Nahrain University, College of Engineering Journal (NUCEJ) Vol.16 No.1, 2013 pp.1 - 8

Speech Denoising Using Mixed Transform

Sadiq J. Abou-loukh Ali K. Ibrahim
Electrical Engineering Department, University of Baghdad

Abstract:

This paper presents a mixed transform based speech denoising technique obtained by the combination of multicircularlet and slantlet transforms and thresholding algorithm. It is well known that denoising is a compromise between the removal of the largest possible amount of noise and the preservation of signal integrity. Mixed transform is an intelligent tool for solving speech processing problems such as speech denoising , the general algorithm of speech denoising using discrete wavelet transform (DWT) is presented, followed by the proposed general algorithm of speech denoising using mixed transform. This paper also discusses the effect of using DWT and mixed transform in speech denoising, theirperformances in terms of mean square error (MSE) and peak signal to noise ratio (PSNR) are assessed. Computer simulation results indicate that the mixed transform offers better MSE and PSNR than DWT.

Key Words: Speech Denoising,Multicircularlet,SlantletTransform, DWT.

1. Introduction

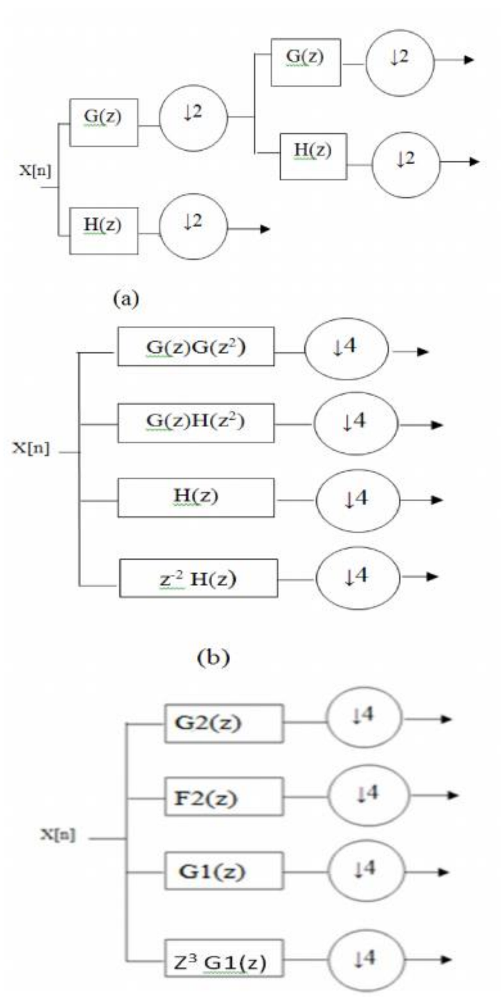
Degradation of the quality of speech caused by the acoustic background noise is common in most ofspeech processing applications such as mobile communication and speech recognition. Therefore, theproblem of removing uncorrelated noise components from the noisy speech, i.e., speech enhancement, has been widely studied in the past and it is still remained as an important issue in the field of speech research [1].Traditional speech denoising techniques are predominantly based on either Wiener filtering or spectral subtraction**.**Although these methods improve the signal tonoise ratio, they distort the signal and alsotend to introduce a perceptuallyannoying residual noise, often referred to as musical noise. Recent noise reduction techniques [2] have exploited themasking properties of the human's auditory system and have resulted in good quality speech with reduced levels of musical noise.

The problem of denoising consists of removing noise from corrupted signal without altering it. Wavelet domain was long been the method of choice to suppress noise. Recently however, methods based on the mixed transformation have become increasingly popular [3]. Utilizationof mixedtransformation in signal and image processing has been found avery useful tool for solving various engineering problems, denoising is one of them. The mixed transform denoising technique is called thresholding; it is a non linear algorithm, which can be decomposed in three steps. The first one consists in computing the coefficients of the mixed transform which is a linear operation. The second one involves in thresholding these coefficients. The last step is the inversion of the threshold coefficients by applying the inverse mixed transform, which leads to the denoised signal. This technique is simple and efficient. However it relies heavily on the choice of the threshold, which in its turn depends on the noise distribution .

