Diyala Journal For Pure Sciences Vol:6 No:1
Jan. 2010

Identification of Noisy Arabic Speech Utterance using Multiwavelet Compression

BY

Taha Mohammed Hasan

Diyala University / Science College / Computer Science Department

Abstract

This paper presents a method that uses the multiwavelet transform in compressing a noisy speech signal and then, testing its effect on the reconstruction of noisy speech utterances after decompression. The ability of multiwavelet transform in compacting signal energy is employed to separate the significant signal features from the noise contribution. The proposed technique showed an improved SNR for the processed speech samples. In each performed experiment, the correlation coefficients are computed and compared to the correlation coefficient yielded by a similar compression scheme based on the cosine transform. The use of the multiwavelet based compression technique is superior in identification of noisy Arabic speech utterance especially at high input SNR values.

Keywords: Speech Compression, multiwavelet Transform, Pattern Recognition.

الخلاصة:

يقدم هذا البحث طريقة تستخدم تحويل المويجات المتعدد في ضغط إشارة كلامية ضوضائية وبعد ذلك يقوم باختبار تأثيره على إعادة بناء نطق الكلام الضوضائي بعد إزالة الضغط تم استخدام قابلية تحويل المويجة المتعددة في ضغط طاقة الإشارة لفصل ميزات الإشارة الهامة من تداخل الضوضاء.تظهر التقنية المقترحة بتحسين نسبة الإشارة إلى الضوضاء للعينات الكلامية المعالجة. تجربة تم اختبارها تم حساب معاملات الارتباط و تقارن إلى معامل الارتباط التي أنتجت من قيل مخطط ضغط مماثل مستند على عملية تحويل كوساين. إن استعمال تقنية ضغط المعتمد على تحويل المويجات المتعدد ساعد على تمييز نطق الكلام العربي المشوه خصوصا في نسبة الإشارة الصحيحة إلى الإشارة المشوهة .

1. Introduction

Speech signals are one of the most important means of communication among the human beings. The speech signal is a slowly time varying signal in the sense that, when examined over a sufficiently short period of time, its characteristics are fairly stationary; however, over long periods of time the signal characteristics change to reflect the different speech sounds being spoken [1]. Consequently, speech signals are represented in the time domain by relatively long sequences which reveal the speech signal energy of the spoken utterances

Speech recognition is affected by noise. Therefore, the probability of detection of an unknown sample within a library decreases as the noise power increases. This means that comparisons of an unknown sample to samples that are very similar to each other leads to a higher probability of error; consequently, to false recognition and an increased error rate. Wavelet-based compression technique can be used to reduce the effect of the noise in order to increase the probability of recognition of the input sample [2]. In general, waveletbased processing techniques of noisy data are important methods that can be applied in data analysis and of great significance in many applications. Such as, signal identification, and pattern recognition. Speech compression is important in mobile communications, to reduce transmission time, and in digital answering machines [3].

An approximation of the speech signal after the compression can be reconstructed by the inverse wavelet transform using a selected number of the wavelet coefficients. Then, the cross-correlation is computed and used for performing speech recognition.

2. Multiwavelet Transform

In multiwavelet transform, we use multiwavelet as transform basis. Multiwavelet functions are functions generated from one single function ψ by scaling and translation:

$$\psi_{a,b}(t)= \frac{1}{\sqrt{a}} \psi_a\left(\frac{t-b}{a}\right) \quad \ldots (1)$$

