The SOM Robustness Capacity for Phonemes Recognition in Adverse Environment

Mohamed Salah Salhi
National school for Engineers
LP 37, the Belvédère, 1002
Tunis,Tunisia

Najet Arous
National school for Engineers
LP 37, the Belvédère, 1002
Tunis,Tunisia

Noureddine Ellouze
National school for Engineers
LP 37, the Belvédère, 1002
Tunis,Tunisia

ABSTRACT

In this paper, we describe the formatting guidelines for IJCA Journal Submission. The SOM, for kohonen Self Organizing Map, has proven to be a classifier of high caliber in the field of speech recognition signals breasts. Thus, several versions and enhancements were applied on this tool such as recurrent SOM 'RSOM', the growing recurrent SOM 'GRSOM' and the growing hierarchical SOM GHSOM, to consider the integration of sound variability. This paper aims to detect the ability of the SOM robustness in terms of phonemes recognition for continuous speech in a noisy environment. This idea represents, in fact, the real case for speech recognition.

General Terms

This paper is generally classified in the profile of Pattern Recognition and signal processing.

Keywords

Noisy environment; phonemes recognition; Self organizing map SOM; SOM robustness.

1. INTRODUCTION

The field of phonemes recognition is a very important area. It participated in the simplification of voice recognition, speaker identification, man-machine speech communication… Several classifiers were used for phonemes recognition since 1982 such as: HMM, SVM and neural networks. The majority of these methods are based on a supervised learning algorithm. In the case of processing a large data dictionary, an unsupervised learning algorithm is most appropriate. In that, the SOM model is proved as a powerful tool for phonemes recognition. In speech recognition, this model has been known since the 80's by the Finn T. Kohonen.

N. Arous used the SOM in speech recognition in the 2000s by the hybridization of GA and SOM. For this, as Kohonen, she uses three central windows of the phoneme.

C. Jlassi in (2006-2010) presents the hierarchical SOM 'HSOM' and the growing hierarchical SOM 'GHSOM', she also has used three central windows of the phoneme. Our contribution is:

-The implementation of the SOM on all windows of the phoneme.

-Improving the robustness of the SOM in a noisy environment.

After this introduction, our paper will deal with the SOM principle and the essentially comparative study for some methods in the second section. In section three we will focus on the examination of SOM robustness in the noise presence.

However, the experimental results and their argumentations are integrated respectively below each section. Finally, the section four is dedicated to the conclusion.

2. SOM OPERATION

Moving to the SOM principle in phonemic recognition, this model is proposed by T. Kohonen in 82. It was inspired by the biology of the human brain. It may be two dimensional or string. It uses the following rules:

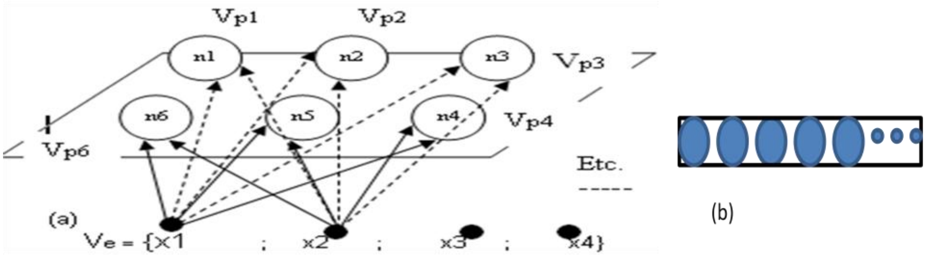

Fig.1 Representation of the SOM map; (a) in dimension 2 and (b) in string

The SOM map is based on the following Rules [1]:
-Each component of the input vector will be sent to each neuron of the SOM map.

-At each learning iteration end, the weight vectors, called prototypes and initialized linearly or randomly in the start, will have each one the same components of the input vector.

