International Journal of Scientific and Technological Research www.iiste.org
ISSN 2422-8702 (Online)
Vol 1, No.2, 2015

Speech Enhancement Based On Dual Tree Complex Wavelet Transform

Haci Tasmaz
Department of Electrical and Electronics Engineering,
Adiyaman University, 02040, Adiyaman, Turkey
E-mail: htasmaz@adiyaman.edu.tr

Abstract

The dual tree complex wavelet transform (DT-CWT) is an efficient tool for many signal processing applications (i.e. image coding, image fusion, image enhancement, pattern recognition and image resolution enhancement etc.), since it has advantages of near shift-invariance, directional selectivity for two or more dimensions and low computational complexity. In his paper, a new speech enhancement method based on the DT-CWT is proposed in order to exploit the aforementioned advantages of the DT-CWT in speech enhancement. An efficient estimator, multiplicatively modified log-spectral amplitude (MM-LSA) estimator, is used for the enhancement of noisy subband wavelet coefficients. The objective tests (SNR, segmental SNR and PESQ-MOS) are used to evaluate the performance of the proposed speech enhancement method. The performance results of the proposed method are compared with those of the other well-known wavelet-based methods. The objective test results and experimental results show that the DT-CWT outperforms the standard wavelet-based methods in speech enhancement.

Keywords: Speech Enhancement, Dual Tree Complex Wavelet Transform, Multiplicatively Modified Log-Spectral Amplitude Estimator.

1. Introduction

When a microphone is used in a noisy media, the quality and the intelligibility of received speech is degraded due to additive background noise. This corruption can be very bothersome particularly in mobile communications. Using speech enhancement algorithms is recommended in such communication devices [1]. The aim is to enhance the quality and intelligibility of speech while decreasing the background noise without causing significant distortion. The areas where speech enhancement is necessary include car interiors for cellular, helicopter and aircraft cockpits, automatic speech recognition systems, hearing aids and cochlear implants. The problem of speech enhancement is still an open area for researchers and engineers.

The spectral subtraction based methods are the most popular among the many single-channel speech enhancement methods existing because of their effectiveness and simplicity [2-3]. These methods are based on subtracting the average noise spectrum from the noisy speech spectrum. The main drawback of spectral subtraction is that, it causes residual and unnatural musical noise.

The single channel speech enhancement algorithms based on minimum mean-square error (MMSE) estimation have received significant attention in the past decades [4-9] and widely used by researchers because of their performance in elimination of musical noise. These methods are based on a priori signal-to-noise ratio (SNR) estimation, Gaussian statistical model and short-time Fourier transform (STFT). The magnitude estimate of clean speech signal can be obtained from the noisy speech signal if a spectral gain is applied to each frequency bin of a short-time frame. The spectral components are assumed to be independent Gaussian variables, so that gain is adjusted as a function of the local SNR at each frequency bin. The difficulty of correctly estimating a priori SNR during noise-only frames is the main limitation of MMSE based estimators.

In the past two decades, the wavelet transform has become a popular tool in many signal processing applications [10-12]. The wavelet transform has been designed to overcome the shortcomings of STFT which provides limited flexibility in time-frequency representation of nonstationary speech signals. There is a tradeoff between time and frequency resolutions in the STFT. Once the frame length is selected, the time resolution is fixed for all frames. On the other hand, the DWT allows more flexible time-frequency analysis of speech signals. Low frequency resolutions of signal can be obtained at high time resolutions while high frequency resolutions can be obtained at low time resolutions. Multi-resolution analysis, perfect reconstruction, good energy localization and compact support properties have made DWT an efficient tool for many signal processing applications [10-12]. Standard DWT is a non-redundant representation in transform domain having $O(n)$ computational complexity.

Decomposition tree structure corresponding to standard DWT does not provide uniform frequency resolutions for all subbands since only approximate coefficient is decomposed at each level of decomposition. If the detail coefficient is also decomposed at each level, the new transform is called as wavelet packets transform (WPT). The WPT provides uniform frequency resolutions for all subbands at each level of decomposition, however; it causes extra computational complexity of $O(n \log n)$.

The WPT allows accurate representation of signals via best basis algorithm using a suitable cost function. These cost functions can be entropy, mean square error (MSE) or number of nonzero coefficients after thresholding [13-15]. Although selection of best basis reduces shift sensitivity, the standard implementation of WPT is still shift-variant.