The mother wavelet $\psi(t)$ has to be zero integral, $\int \psi_{a,b}(t)dt = 0$. From (1) we see that high frequency multiwavelet correspond to $a > 1$ or narrow width, while low frequency multiwavelet corresponds to $a < 1$ or wider width. The basic idea of wavelet transform is to represent any function f as a linear superposition of wavelets. Any such superposition decomposes f to different scale levels, where each level can be then further decomposed with a resolution adapted to that level. One general way to do this is writing f as the sum of wavelets $\psi_{m,n}(t)$over m and n . This leads to discrete wavelet transform:

$$f(t) = \sum_{m,n} \psi_{m,n}(t) \quad \ldots (2)$$

By introducing the multi-resolution analysis (MRA) idea by Mallat [3], in discrete wavelet transform we really use two functions: wavelet function $\psi(t)$ and scaling function $\varphi(t)$. If we have a scaling function $\varphi(t) \in L^2(R)$, then the sequence of subspaces spanned by its scaling and translations $\psi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k)$, i.e

$$Vj = span\ \{ \varphi j,k\ (t), j,k \in Z\} \quad \ldots (3)$$

Constitute a MRA for $L^2(R)$.

$\varphi(t)$ must satisfy the MRA condition:

$$\varphi(t) = \sqrt{2} \sum h(n)\ \varphi(2t-2) \quad \ldots (4)$$

For $n \in Z$. In this manner, we can span the difference between spaces V_j by wavelet functions produced from mother wavelet: $\psi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k)$ Then we have:

$$\psi_{j,k}(t) = \sqrt{2} \sum g(n)\ \varphi(2t-2) \quad \ldots (5)$$

For orthogonal basis we have:

$$g(n) = (-1)^n h(-n+1) \quad \ldots (6)$$

If we want to find the projection of a function $f(t) \in L^2(R)$ on this set of subspaces, we must express it in e as a linear combination of expansion functions of that subspace [4]:

$$f(t) = \sum_{n=-\infty}^{\infty} c(t)\ \varphi(t) + \sum_{n=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d(t)\ \psi_{j,k} \quad \ldots (7)$$

Where $\varphi_k(t)$ corresponds to the space V_0 and $\psi_{j,k}(t)$ corresponds to wavelet spaces.By using the idea of MRA implementation of wavelet decomposition can be performed using filter bank constructed by a pyramidal structure of lowpass filters $h(n)$ and highpass filters $g(n)$[3, 4].

3. General Procedures for Computing DMWT by Using a Critically-Sampled Scheme of Preprocessing First and Second Levels

By using a Critically- Sampled Scheme of Preprocessing (approximation-based scheme of Preprocessing), the DMWT matrix has the same dimensions of the input which should be a square matrix N×N where N must be power of 2. Transformation matrix dimensions which should be equal to image dimensions after preprocessing will be N×N for a Critically-Sampled Scheme of Preprocessing.

There are two orders of approximation types of Critically- Sampled Preprocessing 1st order and 2nd order approximations. For any N×N image matrix, 1st order approximation-based preprocessing can be summarized as follows where every two rows generate two new rows [5].

a) For any odd-row,

$$\text{New odd-row} = \frac{\phi2(1)\ [\text{same odd-row}] - \phi2\ (1/2)[\text{next even-row}] - \phi2\ (3/2)[\text{previous even-row}]}{\phi_2(1)\ \phi_1(1/2)} \quad \ldots (8)$$

b) For any even-row,

$$\text{New even-row} = \frac{\text{Same even-row}}{\phi_2(1)} \quad \ldots (9)$$

The above equations can be written a follows after substituting the values of $\phi_1(1/2)$, $\phi_2(1)$, $\phi_3(1/2)$ for 1st order approximation results:

New odd-row = (0.373615) [same odd-row] + (0.11086198) [next even-row]
+ (0.11086198) [previous even-row] … *(10)*

New even-row = ($\sqrt{2}$ -1) [same even-row] … *(11)*

For 2nd order approximation, Esq. (16) and (17) become,

New odd-row= ($10/8\sqrt{2}$) [same odd-row] + ($3/8\sqrt{8}$) [next even-row]
+ ($3/8\sqrt{2}$) [previous even-row] … *(12)*

New even-row= [same even-row] … *(13)*

It should be noted that when computing the first odd row, the previous even-row in Esq. (6) is equal to zero. In the same manner, when computing the last odd row, the next even-row in Esq. (6) is equal to zero. The same thing is valid for Esq. (8).