-The quantization error associated to a neuron (i) is calculated by the Euclidean distance as follows:

$$E_i = \|x(t) - w_i\| \quad (1)$$

The winner neuron (v) also called BMU, is the unit that minimizes the quantization error from which we have:

$$E_v = \min E_i \quad ; i \in N \quad (2)$$

The learning rule updates the neural weights located in the vicinity of the active neuron by bringing them close to the input vector:

$$\Delta w_i = \gamma . h_{iv} (x(t) - w_i) \quad (3)$$

γ is a learning report and hiv is a neighborhood function, which decreases by the distance between units (i) and (v) on the SOM map.

However, the SOM disadvantage results in its convergence to a local optimum and the algorithm is not initially adapted to process the temporal dimensions of speech signal.

3. SOM EXPERIMENTATION FOR HOLY SPEECH RECOGNITION

Our work is experienced on the TIMIT database by considering its wide dissemination in the international community makes an objective assessment and sharing of desired performance of developed recognition systems.

Applied to phoneme recognition, our program performs the segmentation of sentences into 10ms sound atoms.

We have used the total energy of each phoneme for the data parameterization.

The adopted approach for phonemic recognition is abstracted as follow:

-Reading and segmentation of the TIMIT database sentences.

-Convert sound atoms in acoustic vectors consisting of coefficients MFCCs.

-Sequential presentation of these vectors into the SOM map.

-The MFCCs are calculated on all the phonemes windows where the energy is considered maximum.

The table and figure below show the results of the phonemes recognition in their classes by SOM.

Table I. Phonemic recognition rate of TIMIT database by SOM

Phonemes classes	**Training**		**Test**	
	Nbr of simples	**Rates**	**Nbr of simples**	**Rates**
Vowels	4432	74.95	1339	52.89
Semi-vowels	2717	86.57	844	70.99
Nasales	1109	94.32	332	64.45
Fricatives	1729	82.95	528	59.43
Stops	3946	75.95	1122	48.86
Others	239	75.95	75	52.75
Affriquates	172	90.03	42	82.29
Overall rate	**14344**	**82.96**	**4282**	**61.66**

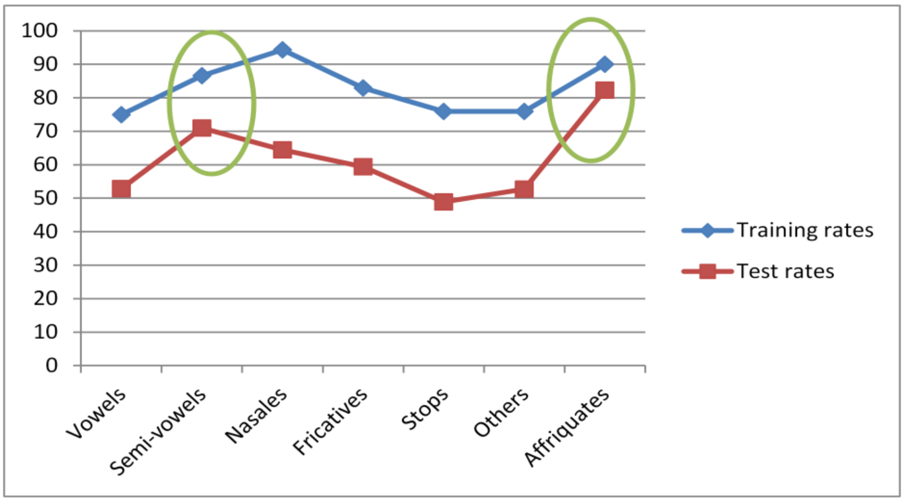

Fig.2 Representation of Training and test rates for all classes of TIMIT database

We note that for certain classes such as semi-vowels and affricates, the generalization (test) rates are closed to self consistency (training) ones due to their transient state.

A comparative study on the vowels basis shows the performance of SOM taking account of all the phoneme windows compared to SOM on three central windows. The obtained results confirm that they are better than those obtained by other models using three central windows.