2. Slantlet Transform

The slantlet transform (SLT) is based on improved version of the usual discrete wavelet transform(DWT);where the support of the discrete-time basis function is reduced[4]. The SLT is an orthogonal discrete wavelet with two zero moments and with improved time localization ,the basis of the slantlet is based on a filter bank structure where different filters are used for each scale. Consider a usual two-scale iterated DWT filter bank shown in Fig.1(a) and its equivalent form Fig.1(b). The slantlet filter bank is based on the structure of the equivalent form shown Fig.1(b),but it is occupied by different filters that are not products. With this extra degree of freedom obtained by giving up the product form, filters of shorter length are designed satisfying orthogonality and zero moment condition [5].

For two-channel case the Daubechies filter is the shortest filter which makes the filter bank orthogonal and has K zero moments. For K=2 zero moments the iterated filters of Fig.1(c) are of length 10 and 4 but the slantlet filter bank with K=2 zero moments shown in Fig.1(c) has filter length 8 and 4 . Thus the two scale slantlet filter bank has a filter length which is two samples less than that of a two-scale iterated Daubechies-2 filter bank. This difference grows with the increased number of stages. Some characteristic features of the slantlet filter bank are orthogonal, having two zero moments and octave-band characteristic. Each filter bank has a scale dilation factor of two and provides a multi-resolution decomposition.The slantlet filter are piecewise linear. Even though there is no tree structure for slantlet it can be efficiently implement like an iterated DWT filter bank. Therefore, computational complexities of the slantlet are of the same order as that of the DWT,but slantlet transform gives better performance in denoising and compression of the signal[5].



(c)

Fig.1 (a) Two-scale iterated filter bank DWT,
(b) Equivalent form using DWT,
(c) Two-scale filter bank using SLT.

2.1 Derivation ofSlantlet Filters Coefficients

The filters that construct the slantlet filter banks are $g_i(n), f_i(n)$ and $h_i(n)$. The I-scale filter bank has 2I channels. The low pass filter is to be called $h_i(n)$. The filter adjacent to the low-pass channel is to be called $f_i(n)$. Both $h_i(n)$ and $f_i(n)$ are to be followed by down sampling by 2^i. The remaining 2I-2 channels are filtered by $g_i(n)$ and its shifted time-reverse for i=1,…..I-1 .Each is to be followed by down sampling by 2^{i+1} [5].

The sought filter $g_i(n)$ is described by four parameters and can be written as:

$$g_i(n) = \begin{cases} a_0 + a_1 n & for\ n = 0, \ldots, 2^i\text{-}1 \\ b_0 + b_1(n - 2^i) & for\ n = 2^i, \ldots, 2^{i+1}\text{-}1 \end{cases} \quad 1$$

To obtain $g_i(n)$ such that sought I-scale filter bank is orthogonal with 2 zero moments requires obtaining parameters $a_0 . a_1, b_0, b_1$ such that:

$$m = 2^i$$
$$s_1 = 6\sqrt{m/((m^2-1)(4m^2-1))}$$
$$t_1 = 2\sqrt{3/(m.(m^2-1))}$$
$$s_0 = -s_1.(m-1)/2$$
$$t_0 = ((m+1).s_1/3 - mt_1)(m-1)/(2m)$$
$$a_0 = (s_0 + t_0)/2$$
$$b_0 = (s_0 - t_0)/2$$
$$a_1 = (s_1 + t_1)/2$$
$$b_1 = (s_1 - t_1)/2$$

$$h_i(n) = \begin{cases} a_0 + a_1 n & for\ n = 0, \ldots, 2^i\text{-}1 \\ b_0 + b_1(n - 2^i) & for\ n = 2^i, \ldots, 2^{i+1}\text{-}1 \end{cases} \quad 2$$

