Research Journal of Applied Sciences 14 (1): 1-6, 2019
ISSN: 1815-932X
© Medwell Journals, 2019

NAM Patterns Identification with Artificial Neuronal Networks

Clavijo Juan, Olga Ramos and Dario Amaya
Universidad Militar Nueva Granada, Bogota, Colombia

Abstract: Sub vocal speech patterns are a unique investigation line, it has strengthened in the las few years. Aiming to identify sub vocal patterns with NAM sensors, this investigation proposes to identify sub vocal samples with Artificial Intelligence (AI) like neuronal networks techniques. Between the results of the research, it can be noticed the capacity to identify phrases and Spanish words.

Key words: Sub vocal speech, silent speech, NAM, FFT, RNA, neuronal networks

INTRODUCTION

The sub vocal speech evidences its importance by the technological advances through times like electronic communications high growth and development. In the verbal communication scope, the electronical equipment is essential and daily used. Man-man communication and man-machine are common in general. Keeping in mind that speech interaction systems are speaker phone and there is a technological gap in silent speech. And this aims to make allow understandable man-machine communications.

Non audible bicker (Nakajima *et al*., 2003) is one way of silent speech that could allow to identify incomplete speech characteristics that could be useful for its subsequent synthetic generation or for an input device replacing or complementing its functions as they do in machines keyboards.

The pattern identification is managed by a series of samples, taken in the fourier transform spectrum (Proakis and Dimitris, 2007). This samples are normalized and transformed in temporal patterns, thanks to a group of successive samples.

The patterns are treated as an image that represents a phrase or a word spectrum (Gonzalez, 2008). As the pattern can be interpreted as an image, it also can be treated and manipulated as one. Image digital treatment allows to make normalizations over the pattern, like its length, contrast contour detection, etc., (Proakis and Dimitris, 2007; Gonzalez, 2008).

With the information discreet in similar characteristic patterns, the next step is to identify its information. Aiming to do this, this research proposes a radial basis neuronal network, identifying 5 groups, this type of network was selected because of its fast training and easy programming. It's necessary to stand out that this neuronal networks strategy has been used before for similar investigations (Leonard and Kramer, 1991; Kim and Park, 2004; Chen, 1994).

MATERIALS AND METHODS

Characteristics of patterns: The patterns are acquired by means of a NAM microphone, developed under implemented characteristics of some other investigations (Nakajima *et al*., 2003; Tran *et al*., 2010; Heracleous *et al*., 2010). The NAM microphone acquired acoustic data is amplified and transformed in frequency domain, to create FFT samples. This process is accomplished by a basic DSP dsPIC30F4013 in which the Fourier transform is done (Heracleous *et al*., 2010) and then is transmitted by WiFi connection to a computer where the strong sample processing is done. The acquired patterns are treated as images that correspond to a word or phrase spectrogram, each of this images can be manipulated by creating normalizations between the parameters in terms of contrast and time. The spectrogram samples are formed by FFT samples taken at a 15, 625 Hz frequency, over 128 samples in the time domain. Each FFT samples, generates 64 samples that corresponds to the contained bands and frequencies in a 0-1 kHz spectrum, due to the sample frequency over the NAM signal is 2 kHz.

A pattern over the spectrogram is shown in Fig. 1. The patterns are normalized in champs of 64 by 64 pixels, generating 4096 dimension patterns. The less strong data that correspond to the most sharped harmonics is replaced by a new characteristic of the data of the pattern in Fig. 1, it can be seen as a shell pattern and mathematically corresponds to the weighted average of the values of all the bands in each sample FFT. The same parameter is noticed in Fig. 2, as a sloped band that represents its own average.

This development can also normalize the harmonic frequencies in function of the first harmonic, this condition enables to make transparent the speaker phone's levels of fundamental frequencies that can vary due to the emotional state.

