Military Technical College
Kobry El-Kobbah,
Cairo, Egypt

7th International Conference
on Electrical Engineering
ICEENG 2010

Wavelet and Neural Network Method for Speech Enhancement Objective Evaluation

By

K. Daqrouq* M. Alfaouri* T. Abu Hilal ** J.O.Daqrouq *** S.El-Hajjar****

Abstract:

Wavelet Neural Network Evaluation method WNNEM is proposed as a powerful tool for enhanced speech signal evaluation. This objective evaluation measure utilizes Feed forward back Propagation Neural Network FFBNN to train the free of noise signal, and then enhanced signal is simulated to the training output results taken for given target. The distance between simulation and the target, over different wavelet sub bands is studied. Four known speech enhancement method for studying the performance of WNNEM are utilized. The advantage of this method is the evaluation of different band passes of frequency based on wavelet transform by neural network, which is very powerful classification tool. Several objective measures are used to investigate the WNNEM compatibility. Results proved the validity of the proposed method.

Keywords:

Wavelets, neural network, objective evaluation and speech enhancement

---

* Electronic and Communication department-Philadelphia Univ.-Jordan-haleddaq@yahoo.com

** Arab Academy for Banking & Financial Sciences, Computer Information Systems tariq_abuhilal@du.edu.om

*** IT Department- Gazco Company-Abu Dhabi, jdaqrouq@gasco.ae

**** Dhofar University, Foundation Department, Oman

1. Introduction:

The types of deformation introduced by speech enhancement algorithms can be broadly divided into two kinds: the distortions that change the speech signal itself (called speech distortion) and the distortions that change the background noise (called noise distortion). Of these two types of deformation, listeners seem to be influenced the most by the speech distortion when making judgments of overall quality [1], [2]. the most accurate method for evaluating speech quality is through subjective listening tests. Although subjective evaluation of speech enhancement algorithms is often accurate and reliable (i.e., repeatable) provided it is performed under stringiest conditions (e.g., sizeable listener panel, inclusion of anchor conditions, etc. [4]–[7]), it is costly and time consuming. For that reason, much effort has been placed on developing objective measures that would predict speech quality with high correlation. Many objective speech quality measures have been proposed in the past to predict the subjective quality of speech [4]. Most of these measures, however, were developed for the purpose of evaluating the distortions introduced by speech codecs and/or communication channels [7]–[12].

Different methods have been proposed for speech enhancement systems evaluation. All of these methods are based on comparison of original signal with enhanced signal by relative ratio measure or distance measure. The most popular measure, which gives a measure of the signal power improvement related to the noise power is SNR [13], and segmental SNR (segSNR) [14]. From spectral domain evaluation algorithm, we can mention Weighted Slope Spectral distance (WSS) [15]

$$d_{WSS}=\frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{I=1}^{K}W(I,M)\left(s_C(I,M)-s_P(I,M)\right)^2}{\sum_{I=1}^{K}W(I,M)} \tag{1}$$

Where $W(I,M)$ is the weight placed on *I*th frequency band, K is the number of bands and M is the number of frames in the signal. $s_C(I,M)$ and $s_p(I,M)$ spectral are the slope of the clean and enhanced signals, respectively. Hu and Loizou in [3], used the value of K as 25.

Cepstrum distance has been used in as a difference of original signal cepstrum and enhanced signal cepstrum [3]

$$d_{CEP}(\vec{C}_C,\vec{C}_P)=\frac{10}{\log 10}\sqrt{2\sum_{k=1}^{p}(C_C(k)-C_P(k))^2} \tag{2}$$

where $\vec{C}_C$ and $\vec{C}_P$ are original signal cepstrum and enhanced signal cepstrum vectors, respectively. In literature, LPC-based objective measures have been utilized, such as log-likelihood ratio (LLR) [14]

$$d_{LLR}(\vec{a}_P,\vec{a}_C)=\log\left(\frac{\vec{a}_P \mathrm{R}_C \vec{a}_P^T}{\vec{a}_C \mathrm{R}_C \vec{a}_C^T}\right) \tag{3}$$

Where $\vec{a}_C$ and $\vec{a}_P$ are LPC vectors of the original and enhanced signals, respectively. R_C is autocorrelation of original signal.

In [3] composite evaluation is proposed, which was obtained as a correlation between objective and subjective evaluation, by using two merits: correlation coefficient and standard deviation.