The same approach work for $f_i(n)$ and $h_i(n)$

$$f_i(n) = \begin{cases} c_0 + c_1 n & for\ n = 0, \ldots, 2^{i-1}\text{-}1 \\ d_0 + d_1(n - 2^i) & for\ n = 2^i, \ldots, 2^{i+1}\text{-}1 \end{cases} \quad 3$$

where:

$$m = 2^i$$

$$u = 1 / \sqrt{m}$$

$$v - \sqrt{(2m^2 + 1)/3}$$

$$a_0 = u.(v+1)/(2m)$$

$$b_0 = u.(2m - v - 1)/(2m)$$

$$a_1 = u / m$$

$$b_1 = -a1$$

$$q = \sqrt{3/(m.(m^2 - 1))} / m$$

$$c_1 = q.(v - m)$$

$$d_1 = -q.(v + m)$$

$$d_0 = d_1.(v + 1 - 2m)/2$$

3. The Mixed Transform

The proposed mixed transform consists of two types of transforms in cascading form to give better coding performance, because cascading transforms further decorrelates the coefficients of the input 2-D signal . The first transform is the proposed multicircularlet transform and the second one is the slantlettransform. Different schemes of applying the slantlettransform to the multicircularlet transform output was implemented. The optimal scheme was by applying the slantlettransform for the subbands containing the approximation signal which is nine subbands of the overall resultant subbands. Figure (2) shows the block diagram of the proposed mixed transform.

The preprocessing step of the mixed transform is the process of getting input rows for the multifactor bank. For data denoising, where one is trying to find compact transform representation for a data set, it is not suitable to use repeated rows scheme, because one is seeking to remove redundancy, not to increase it . It is imperative to find critically sampled transform schemes. This scheme maintains acritically sampled representation. It was shown that the critical sampling preprocessing scheme where the multifilter processes two N/2-point data streams using an approximation method suggested by Geronimo, and adapted here for the multicircularlet transform and will be illustrated in the next section[7].

3.1 A Computation Algorithm of Discrete MulticircularletTransform

By using a critically-sampled scheme of preprocessing (approximation-based scheme of preprocessing), the discrete multicircularlet transform matrix has the same dimensions of the input which should be a square matrix NxN, where N must be a power of 2. Transformation matrix dimensions which should be equal to image dimensions after preprocessing will be NxN for a critical-sampled scheme of preprocessing.

There are two orders of approximation types of critically-sampled preprocessing 1^{st} order and 2^{nd} order approximations [8] &[9].

For 1^{st} order approximation-based preprocessing, where every two rows generates two new rows, can be summarized as follows:-

Newoddrow= (0.373615) [same (oddrow)]
+(0.1108619)[next(evenrow)] 4
+(0.1108619)[previous(evenrow)]

New evenrow =($\sqrt{2}-1$) [same(evenrow)] 5

For 2^{nd} order approximation-based preprocessing,can be summarized as follows:-

Newoddrow= (10/8 $\sqrt{2}$) [same (oddrow)]
+(3/8 $\sqrt{2}$) [next (evenrow)] 6
+ (3/8 $\sqrt{2}$)[previous(evenrow)]

New even-row = [same (even-row)] 7

It should be noted that when computing the first odd-row, the previous even-row in eq.(4) is equal to zero. In the same manner, when computing the last odd-row, the next even-row in eq. (4) is equal to zero. The same thing is valid for eq. (6).