Corresponding Author: Clavijo Juan, Universidad Militar Nueva Granada, Bogota, Colombia

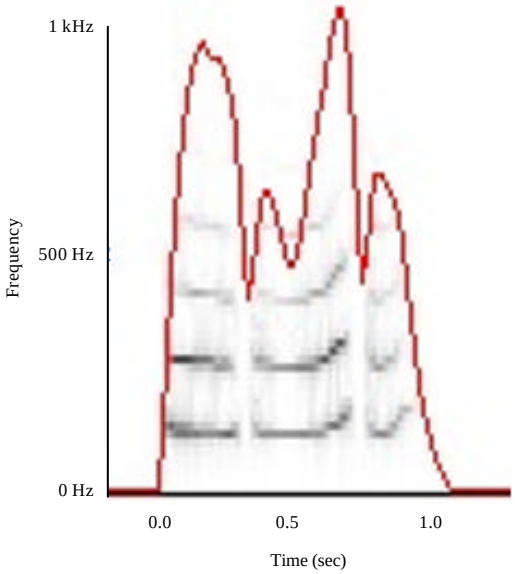


Fig. 1: Spectrogram

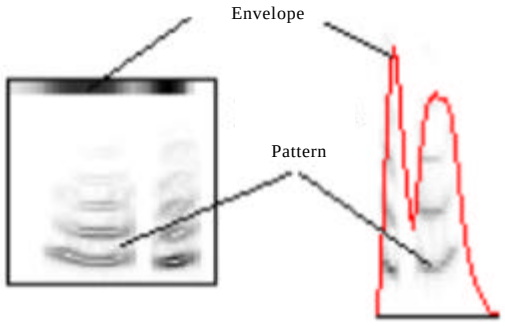


Fig. 2: Normalized pattern

The effect of normalization regarding the fundamental frequency is noticed in Fig. 3a, b represents the non-normalized spectrum and the normalized spectrum. With appropriately characterized patterns some identification process can be applied.

Pattern identification: To identify patterns this investigation implements, artificial intelligent techniques based in artificial neuronal networks (Leonard and Kramer, 1991; Kim and Park, 2004; Chen, 1994). In general terms there are many useful techniques for this purpose, like unidirectional, radial, self-organized, fed back networks, etc.

In first instance this research implements radial basis neuronal networks, easing the training process due to its fast iteration capability (Yousefian *et al*., 2008; Sankar and Sethi, 1997; Diaz-de-Maria and Figueiras-Vidal, 1995; Tao *et al*., 2010).

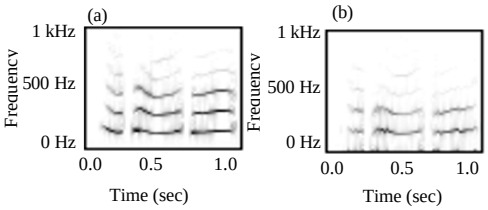


Fig. 3: a, b) Harmonic comparative

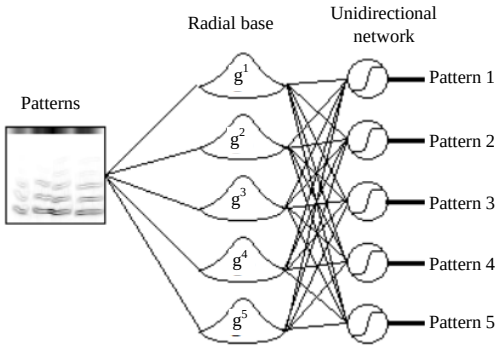


Fig. 4: Neuronal architecture

An artificial neuronal network is a connection arrangement with synaptic weights that allows the simplified simulation of biological neurons behavior (Georg, 2005).

The radial basis network is characterized by the implementation of a sensitive layer in first place in second place a radial basis layer in which each neuron is a Gaussian identifier, specialized in vector space portions which may belong to a particular pattern (Simon, 2009). This proves that for each existent neuron present in the layer, a Gaussian centroid would exists and it may represent a group of patterns.

This Gaussian layer is followed by one or many layers with stagger activation or sigmoid. This last layer allows to depurate the Gaussian classification (Georg, 2005; Simon, 2009).