It is obvious now why the dimension of the resulting matrix after approximation-based preprocessing has the same dimension as before preprocessing.

The following procedure for computing DMWT by using approximation-based preprocessing is valid for both 1st and 2nd order of approximation with one exception of using Esq. (6) and (7) for 1st approximation preprocessing step and Esq. (8) and (9) for 2nd approximation preprocessing step:

1. *Checking image dimensions*: Image matrix should be a square matrix, N×N matrix, where N must be power of 2. So the first step of the transform procedure is checking image dimensions. If the image is not a square matrix, some operation must be done to the image like resizing the image or adding rows or column of zeros to get a square matrix.
2. *Constructing a transformation matrix*: Using the transformation matrix (12) format, an N/2×N/2 transformation matrix should be constructed using GHM low- and high-pass filter matrices given in (5) and (6) respectively. After substituting GHM matrix filter coefficients values as given by (13), an N×2N transformation matrix results with the same dimensions as the input image matrix dimensions after preprocessing.
3. *Preprocessing rows*: Approximation-based row Preprocessing can be computed by applying Esq. (16) and (17) to the odd- and even-rows of the input N×N matrix respectively for the 1st order approximation preprocessing. For 2nd order approximation preprocessing, Esq. (16) and (17) are replaced with Esq. (18) and (19) for preprocessing for preprocessing the odd- and even-rows of the input N×N matrix respectively. Input matrix dimensions after row Preprocessing is the same N×N.
4. *Transformation of image rows* : can be done as follows:
   I. Apply matrix multiplication to the N×N constructed transformation matrix by the N×N row preprocessed input image matrix.
   II. Permute the resulting N×N matrix rows by arranging the row pairs 1, 2 and 5, 6… N-3, N-2 after each other at the upper half of the resultant matrix rows, then the row pairs 3, 4 and 7, 8… N-1, N below them at the next lower half.
5. *Preprocessing columns*: to repeat the same procedure used in row preprocessing,

I. Transpose the row transformed N×N matrix resulting from step 4.

II. Repeat step 3 to the N×N matrix (transpose the row transformed N×N matrix) which results in N×N columns preprocessed matrix.

6. *Transformation of image columns*: Transformation of image columns is applied to the N×N column preprocessed matrix as follows:

I. Apply matrix multiplication to the N×N constructed transformation matrix by the N×N columns preprocessed matrix.

II. Permute the resulting N×N matrix rows by arranging the row pairs 1, 2 and 5, 6… N-3, N-2 after each other at the upper half of the resultant matrix rows, the row pairs 3, 4 and 7, 8… N-1, N below them at the next lower half.

7. *The final Transformed matrix*: to get the final Transformed matrix the following steps should be applied:

I. Transpose the resulting matrix from column transformation step.

Apply coefficients permutation to the resulting transpose matrix. Coefficients permutation is applied to each of the basic four sub bands of the resulting transpose matrix so that each sub bands permute rows then permute column. The final DMWT matrix using approximation-based preprocessing has the same dimensions, N×N of the original image matrix. The wavelet transform is applied to an input signal at different levels. This is often known as the analysis stage. After passing the signal through the first level filters, the corresponding wavelet coefficients are generated. The coefficients are represented by two sequences. The first sequence corresponds to the low frequencies of the signal, while the second sequence represents the high frequency components.

Similarly, the multiwavelet coefficients of the second level are computed, but taking the low frequencies sequence of the first level as an input to the second stage of the wavelet structure. Fig.(1) Shows the time multiwavelet form of the speech signal corresponding to the sounds in the phrase, and its wavelet transform at the first and second

levels.