Here, a new evaluation measure is proposed by Continuous Wavelet Transform (CWT). This measure is obtained by calculating the differences between CWT of the original signal and the enhanced signal over three levels: low, medium and high. And then, average of standard deviations is obtained

$$d_{CWT} = \frac{\sum_{j}^{J} \sqrt{E\left[(C_j - \overline{C}_j)^2\right]}}{3} \quad \text{for } j = 5, 10 \text{ and } 15 \tag{4}$$

Where $C_j = CWT_j(s) - CWT_j(\tilde{s})$ and $\overline{C}$ is a mean value. The level determination as 5, 10 and 15 is according to the sampling frequency of the speech signal [16]-[17]. These levels present low, medium and high pass bands of the signal frequency. Thus, the utilizing this measure helps studying the difference between filtered and clean signals via three bands, instead of whole signal overlapped bands.

2. Applied Speech Enhancement Methods:

In this paper we utilize four published speech enhancement method for studying the performance of WNNEM:

1. Discrete Wavelet Filtration Method (DWFM)

This method involves multistage wavelet filtration based on convolution with Reverse Biorthogonal Wavelets [18]. This method is based on filtration the low frequency and high frequency parts separately, without thresholding (cutting) the values, which leads to lose the essential speech information.

2. Donoho Thresholding Method (DTM)

Donoho and Johnstone in [19] presented soft thresholding function as follows

$$T_S(\lambda, w_k) = \begin{cases} \operatorname{sgn}(w_k)(|w_k| - \lambda) & \text{if } |w_k| > \lambda \\ 0 & \text{if } |w_k| \le \lambda \end{cases} \tag{5}$$

Where w_k is the wavelet coefficient, and λ is the universal threshold for WT

$$\lambda = \sigma\sqrt{2\log(N)} \tag{6}$$

Where $\sigma = {MAD}/{0.6745}$ is the noise level, *MAD* is the absolute of median estimated on first scale, and N is the length a speech frame (de-noised) signal. For Wavelet Packets Transform, the threshold is calculated by

3. Massart Thresholding Method (DTM)

Birgé and Massart in [20] proposed a level-dependent threshold, which can be explained by the following sequent concepts

- $[C,L]$ is the wavelet structure of the decomposed signal to be enhanced (de-noised), at level $j = length(L) - 2$.
- α and M are real numbers greater than 1.
- T is a vector of length j; $T(i)$ contains the threshold for level i.
- N_{KEEP} is a vector of length j; $N_{KEEP}(i)$ contains the number of coefficients to be kept at level i.

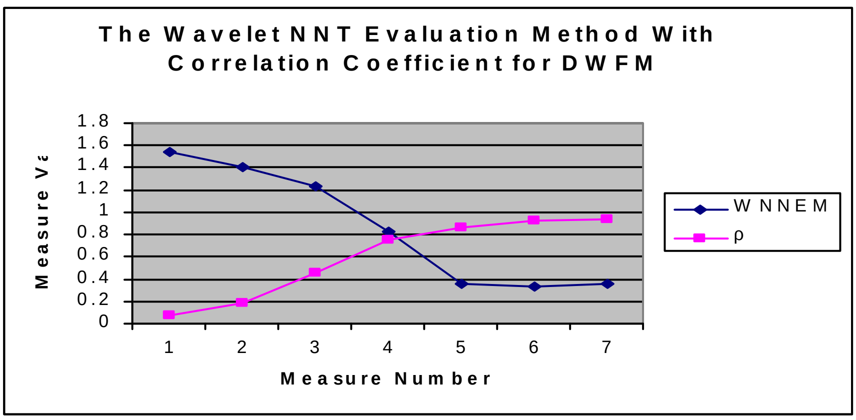
The strategy definition:

1) For level $j+1$, everything is kept.