For this particular case the third and fourth unidirectional with sigmoid activation layers were implemented, they had as much outputs as patterns to identify.

Figure 4 shows the implemented architecture of the accomplished tests. The implementations of 5 radial basis networks has as purpose, make each of the Gaussian neurons an specialist in one training patterns, this means that for the process of training it is required to have at least 5 different patterns, Fig. 4 shows a synthesized architecture, as the number of dimensions of each pattern is 4096 and that's the input number for each neuron whereas that the next neuronal layer only has 5 inputs per neuron.

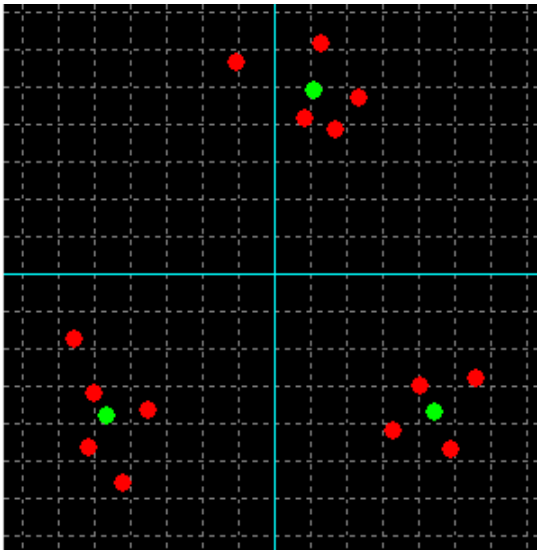


Fig. 5: Fuzzy clustering

Whenever the number of training patterns is even to the Gaussian neurons, the arrangement values in each centroid change into the same, simplifying the training of this layer, however, this way of training is useful just a couple of times and the networks is usually trained with so, much more patterns, so, the network could generalize the input patterns in a better way.

When the number of patterns exceed the number of Gaussian neurons, a neuronal grouping algorithm is required to find the center of the pattern assembly, to show this situation, Fig. 5 shows the classification result of 14 elements or 2-dimensional patterns grouped into 3 classes. The red points represents the elements and the green points represents the group's centers.

The location process of the centers is achieved by means of a fuzzy clustering, it is the same basic neuronal training technique used for the programmed radial basis neuronal network (Georg, 2005). The last parameter required to train the Gaussian neurons is the definition of its range, this is the distance between its center and its nearest neighbor center. The set of centroid and radius creates the action of each Gaussian neuron, creating a Gaussian bell that wrap the groups.

The second stage of training of the artificial neuronal network is done by a supervised method known as Back Propagation (Simon, 2009), this one consist in propagating the error variations in an iterative way over time but backwards, this way it creates error values in the hidden layers and with this proportions a Hebbian training procedure can be made in each of them. After providing 10 patterns to the network, it is ready to identify the provided patterns.

Fig. 6: Voluntary test

Fig. 7: Software view

Tests: The tests of this investigations are consider with the following pattern characteristics:

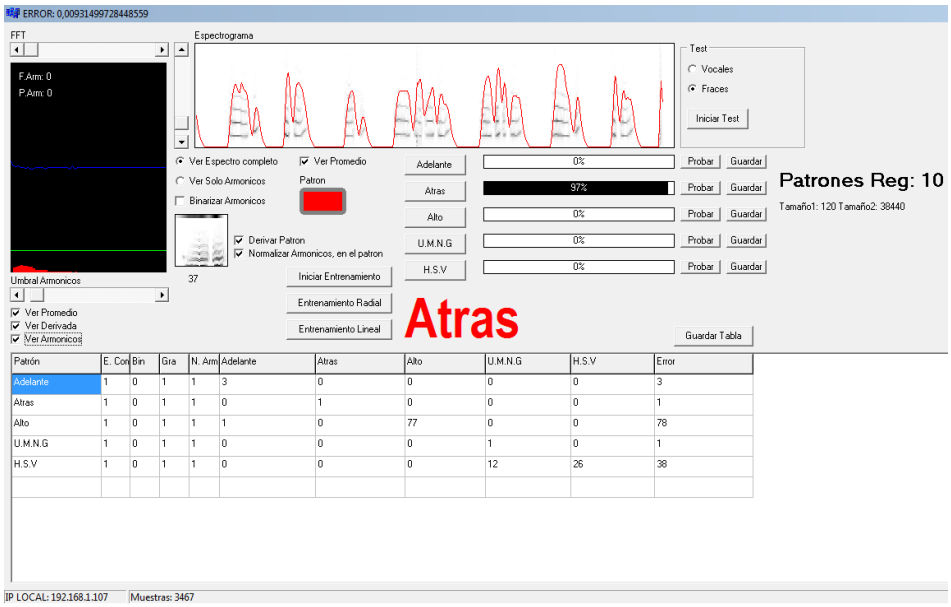
- Harmonic frequencies normalizations
- Gradient calculation for each pattern
- Hybrid training with 10 patterns
- Tests with vowels and sentences