Table II. Comparative recognition rates over models on vowels of TIMIT database by SOM

Recognition model	Phoneme Support	Global generalization rates
SOM N.Arous (2003)	3 central windows	**56%** 81% for training
GHSOM C.Jlassi (2010)	3 central windows	**55%** 75% for training
SOM M.S.Salhi (2011)	All windows	**62%** 83% for training

We have noted that In light of using all the phoneme windows, we have made a gain in generalization rate of 6 % to 7 %.

These results are interpreted as follows:

-The SOM is a powerful tool in the representation of static data.

-The effectiveness in phoneme recognition, is strongly linked to the learning quality (number of windows, extended, overlap, etc..).

The study of the SOM robustness in the noise presence is another contribution that will be the subject of the next section.

4. The SOM MODEL WITH NOISE REDUCTION IN PHONEMIC RECOGNITION

Several noise reduction techniques have been known to use on the speech signal. Some techniques are classified in the spectral attenuation. Others belong to the atomic decomposition. The others are supported to statistical and hearing modeling [2-17].

In our strategy of speech recognition, we have adopted an atomic decomposition technique, which is the technique of wavelet packet denoising.

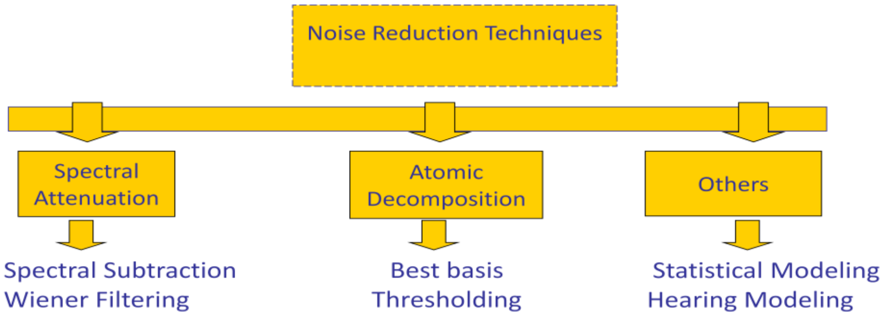

Fig.3 Representation of noise reduction techniques

5. DENOISING BY WAVELET PACKET ANALYSIS AND ADAPTED THRESHOLDING

The method of wavelet packet decomposition that we used is described by the following explanatory diagram.

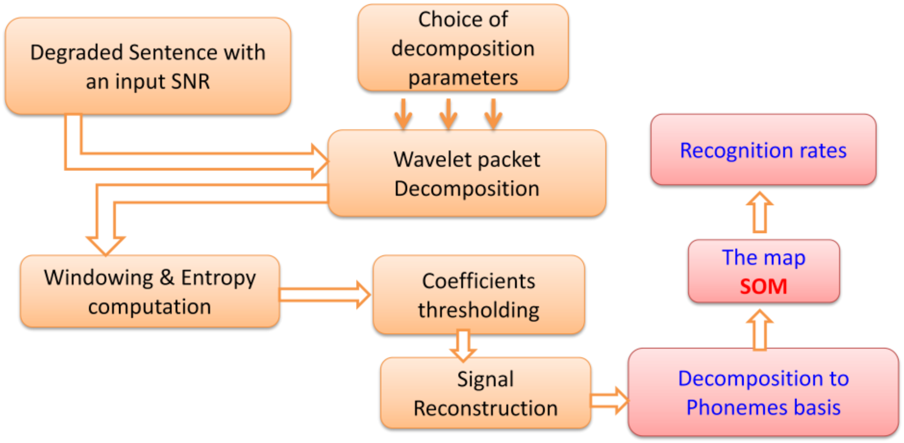
Degraded Sentence with an input SNR
Choice of decomposition parameters
Wavelet packet Decomposition
Windowing & Entropy computation
Coefficients thresholding
Signal Reconstruction
Recognition rates
The map SOM
Decomposition to Phonemes basis

Fig.4 Diagram of adopted recognition algorithm

Contrary to the dyadic wavelet transform (DWT), wavelet packets decompose the signal by low-pass filters and high pass filters [18].

The following figures illustrate well this idea.