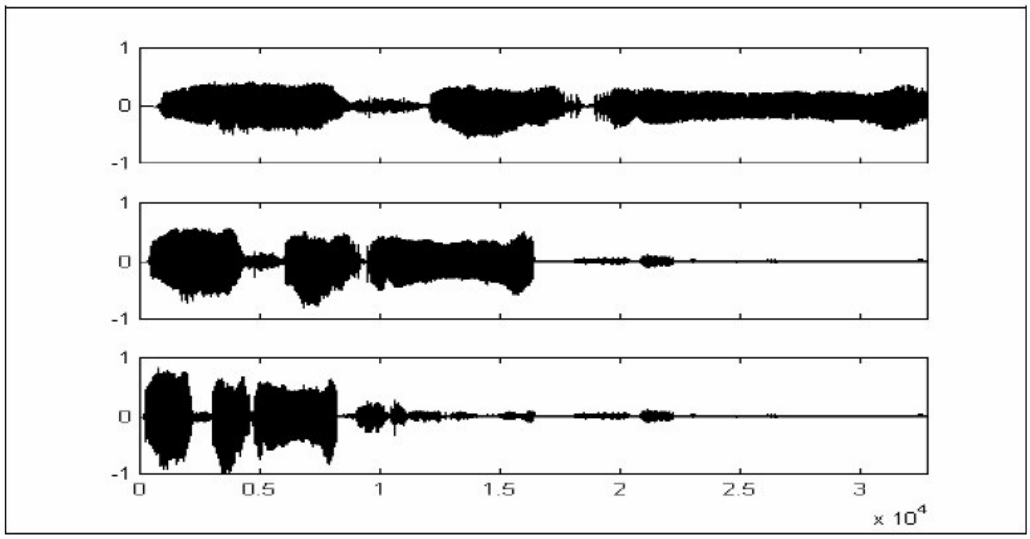

Figure (1): Speech signal sample and its multiwavelet transform at the first and second levels.

In a similar fashion, the transformation process continues until the desired multiwavelet transform level is reached. The number of levels the transformation can take place is depending on the signal size. The first level of the multiwavelet transform is known as the finest scale and the last level in the decomposition is known as the coarse scale.

4. Multiwavelet Compression computation

The energy compaction feature of the *WT* plays an important role in the performance of many of the multiwavelet-based processing techniques. A wavelet-based compression technique works well because the wavelet transform compresses most of the signal energy in a restricted number of large coefficients [4], [8], [9], [10]. In other words, the wavelet transform represents the signal low frequencies with few large coefficients, while it represents the high frequency components by different coefficients [8], [9], [10]. Since noise is related to high frequencies, its effect can be diminished by suppressing the terms associated with these components. After restoring the signal using only the coefficients corresponding to the low frequencies,

two goals can be achieved. First, the signal can be packed with high compression ratios. Second, the signal corrupting noise can be reduced to a minimum level by using an over-sampled scheme of preprocessing (repeated row preprocessing), the DMWT matrix is doubled in dimension compared with that of the input which should be a square matrix N×N where N must be power of 2. Transformation matrix dimensions equal image dimensions after preprocessing. To compute a single-level 2-D Discrete Multiwavelet Transform, the next steps should be followed:

The compression ratio (*CR*) is defined as

$$CR = \frac{L(f(x))}{L(g(x))} \quad \ldots (14)$$

Where *L* is the length of the processed speech signal.

Since the reconstructed signal represents a portion of the original signal, equation (6) can be rewritten as

$$CR\ 1 = \frac{1}{Cp} \quad \ldots (15)$$

Where C_P is the compression percentage of the retained coefficients.

4. System Results

The detection of a noisy speech sample using multiwavelet-based compression technique to reduce the noise effect was used in the identification of unknown speech signal. Fig.(2) illustrates the time multiwavelet forms of three speech utterances that were used in the demonstration of the technique discussed in this paper. The first multiwavelet form represents one of the letters of the "*ba*" spoken by one speaker. The second multiwavelet form is corresponding to the word "*As-salah*", (means: pray) as spoken by a second speaker, and the third waveform is corresponding to the sound, "*Muhamed*", (means: name of prophet) as delivered by a third speaker.