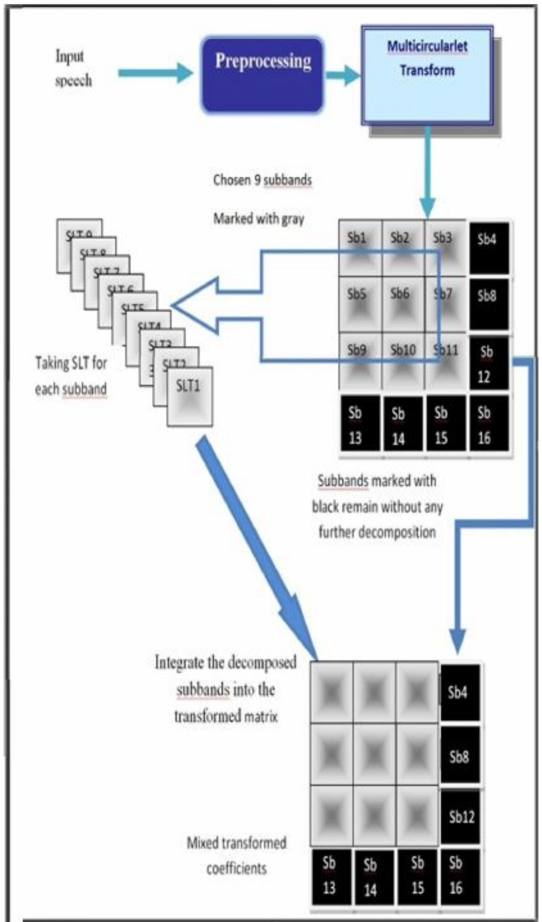


Fig. 2 Mixed transform block diagram

The following procedure for computing algorithm using approximation-based preprocessing is valid for both 1st and 2nd order of approximations with one exception of using eqs.(4) and (5) for 1st order approximation preprocessing step and eqs. (6) and (7) for 2nd order approximation preprocessing step:

a- Checking input dimensions: input vector should be of length N, where N must be a power of two, so resize the signal to the nearest power of 2 value if not a power of 2.

b- Constructing a transformation matrix: using the transformation matrix format, an N/2×N/2 transformation matrix should be constructed using proposed multifactor banks matrix coefficients. After substituting the matrix filter coefficients values, an N×N transformation matrix results with same dimensions of input signal dimensions after preprocessing.

c-Preprocessing rows: approximation-based row preprocessing can be computed by applying eqs. (4) and (5) to the odd- and even-rows of the input N×N matrix respectively for the 1st order approximation preprocessing. For 2nd order approximation preprocessing, eqs. (6) and (7) are used for preprocessing odd- and even-rows of the input N×N matrix respectively. Input matrix dimensions after row preprocessing is the same N×N.

d- Transformation of rows: apply matrix multiplication to the N×N constructed transformation matrix by the N×N row preprocessed input signal matrix.

e-Preprocess columns: to repeat the same procedure used in preprocessing rows,

i- Transpose the row transformed N×N matrix resulting from step d.
ii- Repeat step **c** to the N×N matrix (transpose of the row transformed N×N matrix) which results in N×N column preprocessed matrix.

f-Transformation of columns: transformation of columns is applied next to N×N column preprocessed matrix as follows: apply matrix multiplication to the N×N constructed transformation matrix by the N×N column preprocessed matrix.

g- Final transformed matrix: to get the final transformed matrix:

i-Transpose the resulting matrix from column transformation step.
ii- Apply coefficients permutation to the resulting transpose matrix. The final transformation matrix using approximation-based preprocessing has the same dimensions, N×N, of the original matrix.

4. Proposed Algorithm

The block diagram for implementing the proposed transform in speech denoising is given in figure (3**).**

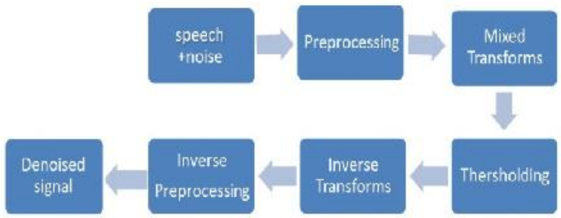


Fig. 3 Block diagram ofthe proposed algorithm

The following steps will be followed in implementation of this algorithm:
Step1: preprocessing, which includes:-
1.input of the speech signal after selecting the sampling rate. Then, the speech signal will be converted to a vector of data (samples).
2. Convert the vector of one dimension to square matrix form (2-dimension).