Moreover, WPT allows perceptual filterbank representation employing critical-band decomposition. Adjusting the subbands of the WPT according to critical-bands of the human auditory system, a perceptual filterbank which leads to efficient speech enhancement algorithms can be designed [16-17].

The difficulties encountered during implementation have led to developing lifting wavelet transform (LWT) or “lifting scheme” [18-20]. The lifting scheme is a spatial domain method where dilation and translation in the classical DWT are not needed. The lifting scheme has the following advantages over the classical wavelet transform: i) It is a spatial domain method, ii) it allows easier implementation, iii) it allows faster and in-place calculations, iv) it allows nonlinear, adaptive, irregularly sampled and integer to integer wavelet transforms.

Though DWT is a powerful tool; it has the following drawbacks for certain signal and image processing applications [21]. a) Lack of shift-invariance, b) poor directionality, c) oscillation of wavelet coefficients at singularities, d) use of short support wavelets only, e) use of very redundant representations, f) aliasing due to down sampling and non-idle filtering during analysis.

One way of overcoming these limitations is to use stationary (or undecimated) wavelet transform. The stationary wavelet transform (SWT) has similar tree structure as DWT but without any down-sampling stage after filtering. The perfect reconstruction criterion is preserved through zero-padding interpolation of low-pass and high-pass filters. The SWT results in approximate and detail coefficients of the same length at each level of decomposition. Though SWT is shift-invariant, it has large redundancy and high computational complexity of $O(n^2)$ [22-23].

Another way is employing dual tree complex wavelets transforms (DT-CWT) [24-28]. The DT-CWT uses real filtering which decomposes the signals into real and imaginary parts. The DT-CWT implementation results in two DWT trees (real and imaginary trees) operating in parallel. It uses real filters with some special constraints in order to satisfy perfect reconstruction and near shift-invariance properties. The DT-CWT implementation results in a reasonable computational complexity of $O(2n)$.

Although, successful results are obtained via the DT-CWT in many signal processing applications (i.e. image coding [15], image fusion [21], image enhancement [20, 24], pattern recognition [30] and image resolution enhancement [31]), (as much as we know), there is no study published on application of the DT-CWT in speech enhancement.

In this paper, a speech enhancement method based on the DT-CWT is proposed in order to overcome the disadvantages of classical DWT. The MM-LSA estimator is employed for the enhancement of noisy subband coefficients.

The performance of the proposed method is evaluated objectively using SNR, SegSNR and PESQ-MOS tests and subjectively by visual inspection on the enhanced speech signals and spectrograms. The performance results of the proposed method are compared with those of the other wavelet-based methods (DWT, WPT, SWT and LWT). The experimental results and the objective test results prove the superiority of the DT-CWT over the standard wavelet transform based methods in speech enhancement.

2. 1-D Dual Tree Complex Wavelet Transform

To overcome the above mentioned disadvantages of the standard DWT, Kingsbury [24-28] introduced the DT-CWT which has the advantages of: a) near shift-invariance, b) good directional selectivity (for two or higher dimensions) with Gabor-like filters, c) perfect reconstruction (PR) using short linear-phase filters, c) limited redundancy, independent of number of scales, 2:1 for 1-D, 2^m:1 for m-D, d) efficient computation, only twice the DWT for 1-D, 2^m times for m-D.

The DT-CWT uses specially designed real filters which are different in the first level and remaining levels of transform and produces two parallel trees representing real and imaginary coefficients.

There are various methods for the design of DT-CWT filters which satisfies the following desired characteristics: i) approximate half-sample delay condition, ii) PR (orthogonal or biorthogonal), iii) finite support (FIR filters), iv) vanishing moments/good stopband, v) linear-phase filters [25-28].

An *odd/even filter design method* is suggested in [24, 26] for achieving good shift-invariance which requires that the samples are evenly spaced. This is achieved by eliminating down-sampling by 2 after the first level filters. The filters in one tree must provide half a sample delay different from those in the other tree, which requires odd-length filters be in one tree and even-length filters in the other tree for linear phase filters.

Unfortunately the following problems are encountered with the odd/even filter design approach: The sub-sampling structure is not very symmetrical, the two trees have slightly different frequency responses and the filter sets are not biorthogonal [25, 26].