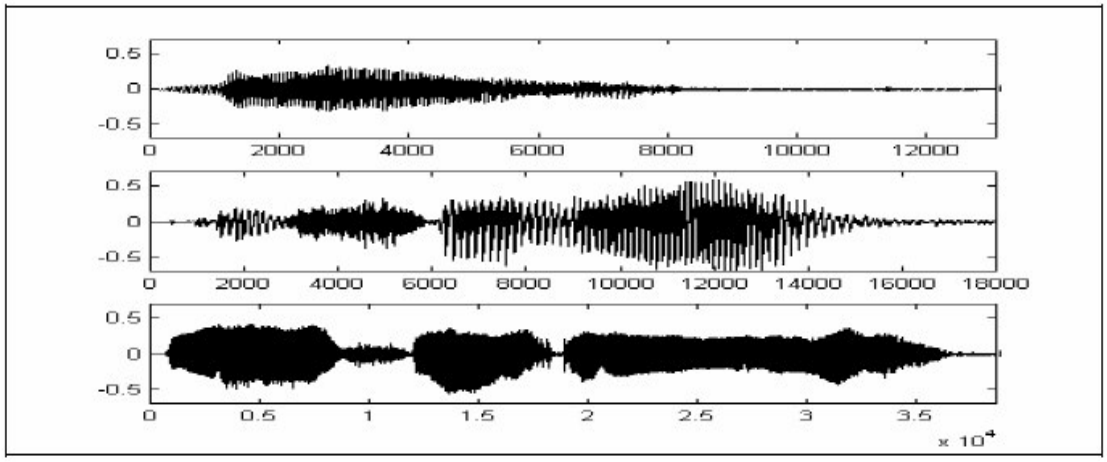

Figure (2): The time waveforms of the speech samples used in the analysis.

For each speech signal used in the analysis, the multiwavelet transform from the fine scale to the coarse scale was computed to generate the wavelet coefficients of the entire signal. The Coiflet multiwavelet basis [4], of order: 5 were used as the mother multiwavelet for the analysis of all the speech signals.

The input speech sample was corrupted with a noise signal of Gaussian distribution. The simulation was performed at different signal to noise ratio values. After the completion of each experiment, the correlation coefficient between the original sample and the reconstructed sample was computed. At each SNR value, the procedure was repeated for compression ratio values that fall in the range (5.00–100.00). Fig.(3) represents typical results of using the second utterance of Fig.(2) in the analysis of the compression technique discussed earlier with an input SNR value of 30 and a at compression ratio of 67.79 (corresponding to retaining 11% of the wavelet coefficients). The subplots of Fig.(3) from left to right and from top to bottom are representing: $^{(1)}$the original, $^{(2)}$noisy, $^{(3)}$restoration after the wavelet compression.

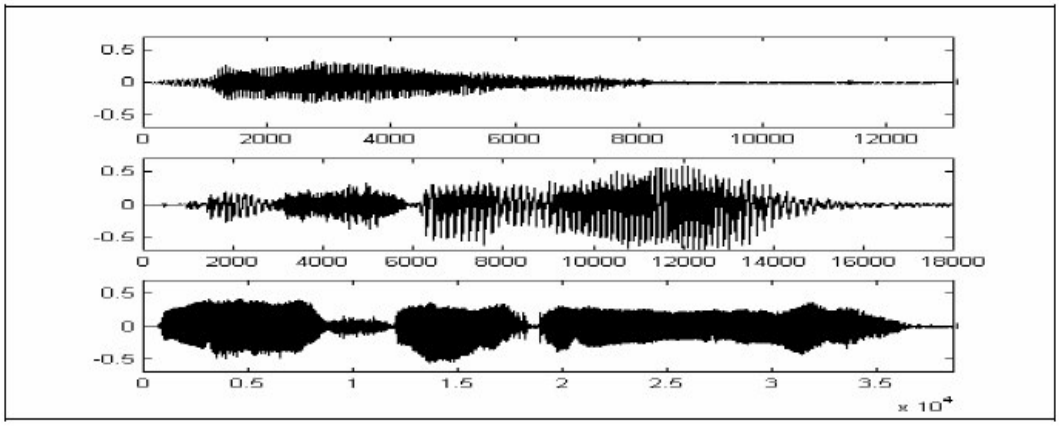

Figure (3): The Multiwavelet forms of a sample speech signal: the original, noisy, reconstruction after Multiwavelet