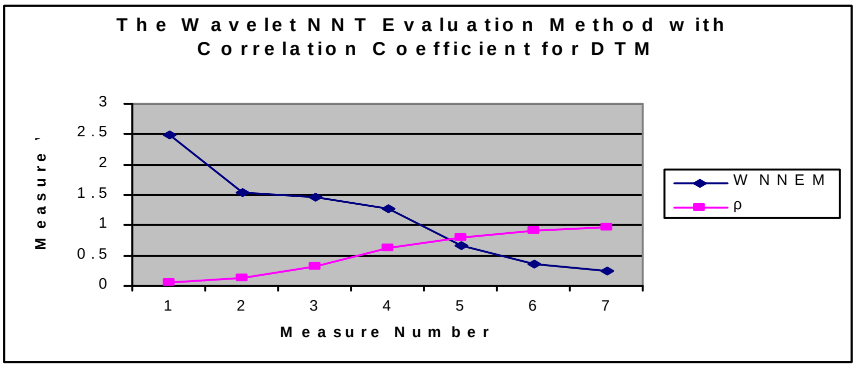
2) For level i from 1 to j, the ni largest coefficients are kept with $ni = M \ (j+2-i)^{\alpha}$. Typically α = 3 for de-noising. Recommended values for M are from $L(1)$ to $2*L(1)$.

1. Kalmen Filter Method (KFM)

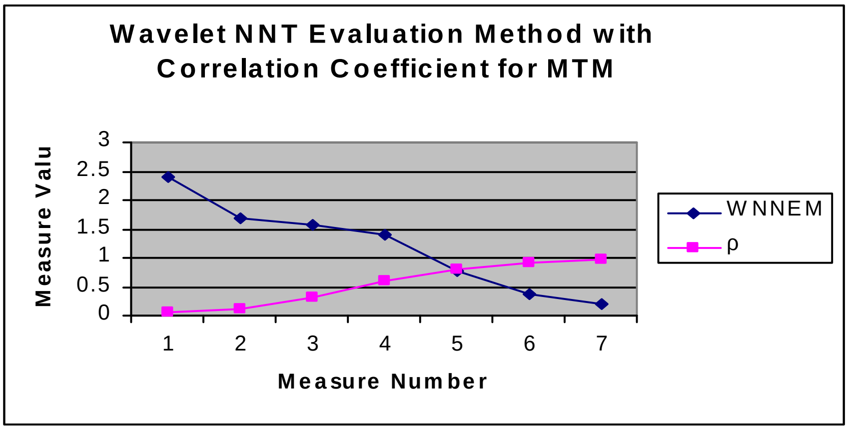
The time-varying Kalman filter is a generalization of the steady-state filter for time-varying systems or LTI systems with nonstationary noise covariance. More about This filter can be found in [20].



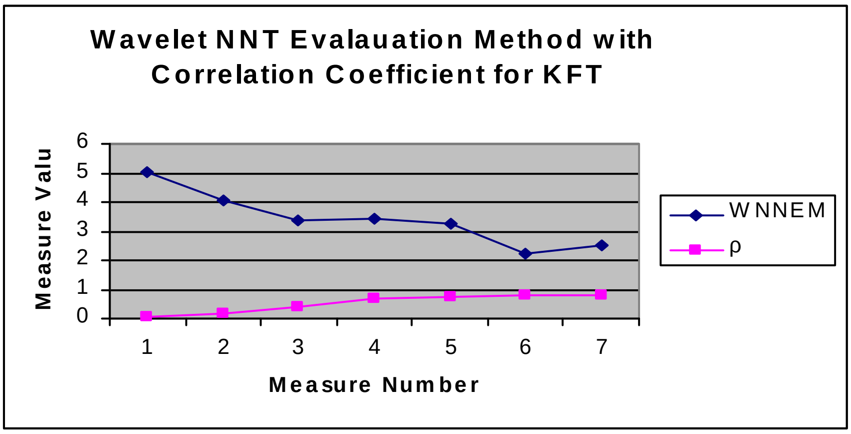
a.



b.



c.



d.

Figure (1): *WNNEM with correlation coefficient for enhanced signal by a. DWFM. b. DTM. c. MTM. and KFT*

Table (1): *Objective measures results*

OBJ. EVALUATIOM	DWFM	DTM	MTM	KFM
SNR	5.1084	2.0276	2.0074	5.5568
P	0.8083	0.7112	0.6991	0.7033
MSE	0.0001	0.0001	0.0002	0.0005
Md_{CEP}	0.39	0.47	0.1869	0.3723
d_{CWT}	0.0204	0.0212	0.0214	0.0389
WNNEM	0.8465	0.935	0.9458	2.5682
Corrupted Signal SNR=-4.5366 dB				