Step2:apply the multicircularlettransform on the resultant matrix from preprocessing.

Step3: subband decomposition, as in figure(2).

Step4: applyslantlet transform for the nine subbands and the other padding to zero.

Step5:for each subband apply the following:

Obtain each coefficients of noisy speech, the coefficients of each subbandare $g^i{}_j$.

1.Select the threshold type. If the selection is of soft thershold type, then filter these coefficients using the following equations [10]:

$$G^i{}_j = T_s(g^i{}_j, Thv) = \begin{cases} sign(g^i{}_j)^*(|g^i{}_j| - Thv), |g^i{}_j| > Thv \\ 0\ldots\ldots otherewise \end{cases} \quad 8$$

Or the other form

$$G^i{}_j = T_h(g^i{}_j, Thv) = sign(g^i{}_j)^*(|g^i{}_j| - Thv), |g^i{}_j| > Thv \quad 9$$

Where

$$sign(g^i{}_j) = \begin{cases} +1, g^i{}_j > 0 \\ 0, g^i{}_j = 0 \\ -1, g^i{}_j < 0 \end{cases} \quad 10$$

Th_v is the threshold value, $(Th_v = \sqrt{2\sigma^2{}_z log_e}(N))$ is used,and σ^2 is the variance of the samples.
Hence, softthresholding is a mean of translating all coefficients towards zero byacertain amount defined by Th_v.

2. If the selection is hard threshold “kill and keep” strategy or “gating” type, then filter the subband coefficients using this equation:

$$G^i{}_j = T_h(g^i{}_j, Thv) = \begin{cases} g^i{}_j, |g^i{}_j| > Thv \\ 0, otherwise \end{cases} \quad 11$$

Step6: for each new subband apply inverse slantlet transform.
Step7: for the resultant subbands, inverse multicircularlet transform had been taken.
Step8: reconstruct the signal by reshaping it into 1D signal from 2D-signal.

5. ExperimentResults

The proposed speech enhancement algorithm has been tested on the spoken English passages. The speech signals corrupted by additive white Gaussian noise. figure (4) shows an example of the application of OUTmethod to a signal corrupted with random noise.