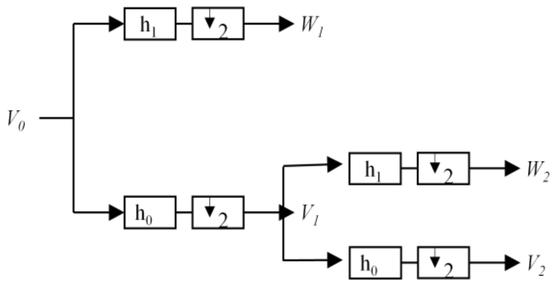

Fig.5 Dyadic wavelet decomposition

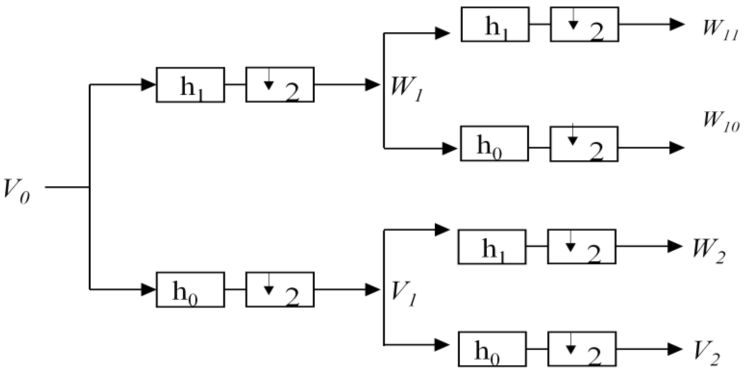

Fig.6 Wavelet packet decomposition

6. DENOISING WITH WAVELET COEFFICIENTS THRESHOLDING

The algorithm we have followed for denoising with wavelet coefficients thresholding is presented below.

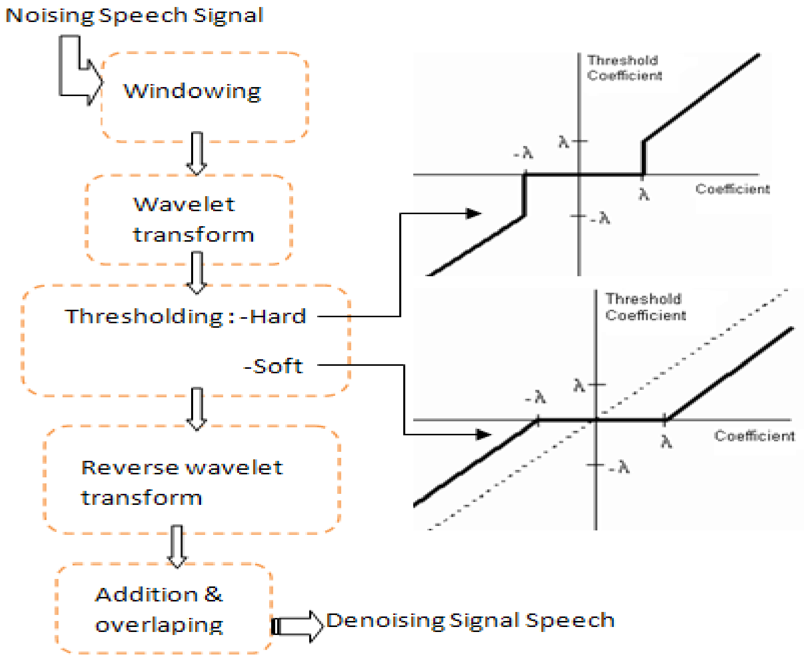

Fig.7 Wavelet analysis algorithm

7. THRESHOLD DETERMINATION

The threshold determination is performed using the following nonlinear function of contraction [19]:

$$THR_h(X,T)=\begin{cases} X & , \quad \text{pour}|X|>T \\ T\left(\dfrac{1}{\mu}\left[(1+\mu)^{|X/T|}-1\right]\bullet sng(X)\right) & , \quad pour\,|X|\prec T \end{cases} \qquad (4)$$

For each node N, the threshold is defined by the following formula:

$$T_{j,k}=\hat{\sigma}_{j,k}\sqrt{2\log(N)} \qquad (5)$$

The parameters j and k represent respectively the scale and the index of the band [20].