It can be noticed from the multiwavelet forms that the noise level was reduced from the input sample using the multiwavelet-based compression technique. Visually it is apparent that the restored speech signal after the wavelet compression resembles the original signal. The error associated with this method is much less than the error.

The computations using the two compression methods at different values of the compression ratio are tabulated in Table (1). It can be noticed that the correlation coefficient was improved using the wavelet and the DCT techniques. It can be also observed that as the noise power increases, the wavelet compression technique, performed much better than the DCT technique in improving the correlation coefficient value. This improvement can lead to enhancements in the identification and recognition of the input speech signal.

Table (1): Performance of restoring the second utterance with an input SNR=30.

Compression Ratio (*CR*)	MSE ($\times 10^{-4}$)		SNR		Correlation Coefficient ($\times 10^{-2}$)	
	Multiwavelet	DCT	Multiwavelet	DCT	Multiwavelet	DCT
5.00	2.58	6.38	49.8	20.1	99.01	97.50
6.25	2.25	7.25	57.1	17.7	99.13	97.15
7.69	2.00	8.24	64.4	15.6	99.23	96.75
8.33	1.94	8.66	66.3	14.8	99.25	96.57
10.00	1.94	9.84	66.3	13.1	99.25	96.10
12.50	2.39	11.7	53.8	11.0	99.07	95.34
14.29	2.94	13.0	43.7	9.9	98.85	94.81
16.67	3.72	14.8	34.6	8.7	98.54	94.06
20.00	4.90	17.4	26.2	7.4	98.07	93.00
25.00	6.72	21.1	19.1	6.1	97.35	91.42
50.00	15.0	35.3	8.6	3.6	93.98	85.17
100.00	27.9	51.3	4.6	2.5	88.51	77.51

The graph of the SNR values as a function of the percentage of retained coefficients is plotted in Fig.(4). The chis showing an essential improvement in the signal identification as the number of retained coefficients is decreased. This is due to the suppression of the of the noise contributing coefficients. It is observed that retaining about 10% of the coefficients yields the maximum SNR value, but by constructing the signal from a number of coefficients fewer than this percentage, the SNR value starts to fall down again; and hence the reconstructed speech signal is degraded. It can be also noticed from the graph that the results of using the wavelet-based compression are showing superiority when compared to the DCT.