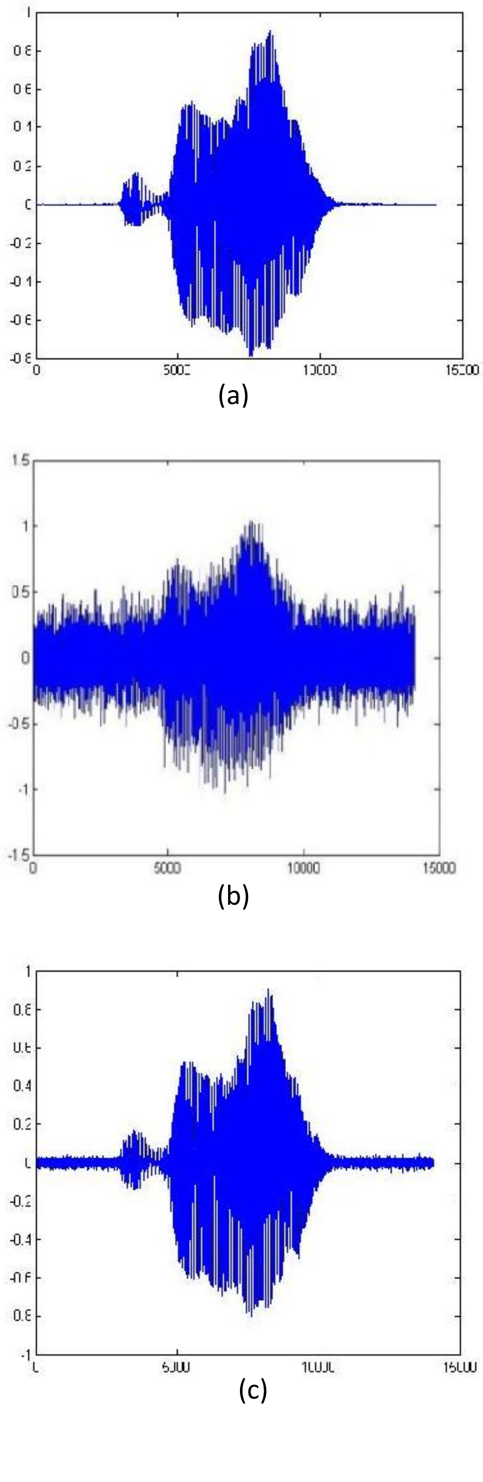


Fig. 4 (a) Clean speech signal,
(b) Corrupted signal with noise,
(c) Enhanced speech signal

To evaluate the performance of a denoising technique the criteria of measuring distortion in reconstructed sound files are defined. This criterion is necessarily applied,which includes signal to noise ratio, peak signal to noise ratio and mean square error.

The above quantities are calculated using the following formats:

1. Peak Signal to Noise Ratio (PSNR):

$$PSNR = 10\log_{10}\frac{MX^2}{\|X-r\|^2} \qquad 12$$

Where, M is the length of reconstructed signal, X is the maximum absolute square value of the signal x and $\|X-r\|^2$ is the energy of the difference between the original and reconstructed signals.

2.Mean Square Error (MSE) is used as a performance index to assess the quality of denoising, and is defined as

$$MSE(dB) = 10\left[\log_{10}\left(\frac{1}{M}\sum_{i=1}^{M}\|x(i)-\hat{x}(i)\|^2\right)\right] \qquad 13$$

Where,$x(i)$is the original signal and $\hat{x}(i)$ is the reconstructed signal.

In the proposed scheme both the complexity of computation and the quality of denoising is considered.

In order to evaluate the performance of the proposed system,different tests had been performed and the results show significant improvementin *SNR* ofreconstructed signals. The properties of tested speech samples are presented in Table1. While Table 2 shows the comparison between the proposed algorithm and DWT on PSNR and MSE

.

Table 1:Properties of tested speech samples				
Signals	The length	Matrix form	Near square power of 2 value	Added or cansul samples
YAM1	6205	80×77	64×64	2109
YAM2	6880	80×86	64×64	2784
YAM3	6913	80×86	64×64	2817
YAM4	8295	80×103	64×64	4199
YAM5	6014	80×75	64×64	1918
YAM6	8935	80×111	64×64	4839

Table 2 : Performance measures of speech signals

Signals	PSNR		MSE	
	DWT	Mixed Transform	DWT	Mixed Transform
YAM1	36.721	39.482	0.5084	0.3603
YAM2	38.58	42.19	0.5676	0.3945
YAM3	33.59	37.53	0.3841	0.2003
YAM4	40.39	45.318	0.4113	0.2377
YAM5	37.613	41.37	0.6354	0.3498
YAM6	38.336	44.173	0.5676	0.3151

6. Conclusions

This paper presents a strategy for digitally implementing a new mixed transform (multicircularlet and slantlet transforms) for speech denoising. The resultant implementations have the approximate reconstruction property, and give stable reconstruction under perturbations of the coefficients. There are, of course, many computing strategies to translate the theoretical results on multicircularlet and slantlet transforms into digital representations.

There are several innovative choices which we now highlight:

1.Mixed transform is important technique in speech denoising applications due to its ability to eliminate the noise. It takes the advantages of multiple transforms.

2.The level of decomposition of multicircularlet transform and the thresholding scheme (e.g. soft or hard) are two other important parameters in multicircularletthresholdingdenoising algorithm. Number of decomposition level of multicircularlet transform depends on the size of the observed data. However, as a rule of thumb, the number of decomposition levels of multicircularlet transform is taken when the corresponding approximate mixedtransform coefficients band at that scale is noiseless. For thresholding scheme, its choice depends on the application, considering if over smoothness or MSE is tolerated more.