We take: $\beta=0{,}9$ et $\mu=255$

- Daubechies wavelets of order 4.
- 3 decomposition levels.

8. DENOISING ON TIMIT VOWELS BASIS

The gain in dB for different denoising techniques is given in the following table.

Table III. Comparative SNR over techniques of noise reduction

Input SNR in dB	Output SNR in dB			
	Spectral subtraction	Hard threshold	Soft threshold	Adapted threshold
10dB	13.20	10.10	10.30	11.60
5dB	9.81	12.40	12.77	5.38
0dB	6.74	10.65	11.41	1.72

We note that the output gain increases with decreasing of input SNR ratio.

9. RECOGNITION RESULTS BY SOM WITH NOISE REDUCTION

Recognition results on the vowels basis with noise reduction are illustrated by the following set of curves.

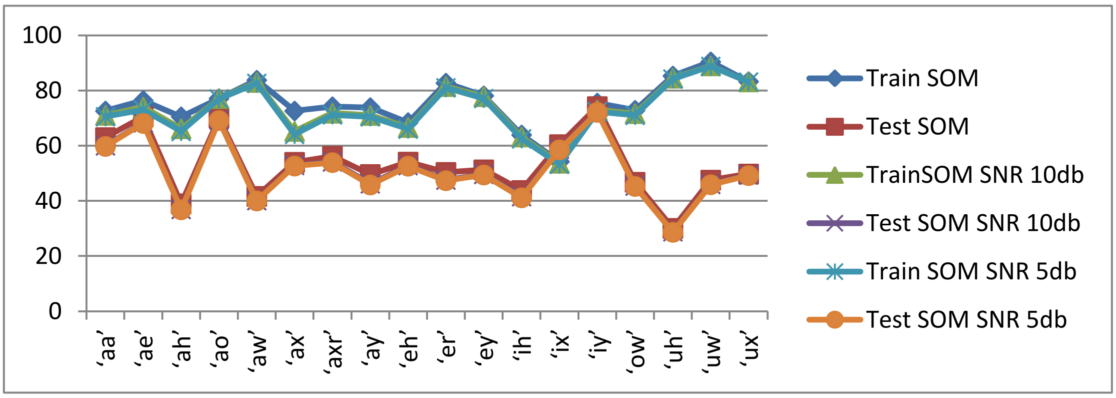

Fig.8 Vowels recognition rate of TIMIT database by the use of SOM joint to wavelet packet technique

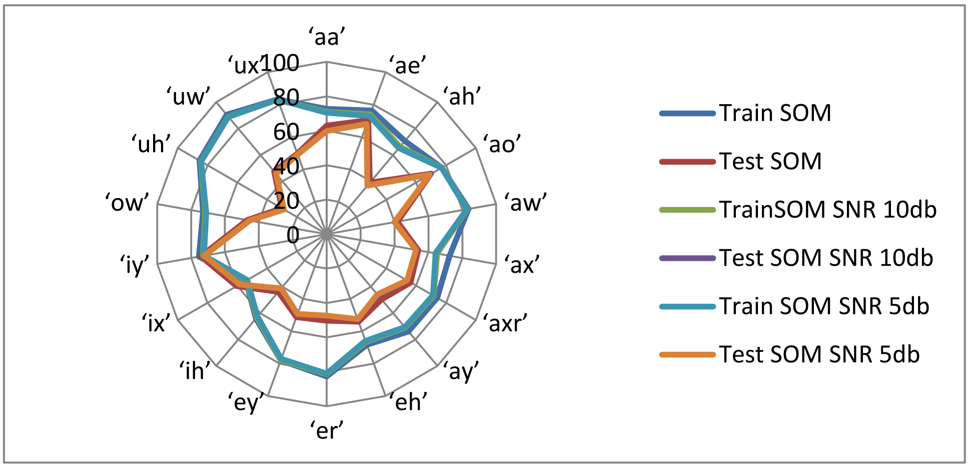

Fig.9 Comparative Vowels recognition rate of TIMIT database by the use of SOM joint to wavelet packet technique in radar form