The test evaluates 5 adults at the same conditions. Each guy uses a NAM microphone, located in the sternocleidomastoid, behind the pinna. It is kept in position with elastic band, this is shown in Fig. 6. The partakers should make the aleatory pronunciation of 5 vowels and 5 phrases or words. The used phrases were:

- Ahead
- Stop
- Behind
- Sub-vocal speaks
- Military University

To standardize the test, the partaker has to read the phrase or the vowel from a computer screen. The participants also see randomly the words or phrases that must be murmured, doing it with the mouth shut, only allowing poor pronunciation and pressure generated out only through the nostrils.

Figure 7 shows the screen that shows the participants the word or phrase that should be pronounced. This view



Patrón	E. Con	Bin	Gra	N. Arm	Adelante	Atras	Alto	U.M.N.G	H.S.V	Error
Adelante	1	0	1	1	3	0	0	0	0	3
Atras	1	0	1	1	0	1	0	0	0	1
Alto	1	0	1	1	1	0	77	0	0	78
U.M.N.G	1	0	1	1	0	0	0	1	0	1
H.S.V	1	0	1	1	0	0	0	12	26	38

Fig. 8: Management software view

is accomplished thanks to an application made in C# language, it has connectivity by a COM serial port. The text and time view is commanded trough this port. Other computer send the commands it also runs the acquisition algorithm for the pattern identification.

The patterns are captured by the mentioned neural network in a C# language application where the whole process is managed. The investigator in charge can choose the normalization patterns, the patterns set of words or phrases, he can also decide how many patterns may train the neuronal network and save the results in a file format*. csv, this file can be opened in Excel to export the data and create a stadistic analysis of the resultant data.

A general view of the management application is shown in Fig. 8. At the end of the test 11 samples of each vowels were collected making a total of 55 test over the vowels. With the 5 mentioned words and phrases, 16 samples were collected for a total of 80 samples over the phrases. At the end 135 samples were analyzed without counting the training samples and for each partaker the test was about 20 between vowels and phrases in brief 235 samples were required, to finish the test proposed for the investigation.

After applying the neuronal network and use a new test pattern, the system register the percentage of similarity against the training patterns, Fig. 9 shows this information in a bar graphic. After determining the obtained data and knowing percentages in each of the presented patterns, a percentage value for the error is created as show in the Table 1.

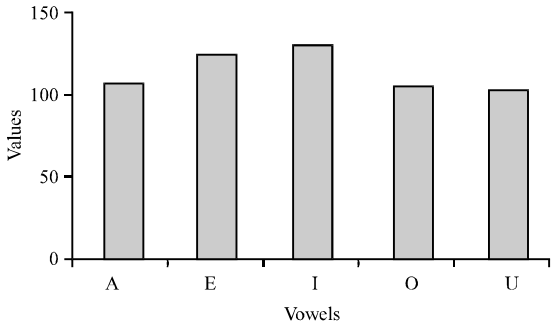


Fig. 9: Error in vowel identification

Table 1: Example of error calculation

Ahead	Behind	Stop	M.U.N.G	S.V.S	Error
80-100	10-0	2-0	0-0	1-0	33

The error is expressed as the summary of the absolute error value of each pattern. In Table 1, the presented pattern was Ahead and the neuronal network identified it as ahead in 80% as behind in 10% as stop in 2% and in 1% as sub-vocal speak. The absolute error value is the 33%, however is clear that the neural network classifies the input as ahead, this example shows an error relatively big but it can be classified as a successful identification.