Table (2): *The relation between DWFM and SNR*

SNR [dB]	DWFM		DTM		MTM		KFM	
	SNR	WNNEM	SNR	WNNEM	SNR	WNNEM	SNR	WNNEM
-12.3455	1.5826	1.2313	-1.5354	1.4750	-1.1732	1.5387	1.2194	3.3654
-4.5366	5.1084	0.8465	2.0276	0.935	2.0074	0.9458	5.5568	2.5682
-1.0147	6.6735	0.5089	5.0762	0.7773	4.1444	0.7296	8.3791	3.5145
0.7086	7.7309	0.4732	6.1513	0.4987	6.1421	0.5776	9.9932	3.9716
4.4268	9.9919	0.4663	7.8958	0.5267	8.3389	0.3450	13.5011	3.0626
14.3954	15.0678	0.2554	12.0697	0.2185	13.6781	0.1527	23.3178	2.5643
20.9316	17.0567	0.2991	12.795	0.1968	14.727	0.1630	29.7811	2.3318

3. Method:

In this paper, we use FFBNN for enhanced signal evaluation by comparing with original free of noise signal. The input P matrix contains N columns of wavelet coefficients; each column presents 2500 wavelet coefficients:

$$P = \begin{bmatrix} I_{F0} & I_{F0} & \cdots & I_{F0N} \\ I_{F1} & I_{F1} & \cdots & I_{F1N} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ I_{F4} & I_{F4} & \cdots & I_{F4N} \end{bmatrix} \quad (7)$$

where *N* is the column of 2500 wavelet coefficients . To take matching decision, this matrix is given to a FEBNN to be trained with the following binary target for N=9.

$$T = \begin{bmatrix} 1\ 0\ 1\ 0 \ldots \\ 0\ 1\ 1\ 0 \ldots \\ 0\ 0\ 0\ 1 \ldots \\ 0\ 0\ 0\ 0 \ldots \end{bmatrix} \tag{8}$$

To implement FFBNN, we can use matlab neural network toolbox by function newff, tansig transfer function and trainlm back propagation training function:

```
net=newff(minmax(P),[5 4],{'tansig'},'trainlm');
```

This commend builds a network of three layers: 5 neurons input layer, 5 neurons hidden layer and 4 neurons output layer. After training with the target by [net,tr]= train(P, T);

We simulate the network outputs (the weights and the biases) with enhanced signal to be evaluated, by

```
T_result=sim(net, pt);
```

Now T_result indicates the net output of enhanced signal according to free of noise signal. Now the quality measure is calculated by the distance between T_result and the target.

4. Results and Discussion:

Tested speech signals were recorded via PC-sound card, with a spectral frequency of 4000 Hz and sampling frequency 16000 Hz, over about 2 sec. time duration. For each speaker, the Arabic expression, which sounds "besmeallahalrahmanalraheem", that means in English "In the Name of God the merciful, the compassionate", was recorded 10 times by each speaker. 4 females and 18 males participated in utterances recording. The recording process was provided in normal university office conditions.

The experimental part of this research is introduced by utilizing several objective measures such as d_{CWT}, modified Cepstrum distance

$$Md_{CEP}(\vec{C}_C, \vec{C}_P) = \log 10 \sqrt{2\sum_{k=1}^{p}(C_C(k) - C_P(k))^2} \tag{9}$$

and modified LPC-based log-likelihood ratio Md_{LLR}

$$Md_{LLR} = \left| \mathrm{Re}\left(\log\left(\frac{\sum_{n}^{N} a_s(n)\mathrm{R}_s}{\sum_{n}^{N} a_{\tilde{s}}(n)\mathrm{R}_{\tilde{s}}} \right)\right)\right| \qquad (10)$$

Where $a_s(n)$ and $a_{\tilde{s}}(n)$ are LPC of the original and the enhanced signals, respectively. R_s, $\mathrm{R}_{\tilde{s}}$ are autocorrelation of original and enhanced signals. The modification is done to make the two measures more suitable for our research. Correlation coefficient and MSE are also used.
Table 1 contains objective measures results taken for corrupted signal SNR equal to -4.5366 dB. These results were calculated for four enhancement methods mentioned in section three. We can see clearly the correlation between the conventional objective methods and the proposed method DWFM.