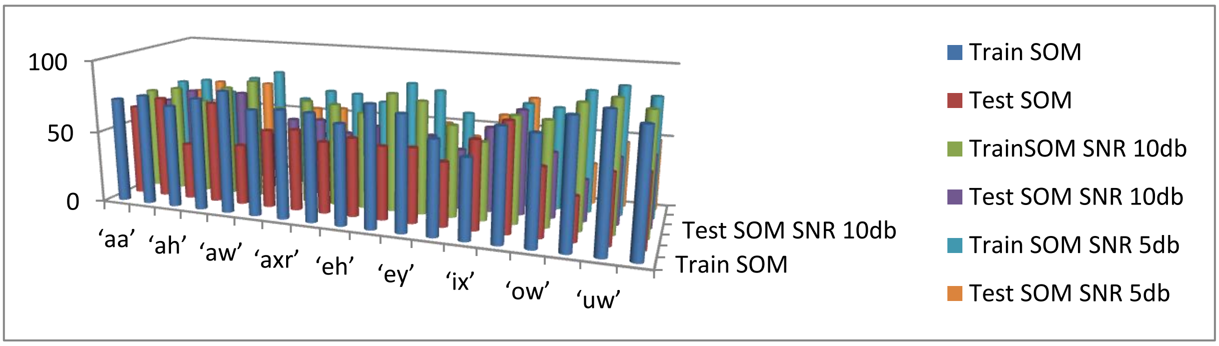

Fig.10 Comparative Vowels recognition rate of TIMIT database by the use of SOM joint to wavelet packet technique in histogram form

We note that there are areas of rapprochement between the generalization (test) rates and the self consistency (train) rates. The averages of generalization rate results based on vowels with and without noise reduction are listed in the table below. The technique of noise reduction is based on the method of adaptative wavelet packet thresholding.

Table IV. Overall rate of SOM generalization based on the vowels

Original Signal	SNR = 0dB without noise reduction	SNR = 0dB with noise reduction	SNR = 5dB with noise reduction	SNR = 10dB with noise reduction
53%	30%	49%	51%	51.5%

In this effect, we note that the lost rates, by the noise presence, are earned via the technique of noise reduction by adaptative wavelet packet thresholding that we have designed.

Over the following figures, we present the results of vowels classification by SOM. The first one is used in adverse environment; the second one is used for a clean speech signal.

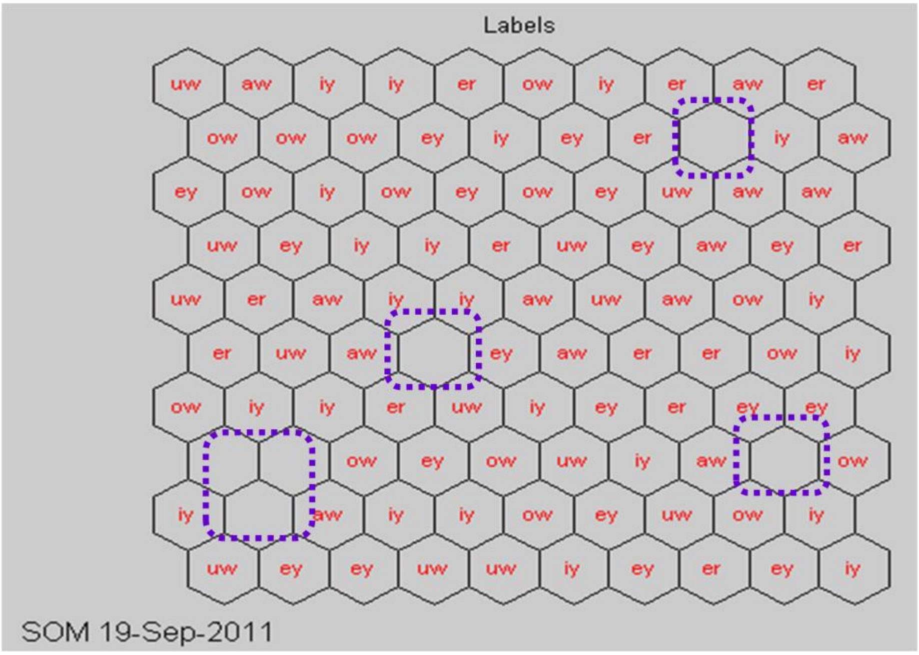
Labels
uw aw iy iy er ow iy er aw er
ow ow ow ey iy ey er iy aw
ey ow iy ow ey ow ey uw aw aw
uw ey iy iy er uw ey aw ey er
uw er aw iy iy aw uw aw ow iy
er uw aw ey aw er er ow iy
ow iy iy er uw iy ey er ey ey
ow ey ow uw iy aw ow
iy aw iy iy ow ey uw ow iy
uw ey ey uw uw iy ey er ey iy
SOM 19-Sep-2011