RESULTS AND DISCUSSION

After testing the entire population, the following results were determined over the corresponding vocal

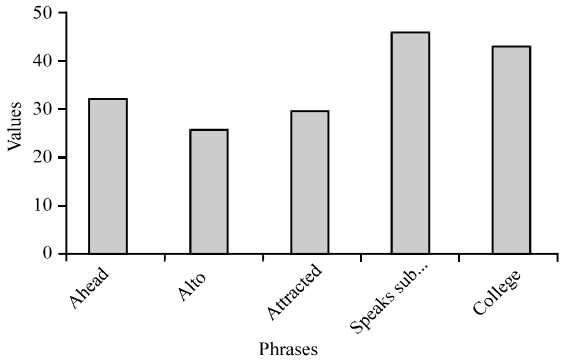


Fig. 10: Error in phrase identification

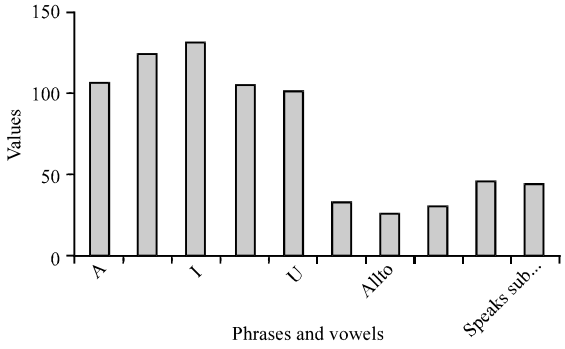


Fig. 11: Comparative error

patterns. This is shown in Fig. 9. The error result is obtained the same way and it can be observed in Fig. 10.

Figure 11 illustrates in a general and comparative way, the result between vowels and phrases. Figure 11 can be noticed that the associated pattern to the vowels has accumulative errors above 100% meanwhile the accumulative phrases error is close to 40% values.

From the results of the experimentation process it can be said that this system can identify complex phonetic forms as phrases or words but there's no success with vowels or simple phonemes. That condition has been shown, at the graphic 3.

This situation is analyzed by the authors of this work as a trouble, due to the similitude of the vowels, this observation can be seen in Fig. 12 which illustrates the 5 vowels spectrogram and where is hard to find notable differences.

Figure 10 shows the similarity between the vowels patterns A, E, O, U, isolating slightly I pattern due to its higher error rate comparing to graphics 1 and 3. The others vowels are hard to differentiate.

On the other hand Fig. 13 illustrates the spectrum of the phrases and words test patterns where the differences are clear enough to make a classification.

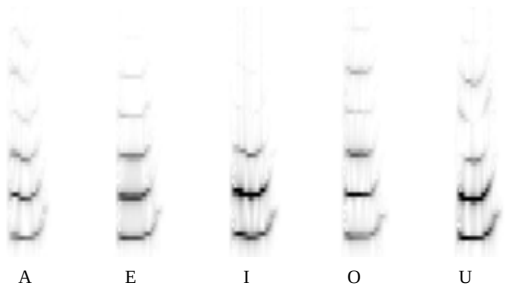


Fig. 12: Vowel spectrogram

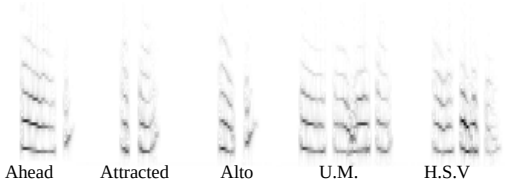


Fig. 13: Words and phrases spectrum

The comparison of the results between the results of words and vowels evidences the possibility to use sub vocal speech as a way of man-machine interaction, putting away such simple interpretations as vowels or consonants.

