 367-379.
[17] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert and J. S. Brumberg, "Silent speech interfaces", Speech Communication, vol. 52, **(2010)**, pp. 270-287.
[18] J. Proakis and D. Manolakis, "Diseño de filtros digitales", Tratamiento digital de señales, 4th Ed., Pearson-Prentice Hall Editorial, **(2007)**, pp. 584-688.
[19] D. Berlo, "El proceso de la comunicación: una introducción a la teoría y a la práctica", El Ateneo Editorial, Argentina, **(1999)**.
[20] I. Selesnick and C. S. Burrus, "Generalized Digital Butterworth Filter Design", IEEE Transactions on Signal Processing, vol. 46, no. 6, **(1998)**, pp. 1688-1694.
[21] NASA TechBrief. Who´s who at NASA: Chuck Jorgensen. Available in: http://www.nasatech.com/NEWS/ May04/who_0504.html, **(2004)**.

Authors

Germán Buitrago Salazar was born in Bogotá, Colombia. He studied at UMNG (Nueva Granada Military University), Bogota, Colombia receiving the B.Sc. degree in Mechatronics Engineering in 2014. Right now is working as a Research Assistant at UMNG in the Research Group GAV in different Mechatronics fields like robotic, image processing and artificial intelligence.

Olga Lucia Ramos Sandoval is originally from Bogotá, Colombia. She was educated at UAN, Bogotá, Colombia receiving the B.Sc. degree in Electronics Engineering in 1998. She got her specialization certified in Electronic Instrumentation in 2000 by UAN and the M.Sc. degree in Teleinformatic in 2007 by the Faculty of Engineering at the Francisco José de Caldas District University, UFJC in Bogotá, Colombia. Currently she is completing the Ph.D. degree in Engineering at UFCJ. Right now she is working as a Teacher at UMNG and as Research in the Research Group GAV in different Mechatronics fields like System Control and Industrial Automation.

Dario Amaya Hurtado was born in Quimbaya, Quindio, Colombia. He was educated at UAN, Bogotá, Colombia receiving the B Sc. degree in Electronics Engineering in 1995 and the M.Sc. degree in Teleinformatic in 2007 by the Faculty of Engineering at the Francisco José de Caldas District University, UFJC in Bogotá, Colombia. He was awarded the Ph.D. degree in 2011 in Mechanical Engineering at Campinas State University,

São Paulo, Brazil, working on hybrid control – He has worked as a professor and researcher at the Military University, Colombia since 2007 been involved in Robotics, Mechatronics and Automation areas.

Jorge Enrique Saby Beltrán was born in Bogotá, Colombia. He was educated at Francisco José de Caldas Distrital University, Bogotá, Colombia receiving the B.A. degree in Linguistics and Literature in 1986 and the M.Sc. degree in Spanish linguistics in 1993 by National University in Bogotá, Colombia. He was awarded the Ph.D. degrees in 1998 in Linguistics and Literature at Pontifical Catholic University of Rio Grande Do Sul, Porto Alegre, Brazil and in 1997 in Science Education at Pontifical Catholic University of Córdoba, Córdoba, Argentina. He has worked as a professor and researcher at the Francisco José de Caldas Distrital University, Colombia since 1989.