Because of the above mentioned disadvantages of odd/even filters**,** the *Q-shift filter design method* is used in this paper. Designing the DT-CWT filters with Q-shift strategy is based on finding a good even-length low-pass filter with a delay of 1/4 sample which also satisfies the orthonormal perfect reconstruction condition. The Q-shift and shift-invariant filter design strategies are given in detail in [26, 27]. The decomposition tree structure of Q-shift filter based DT-CWT implementation is given in Figure 1.

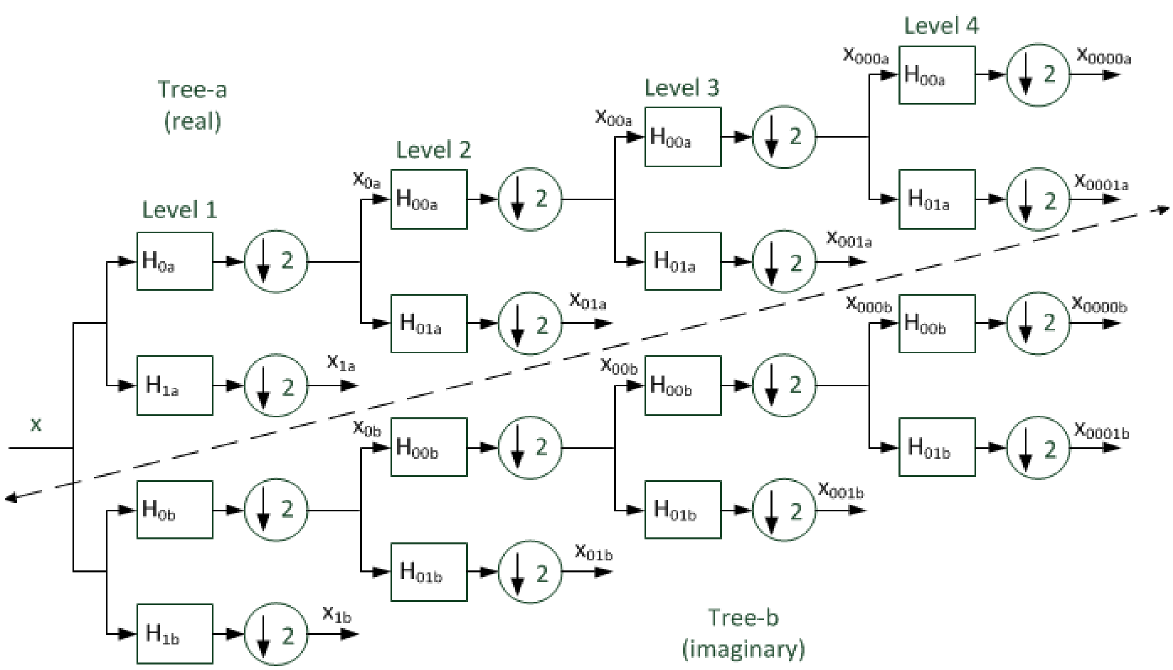


Figure 1. The Q-shift DT-CWT implementation representing real and imaginary parts of complex coefficients from Tree-a and Tree-b.

In this paper, the DT-CWT is based the Farras dual filters [29] which are used for the first level and the Kingsbury Q-shift filters [25-26] which are used for the remaining 3 levels of decomposition. The filter coefficients are given in Table 1 and Table 2 respectively. The DT-CWT filters and software can be obtained in [29].

Table 1. First level DT-CWT filter (Farras) coefficients.

Tree-a		Tree-b	
H_{0a}	H_{1a}	H_{0b}	H_{1b}
0	0	0.01122679	0
-0.08838834	-0.01122679	0.01122679	0
0.08838834	0.01122679	-0.08838834	-0.08838834
0.69587998	0.08838834	0.08838834	-0.08838834
0.69587998	0.08838834	0.69587998	0.69587998
0.08838834	-0.69587998	0.69587998	-0.69587998
-0.08838834	0.69587998	0.08838834	0.08838834
0.01122679	-0.08838834	-0.08838834	0.08838834
0.01122679	-0.08838834	0	0.01122679
0	0	0	-0.01122679

Table 2. Remaining levels DT-CWT filter (Q-Shift) coefficients.