Fig.11 Representation of the SOM outputs for vowels input degraded by white Gaussian noise, SNR = 5 db, then denoised by wavelet packet at level 3, after 20 learning iterations

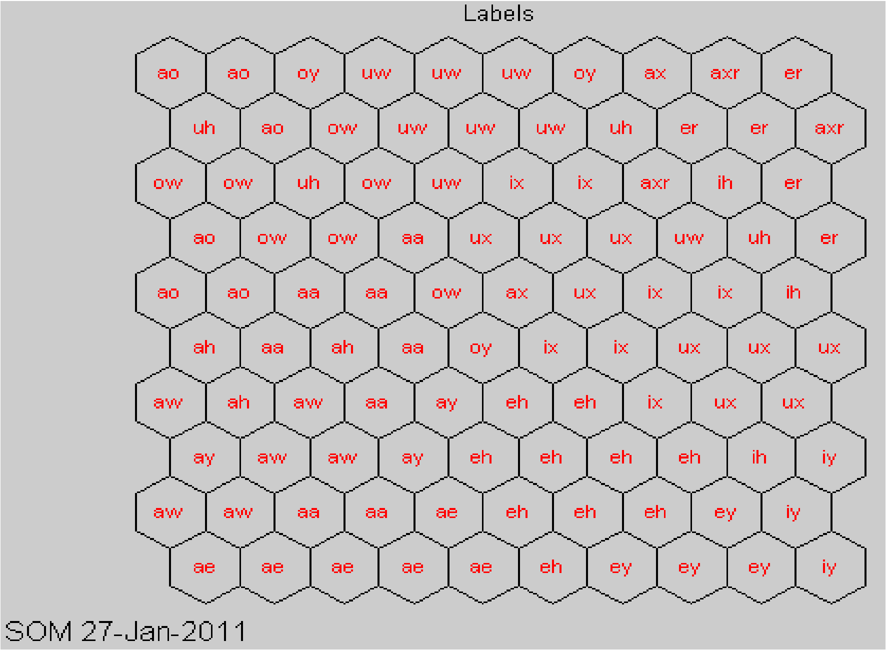
Labels
ao ao oy uw uw uw oy ax axr er
uh ao ow uw uw uw uh er er axr
ow ow uh ow uw ix ix axr ih er
ao ow ow aa ux ux ux uw uh er
ao ao aa aa ow ax ux ix ix ih
ah aa ah aa oy ix ix ux ux ux
aw ah aw aa ay eh eh ix ux ux
ay aw aw ay eh eh eh eh ih iy
aw aw aa aa ae eh eh eh ey iy
ae ae ae ae ae eh ey ey ey iy
SOM 27-Jan-2011

Fig.12 Representation of the SOM outputs for vowels input without noise, after 20 learning iterations

10. CONCLUSION

We have shown that the kohonen SOM map is a tool for unsupervised algorithm so it can be used in the recognition of a dictionary of a large data amounts.