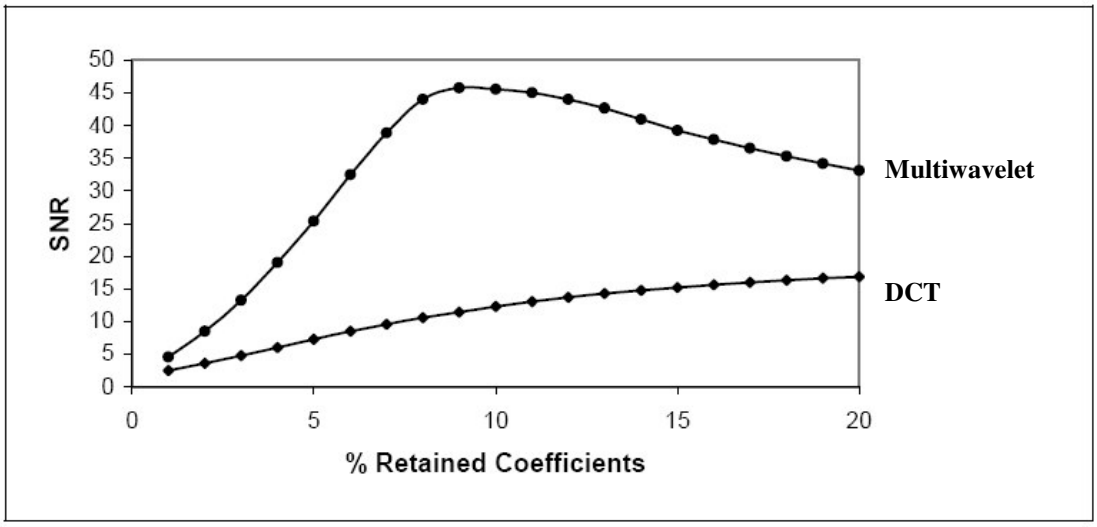

Figure (4): Graph of SNR versus percentage of retained coefficients using multiwavelet and DCT.

The graph of the output SNR values versus the percentage of the retained wavelet coefficients at different values of the input SNR (5, 10, 15, and 20) is depicted in Figure (5).

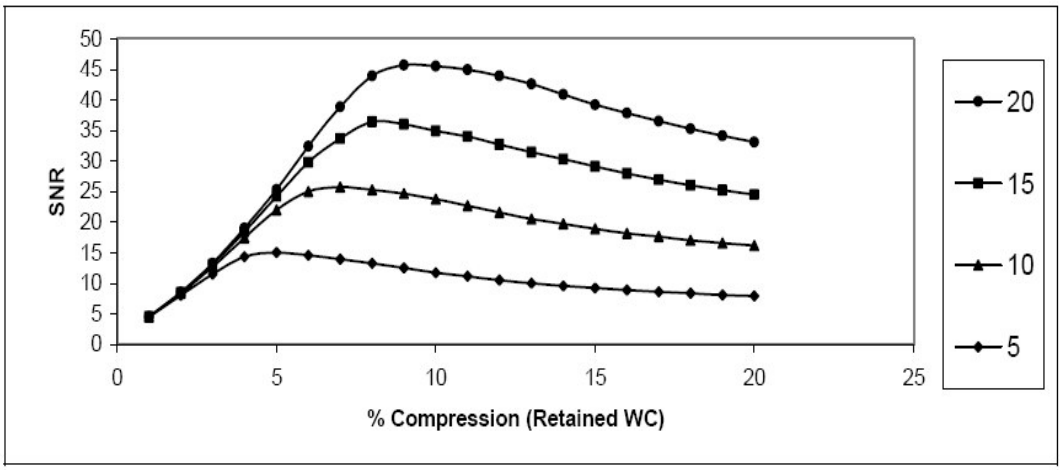

Figure (5): SNR as a function of the percentage of the retained Multiwavelet coefficients at different input SNR (5, 10, 15, and 20) using the second utterance.

It can be noticed form the graph that the compression technique improved the output SNR at a certain percentage value of the retained wavelet coefficients (~5 – 10%). At a certain level of the input SNR, increasing the value of the percentage of the retained wavelet

Coefficients, leads to degradation in the reconstructed signal, because of the increase in the noise power contribution to the speech signal. The correlation coefficient as a function of the input SNR before and after the wavelet compression of the second utterance is depicted in Fig. (6).