Tree-a		Tree-b	
H_{00a}	H_{01a}	H_{00b}	H_{01b}
0.03516384	0	0	-0.03516384
0	0	0	0
-0.08832942	-0.11430184	-0.11430184	0.08832942
0.23389032	0	0	0.23389032
0.76027237	0.58751830	0.58751830	-0.76027237
0.58751830	-0.76027237	0.76027237	0.58751830
0	0.23389032	0.23389032	0
-0.11430184	0.08832942	-0.08832942	-0.11430184
0	0	0	0
0	-0.03516384	0.03516384	0

3. Proposed Subband Speech Enhancement Method

The minimum mean square error (MMSE) methods are based on the Gaussian statistical model and *a priori* SNR estimation. The a priori SNR is reported to be significant (rather than noise variance) in the noise removal [4-5]. Furthermore, it is reported [6] that, the MMSE based estimators cause no musical background noise. Since the a priori probability of speech and noise are not known in practice, each probability distribution can be measured or alternatively a suitable statistical model (i.e. Gaussian statistical model) can be assumed in order to derive the minimum mean square error short time spectral amplitude (MMSE-STSA) estimator [7].

It is reported that, a distortion measure which is based on the mean-square error of the log-spectra is more suitable for speech processing [8]. Such a distortion measure is therefore widely used for speech processing. So, it is important to analyze the STSA estimator which minimizes the mean-square-error of the log-spectra in enhancement of noisy speech.

A useful approach is to derive minimum mean square error log-spectral amplitude (MMSE-LSA) estimator which takes into account the probability of speech absence in the noisy speech signal. This modified estimator which results in improved speech enhancement performance is called as multiplicatively modified log-spectral amplitude (MM-LSA) estimator. Such an estimator can be derived based on Gaussian statistical model and assuming that the speech does not always exist in the input speech signal [8-9]. The derivation of MM-LSA estimator used in this paper is given in [9] in detail. The overall block diagram of the proposed speech enhancement method is given in Figure 2.

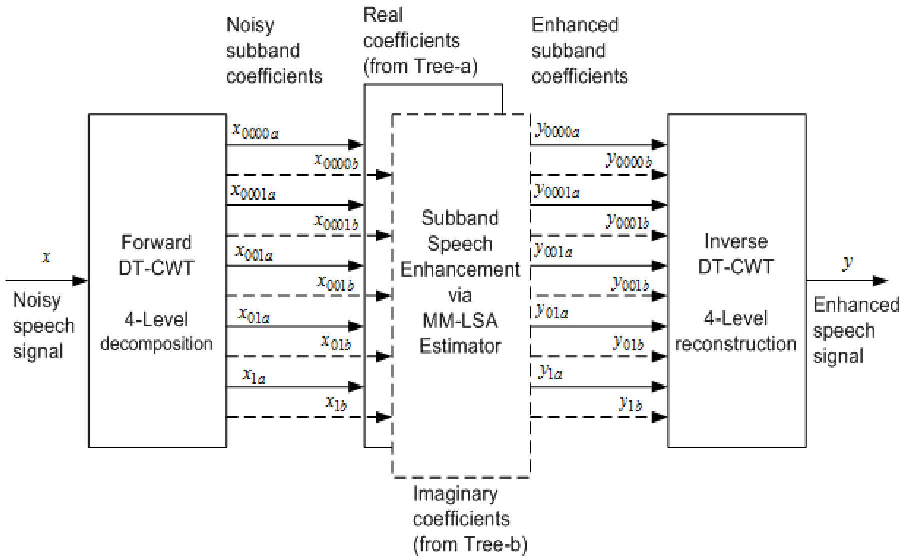


Figure 2. Overall block diagram of the proposed speech enhancement system based on the DT-CWT and MM-LSA estimator.

4. Performance Evaluation

The signal to noise ratio (SNR), the segmental signal to noise ratio (SegSNR) and the quality-mean opinion score for the estimation of subjective mean opinion score (PESQ-MOS) tests are employed for objective performance evaluation of the proposed speech enhancement method [32-33].

The experimental results are used for subjective performance evaluation by visual inspection on the output speech signals waveforms and spectrograms. In the evaluations, a 3 seconds duration English sentence “Hurdle the pit with aid of a long pole.” pronounced by a man speaker and the noise signals (GWN, car, airport and train) sampled at 8000 sample/second taken from the Noizeus database are used. The original speech signal is corrupted by adding the noises at different levels (-5, 0, 5, 10 dB SNRs). The noisy speech signal is enhanced by using proposed speech enhancement method and other wavelet transform based methods. The SNR improvement, SegSNR improvement and PESQ-MOS results of the proposed speech enhancement method (based on the DT-CWT) and the results based on the other wavelet transforms (SWT, DWT, WPT and LWT) are given in Figure 3 through Figure 5 respectively.