3.The proposed algorithm of multidenoising using mixed transform gives better result than the proposed algorithm of denoising using DWT transform in term of PSNR and less MSE.

References

[l] Chang, S., Kwon, Y., Yang, S., and Kim, I., "Speech Enhancement for Non-StationnaryNoiseEnvironment by Adaptive Wavelet Packet", Proc. of IEEE International Conference on Acoustics, Speech, and Signal

Processing (ICASSP), Vol. 1, pp. 561-564, USA, 2002.
[2] Gustafsson, S., Jax, P., and Vary, P.,"A Novel Psychoacoustically Motivated AudioEnhancement Algorithm Preserving Background Noise Characteristics", Proc.of IEEEInternational Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1, pp. 397-400, USA, 1998.
[3] Lallouani, A.,Gabrea, M. & Gargour, C.S., "Wavelet Based Speech Enhancement Using Two Different Threshold-Based Denoising Algorithm", Canadain Conference on Electrical and Computer Engineering, Vol. 1, pp. 315-318, Canada, May 2004.
[4]Chatterje, A.,Maitra,M.&Goswami,S.K., "Classification of Over Current and IinrushCurrent for Power System Reliability using SlantletTransform and Artificial Neural Network",Expert System with Application ,Elsevier,Vol.36,pp.2391-2399,2009.
[5] Selesnick I. W., "The Slantlet Transform" IEEE Trans. Signal Processing, Vol. 47, No. 5, pp. 1304–1313,May 1999.
[6] Gatea, S. M., "A Multi Transform Based Dynamic Time Warping Isolated Word Speech Recognition System", M.Sc. Thesis, University of Baghdad, Electrical Engineering Department, April 2010.
[7] Geronimo, J., Hardin, D., & Massopust, P., "Fractal Functions and Wavelet Expansions Based on Several Scaling Functions", J. Approx. Theory , Vol.78, No.3, PP.373-401, 1994.
[8] Alwan, I.M.,"A Proposal Multicircularlet Mixed Transform and Its Application for ImageCompression", M.Sc.Thesis, University of Baghdad, Electrical Engineering Department, 2009.
[9]Saleh, Z.J., "Video Image Compression Based onMultiwavelets Transform", Ph.D. Thesis, University of Baghdad, Electrical Engineering Department, June 2004.
[10]Mahmoud, W.A. &Ramahi, N., "The Walidlet Transform", Proceeding of 1stInternational Conference of Digital Communication and Computer Application , PP. 382-387, Jordan, 2007

إزالة الضوضاء من الصوت باستخدام التحويلات الخليطة

صادق جاسم أبو اللوخ   علي خليل ابراهيم

قسم الهندسة الكهربائية ، جامعة بغداد

الخلاصة:-

في هذا البحث، نقترح مزيج من التحويلات الخليطة (mixed transforms) المستندة الى ازالة الضوضاء من الصوت و الحاصل عليها بدمج تحويل (multicircularlet) وتحويل المويل (slantlet) مع خوارزمية العتبة (thresholding).تعتبر التحويلات الخليطة أداة ذكية من اجل حل مشاكل الكلام مثل تقليل الضوضاء في الصوت أو المحافظة على شمولية الاشارة.

في هذا البحث تم عرض الخوارزمية العامة لأزالة الضوضاء من الكلام باستخدام تحويل الموجة المتقطع (DWT)، ثم تليها الخوارزمية العامة لتقليل الضوضاء باستخدام التحويلات الخليطة المقترحة. يتطرق البحث مناقشة تأثير استخدام تحويل الموجة المتقطع والتحويلات الخليطة في ازالة الضوضاء. وتم تقييم الاداء بين الطريقتان من خلال حساب معدل مربع الخطأ (MSE) ونسبة الاشارة الى الضوضاء (PSNR). نتائج التمثيل التي تم الحصول عليها تبين ان التحويلات الخليطة افضل من تحويل الموجة المتقطع.