The relation between WNNEM and SNR is presented in table 2. we can see that there is a compatibility between these two measures over four enhancement methods mentioned in section 3. DWFM showed best SNR improvement with best WNNEM (smallest).

In figure 1 we illustrate the relation between WNNEM and correlation coefficient. These results were calculated for seven SNR levels for corrupted signal, vary from -30 dB to 17 dB, for four enhancement methods mention in section 3. The figures illustrate that there are correct relation ship between WNNEM and correlation coefficient, because when correlation coefficient is small then WNNEM as an error is high, but when it is high WNNEM as an error is small.

5. Conclusions:

In this paper, Wavelet Neural Network Evaluation method is presented. Feed forward back propagation neural network is proposed to train the free of noise signal, and then enhanced signal is simulated to the training output results taken for given target. Four published speech enhancement method for studying the performance of Wavelet Neural Network Evaluation method are utilized. The advantage of this method is the evaluation of different band passes of frequency based on wavelet transform by neural network which is very powerful classification tool. Several objective measures are used to compare the proposed method with. Results proved the validity of the proposed method.

References:

[1] Y. Hu and P. Loizou, "Subjective comparison of speech enhancement algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. 1, pp. 153–156.

[2] Y. Hu and P. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, pp. 588–601, 2007.

[3] Y. Hu and P. Loizou, Evaluation of Objective Quality Measures for Speech Enhancement, IEEE Tran. on Audio, Speech, and Language Processing, Vol. 16, N0. 1, Jan. (2008).

[4] S. Quackenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice-Hall, 1988

[5] Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," ITU-T, ITU-T Rec. P. 835, 2003.

[6] P. Kroon, , W. Kleijn and K. Paliwal, Eds., "Evaluation of speech coders," in *Speech Coding and Synthesis*. New York: Elsevier, 1995, pp. 467–494.

[7] L. Thorpe and W. Yang, "Performance of current perceptual objective speech quality measures," in *Proc. IEEE Speech Coding Workshop*, 1999, pp. 144–146. [5] T. H. Falk and W. Chan, "Single-ended speech quality measurement using machine learning methods," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 1935–1947, Nov. 2006.

[8] L. Malfait, J. Berger, and M. Kastner, "P.563-the ITU-T standard or single-ended speech quality assessment," *IEEE Trans. Audio, Speech,Lang. Process.*, vol. 14, no. 6, pp. 1924–1934, Nov. 2006.

[9] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. 2, pp. 749–752.

[10] "Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU, ITU-T Rec. P. 862, 2000.

[11] R. Kubichek, D. Atkinson, and A. Webster, "Advances in objective voice quality assessment," in *Proc. Global Telecomm. Conf.*, 1991, vol. 3, pp. 1765–1770.

[12] Turbin and N. Faucheur, "Estimation of speech quality of noise reduced signals," in *Proc. Online Workshop Meas. Speech Audio Quality Netw.*, (2007), [Online]. Available: http://wireless. feld.cvut.cz/mesaqin2007/contributions.html.

[13] Hansen and J. Pellom B, “An effective quality evaluation protocol for speech enhancement algorithms,” in *Proc. Int. Conf. Spoken Lang. Process.*, (1998), vol. 7, pp. 2819–2822.

[14] Klatt D. , “Prediction of perceived phonetic distance from critical band spectra,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, (1982), vol. 7, pp. 1278–1281.

[15] K.Daqrouq and N.M Abu-Sheikha, Heart Rate Variability Analysis Using Wavelet Transform, Asian Journal for Information Technology, Pakistan, Vol. 4, Number4, 2005.

[16] K. Daqrouq and N Abu-Isbeih Ibrahim N., Arrhythmia Detection Using Wavelet transform, IEEE Region 8, EUROCON 2007, Warsaw, Poland, Sept., 2007.

[17] K. Daqrouq, A. Al-Qawasmi, The Study of Wavelet Filters Speech Enhancement Method, MIC-CCA2009, 26-28 Oct. 2009

[18] D.Donoho, I. Johnstone, (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.

[19] L. Birgé, P. Massart (1997), "From model selection to adaptive estimation," in D. Pollard (ed), Festchrift for L. Le Cam, Springer, pp. 55-88.

[20] M.J. Grimble, Robust Industrial Control: Optimal Design Approach for Polynomial Systems, Prentice Hall, 1994, p. 261 and pp. 443-456.