CONCLUSION

Another important factor is the sample frequency of the audio acquisition system and also the FFT frequency, due to the speed limitations of the device DSP is not possible a higher quality acquisition, however, the usage of devices with higher capacity should be considered to improve and evolve this type of applications.

IMPLEMENTATIONS

The implementation of this type of systems can be applied to control the performance of electronic devices and also the words used in this investigation "Ahead, stop, behind" can be used as simple instructions to control mobile robots.

REFERENCES

Chen, D.W., 1994. Using localized basis function for multi-speaker speech recognition. Proceedings of the 1994 International Symposium on Speech, Image Processing and Neural Networks, April 13-16, 1994, IEEE, Hong Kong, ISBN: 0-7803-1865-X, pp: 734-736.

Diaz-de-Maria, F. and A.R. Figueiras-Vidal, 1995. Nonlinear prediction for speech coding using radial basis functions. Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP-95) Vol. 1, May 9-12, 1995, IEEE, Detroit, Michigan, USA., ISBN:0-7803-2431-5, pp: 788-791.

Georg, F.L., 2005. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6th Edn., Addison-Wesley, Boston, Massachusetts, USA., ISBN:978-0-321-54589-3, Pages: 784.

Gonzalez, R., 2008. Digital Image Processing. 2nd Edn., Prentice Hall, Upper Saddle River, New Jersey, USA., ISBN:978-81-7758-168-3, Pages: 793.

Heracleous, P., V.A. Tran, T. Nagai and K. Shikano, 2010. Analysis and recognition of NAM speech using HMM distances and visual information. IEEE. Trans. Audio Speech Lang. Process., 18: 1528-1538.

Kim, H.I. and S.K. Park, 2004. Voice activity detection algorithm using radial basis function network. Electron. Lett., 40: 1454-1455.

Leonard, J.A. and M.A. Kramer, 1991. Radial basis function networks for classifying process faults. Cont. Syst. Mag., 11: 31-38.

Nakajima, Y., H. Kashioka, K. Shikano and N. Campbell, 2003. Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin. Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal (ICASSP'03) Vol. 5, April 6-10, 2003, IEEE, Hong Kong, China, ISBN:0-7803-7663-3, V708-V711.

Proakis, J.G. and G.M. Dimitris, 2007. Digital Signal Processing. 4th Edn., Prentice Hall, Upper Saddle River, New Jersey, USA., ISBN:9780132287319, Pages: 948.

Sankar, R. and N.S. Sethi, 1997. Robust speech recognition techniques using a radial basis function neural network for mobile applications. Proceedings of the International Conference on Engineering New New Century Southeastcon, April 12-14, 1997, IEEE, Blacksburg, Virginia, USA., ISBN:0-7803-3844-8, pp: 87-91.

Simon, H., 2009. Neural Networks and Learning Machines. 3rd Edn., China Machine Press, Beijing.

Tao, Z., X.D. Tan, T. Han, J.H. Gu and Y.S. Xu *et al*., 2010. Reconstruction of normal speech from whispered speech based on RBF neural network. Proceedings of the 2010 3rd International Symposium on Intelligent Information Technology and Security Informatics (IITSI), April 2-4, 2010, IEEE, Jinggangshan, China, ISBN:978-1-4244-6743-3, pp: 374-377.

Tran, V.A., G. Bailly, H. Lœvenbruck and T. Toda, 2010. Improvement to a NAM-captured whisper-to-speech system. Speech Commun., 52: 314-326.

Yousefian, N., A. Jalalvand, P. Ahmadi and M. Analoui, 2008. Speech recognition with a competitive probabilistic radial basis neural network. Proceedings of the 4th IEEE International Conference on Intelligent Systems Vol. 1, September 6-8, 2008, IEEE, Varna, Bulgaria, ISBN:978-1-4244-1739-1, pp: 7-19.