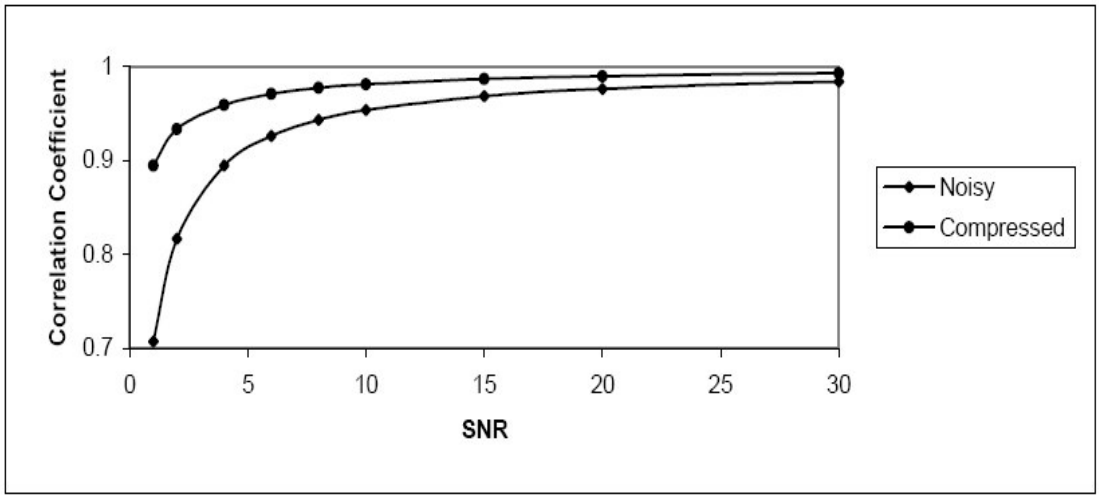

Figure (6): Graph of the Correlation coefficient versus SNR using the second utterance.

The graph shows an improvement in the detection of the input utterance after compression. The plot also illustrates that as the noise power that affects the speech signal is low, in other words at high SNR values, the correlation coefficient is near unity. This leads to better recognition and identification of the input speech signal. It can be also noticed that as the noise power increases, the signal identification starts to drop down. In general, the results out of the wavelet-based compression technique showed superiority when compared with the results of speech signal recognition of the noisy sample especially at low SNR values.

5. Conclusion

In this paper, the identification of noisy Arabic speech utterances using multiwavelet-based compression approach was investigated to study its effect on the speech recognition. The results showed an improved performance using this method when compared to the DCT compression technique. This is concluded from observing the increase in the overall SNR of the signal processed by the wavelet technique in comparison to either the signal before processing, or the signal processed using the DCT. The results showed an improved detection at high noise levels when the wavelet scheme is used over the DCT method. The experiments of the wavelet compression technique showed superiority in recognizing the speech signal when compared with the results of identification of the noisy speech signal especially at low SNR values.

References

[1] Rabiner, L. and Juang, B., *Fundamentals of Speech Recognition*. Prentice Hall, third addition 2005.

[2] Chang, S. Yu, B., and Vetterli, M., "Adaptive Wavelet Thresholding for Image Denoising and Compression.", *IEEE Transactions on Image Processing*, Vol. 9, No. 9, p 1532-1546, September 2000.

[3] Mallat, S. G. "A Theory of Multiresolution signal decomposition: The Wavelet Representation". *IEEE Transform.* 2002*;* 11(7): 674-693.

[4] Hilton, M. L. "Wavelet and Wavelet Packet Compression of Electrocardiograms". *IEEE Trans. Biomed. Eng.*, May 2002; 44: 394-402.

[5] Strang, G. and Nguyen, T., *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 2006.

[6] Donoho, D. and Johnstone, I., "Adapting to Unknown Smoothness via Wavelet Shrinkage." Technical Report, Department of Statistics, Stanford University, 2004.

[7] Young, R., *Wavelet Theory and its Applications*. Kluwer Academic Publishers: Boston, 2003.

[8] Donoho, D., "Wavelet Shrinkage and W.V.D.: A 10-minute tour." Progress in Wavelet Analysis and Applications. *Proceedings of the international Conference on Wavelets and Applications*, 2007-2008.

[9] Djohan, A.; Nguyen, T. Q.; Tompkins, W. J. "ECG Compression Using Discrete Symmetrical Wavelet Transform". *Proc. IEEE Intl. Conf. EMBS*. 2003; 1(2):12-14.

[10] Istepanian, R., Hadjileontiadis, L. and Panas, M., "ECG Data Compression Using Wavelets and Higher Order Statistics Methods." IEEE Transactions on Information Technology in Biomedicine, Vol. 5, No. 2, p 108-115, June 2001.