The experimentation of this tool in the phonemes recognition in TIMIT database has validated its robustness. However, we showed that this robustness is weakened when using the SOM for recognizing a noisy signal. In order to improve the robustness of the SOM capacity in any adverse environment, we adopted a strategy of phonemes recognition by the joint use of SOM to the technique of noise reduction by adapted wavelet packet and we have obtained very encouraging results.

11. Reference

[1] Teuvo Kohonen,:‘ Self-organizing maps’, Helsinki university of technology november 16, 2000.

[2] M. Berouti, R. Schwartz et J. Makhool : ‘Enhancement of Speech Corrpted by acoustic Noise’, Proceedings of the IEEE, pp 208-211, 1979.

[3] R.Boite, H Bourlard, T. Dutoit, J. Hancq et H. Leich : ‘Traitement de la parole’, Collection technique et Scientifique des Télécommunications, 1999.

[4] S.F Boll, : ‘Suppression of Acoustic Noise in Speech Using Spectral subtraction’, IEEE Transactions on Acoustics Speech and Signal Processing, Vol 27, pp 113-120, April 1979.

[5] A. Bron : ‘ wavelet-based denoising of speech”, Master Thesis, Technion Israel Institute of Technology, Haifa, 2000.

[6] B. Burke : ‘ Onde et ondelette la saga d’un outil mathématique’, editions, Hubbard, 1996.

[7] R. R. Coifman et M.V. Wickerhauser : ‘Entropy based Algorithm for best basis selection’, IEEE transactions on Information theory, 1992.

[8] D. L. Donoho : “De-noising by Soft Thresholding”, IEEE Transactions on Information Theory, vol 41, n°3, May 1995.

[9] Y. Ephraïm and D. Malah : “Speech enhancement using a minimum meansquare error short-time spectral amplitude estimator”, IEEE Trans. on Acoustics,Speech and Signal Processing, vol 32, n° 6,pp 1109-1121, Dec 1984.

[10] Y. Ephraïm and D. Malah : “Speech enhancement using a minimum meansquare error log-spectral amplitude estimator”, IEEE Trans. on Acoustics, Speech and Signal Processing, vol.33, n°2, pp. 443–445, Apr 1985.

[11] G. Ju, L. Lee : ‘Speech enhancement and improved recognition accuracy by integrating wavelet transform and spectral subtraction algorithm’, 8th European Conference on Speech Communication and Technology, Eurospeech2003.

[12] J.S. Lim et A.V Oppenheim : ‘Enhancement and bandwidth compression of noisy speech’, Proceedings of the IEEE, pp 1586-1604, 1979.

[13] S. Mallat : ‘Une exploration des signaux en ondelettes’, Editions de l’Ecole Polytechnique, 2000.

[14] T. F. Quatieri : ‘Discrete time speech processing : Practice and Application’, Prentice Hall, New Jersey, 2002.

[15] Stephan Bloehdorn and Sebastian Blohm: “A Self Organizing Map for Relation Extraction from Wikipedia using Structured Data Representations”, Institute AIFB, University of Karlsruhe D-76128 Karlsruhe, Germany, 2006.

[16] F. Truchetet : ‘Ondelettes pour le signal numérique’. Collection Traitement du Signal, Edition Hermes, 1998.

[17] S. V. Vaseghi:‘Advanced Signal Processing and Noise Reduction’, John Whiley & Sons edition, 1996.

[18] N. Whitmal, J . C. Rutledge et J. Cohen: ‘‘Reducing correlated noise in digital hearing aids’’, IEEE Eng. Med. Biol. Magazine, vol15, pp. 88-96, 1996.

[19] S. Chang, Y. Kwon, S. I. Yang and I. J. Kim : ‘Speech Enhancement for non stationary noise Environment by adaptive Wavelet Packet’, Proc. International Conference on Signal and Speech Processing, Vol 1, pp 561-564, 2002.

[20] M.S.Salhi and N. Ellouze : ‘An Adaptative Thresholding SOM-Wavelet Packet Model to Improve the Phonemic Recognition Rate’, Setit 2012_IEEE proceeding.