The MM-LSA estimator is used successfully for the enhancement of noisy subband coefficients with all of the wavelet transform based speech enhancement algorithms. The enhanced speech signal is obtained via the inverse DT-CWT. The SNR results of noisy and enhanced speech signals are evaluated using Equation (1) given below.

$$SNR_{(dB)} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1}[x(n)]^2}{\sum_{n=0}^{N-1}[\hat{x}(n)-x(n)]^2}\right) \tag{1}$$

Where $x(n)$ denotes original speech signal, $\hat{x}(n)$ denotes enhanced speech signal, and N denotes number of samples in original speech signal. The SNR is a common measure of speech quality; however it is not well correlated with human auditory system. The SNR improvement at a given input SNR (-5, 0, 5 or 10 dB) can be simply obtained by subtracting the noisy SNR value from the enhanced SNR value at that input SNR. The frame based segmental SNR is a better measure of speech quality which is obtained by averaging frame level estimates as given in Equation (2).

$$SegSNR_{(dB)} = \frac{1}{M}\sum_{m=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=N*m}^{N*m+N-1}[x(n)]^2}{\sum_{n=N*m}^{N*m+N-1}[\hat{x}(n)-x(n)]^2}\right] \qquad (2)$$

Here $x(n)$ denotes original speech signal, $\hat{x}(n)$ denotes enhanced speech signal, M denotes number of frames and N denotes number of samples in each short time frame. Since the SNR values lower than (-10 dB) and higher than (+35 dB) do not truly reflect the human auditory system, a limit [-10,+35] is chosen. The SNR values lower than (-10 dB) are set to (-10 dB) and similarly the ones higher than (+35 dB) are set to (+35 dB) in the evaluations.

The SegSNR improvement at a specific input SNR (-5, 0, 5 or 10 dB) is obtained by subtracting the noisy SegSNR value from the enhanced SegSNR value at that input SNR.

The ITU standard PESQ (ITU-T Recommendation P.862, 2001) [33] is an advanced version of the perceptual speech quality measure (PSQM which predicts subjective MOS for a wide range of speech distortions in transmission systems.

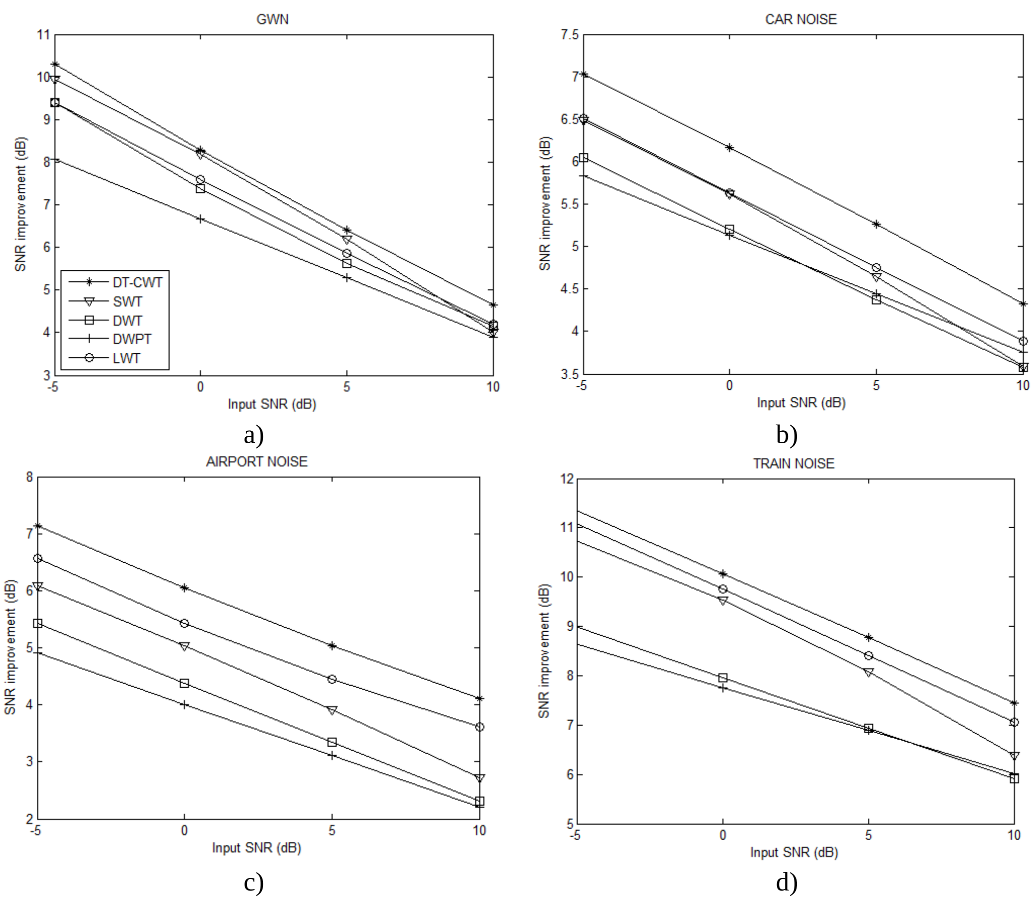


Figure 3. SNR improvement (dB) versus input SNR (dB) for wavelet-based speech enhancement algorithms (DT-CWT (proposed), SWT, DWT, WPT and LWT), for the noise types, a) GWN, b) Car noise, c) Airport noise and d) Train noise.

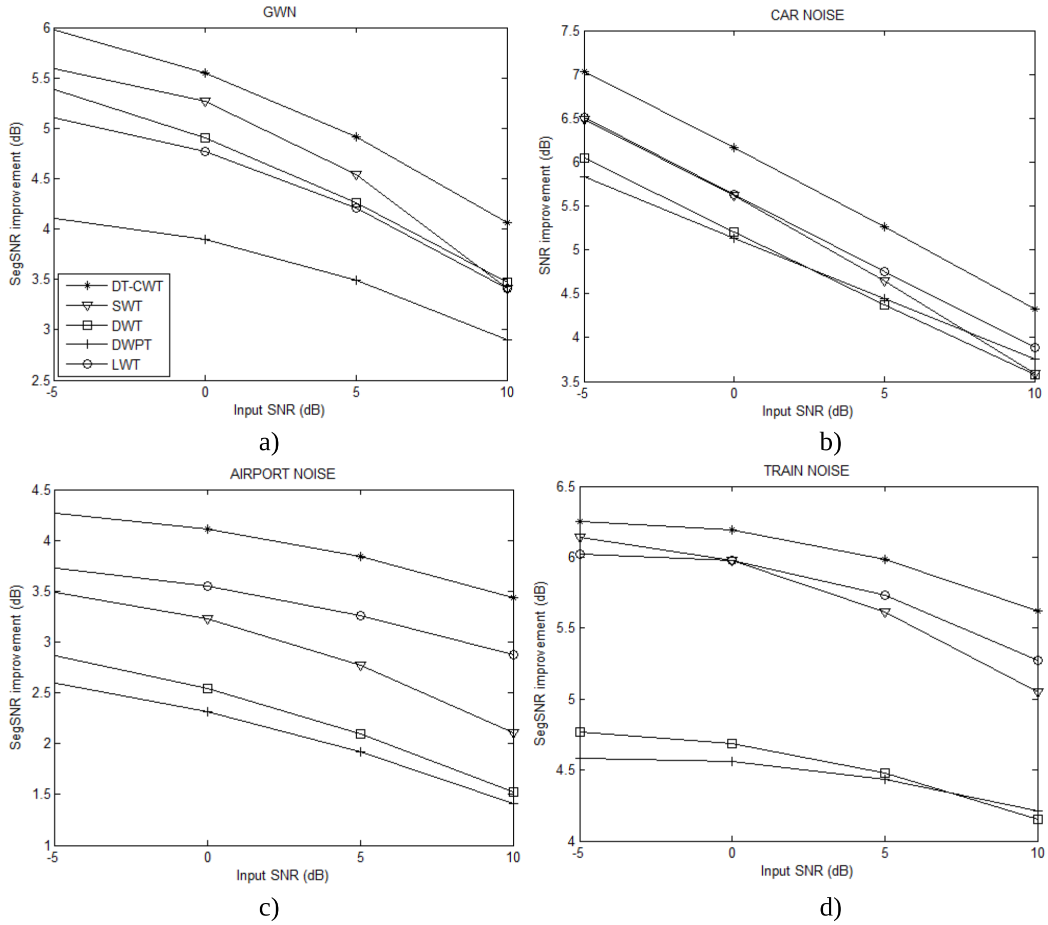
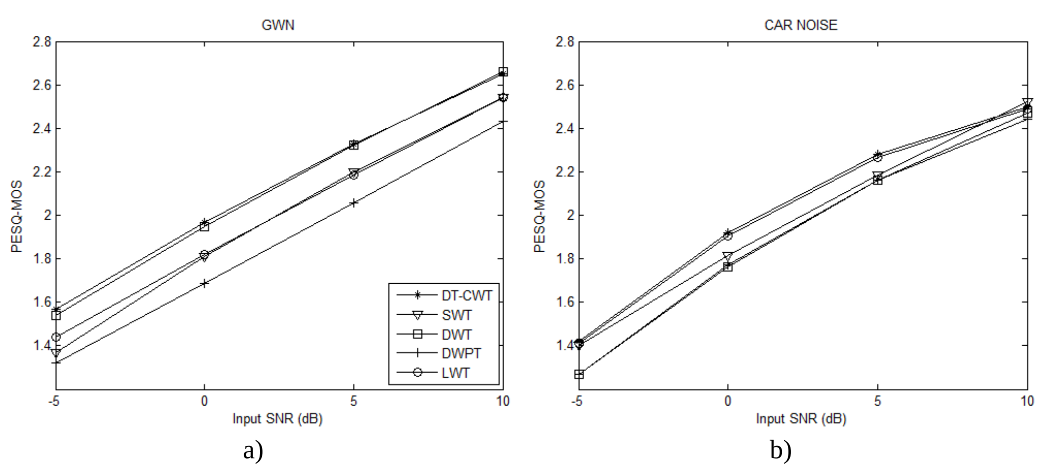


Figure 4. Segmental SNR improvement (dB) versus input SNR (dB) for wavelet-based speech enhancement algorithms (DT-CWT (proposed), SWT, DWT, WPT and LWT), for the noise types, a) GWN, b) Car noise, c) Airport noise and d) Train noise.



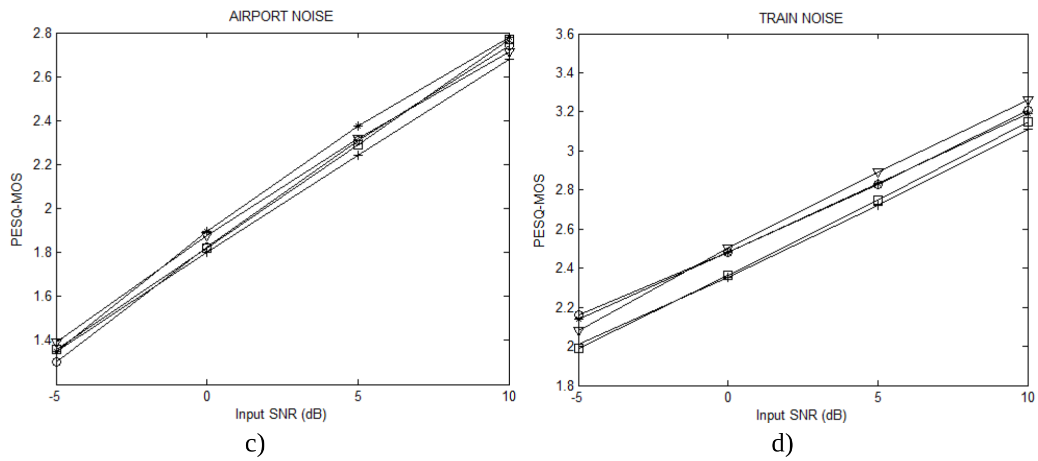


Figure 5. PESQ-MOS versus input SNR (dB) for wavelet-based speech enhancement algorithms (DT-CWT (proposed), SWT, DWT, WPT and LWT), for the noise types, a) GWN, b) Car noise, c) Airport noise and d) Train noise.

5. Experimental Results

This part presents the original, noisy and enhanced speech signal waveforms and spectrograms of the proposed speech enhancement method. The original speech signal is corrupted by various noise types (GWN, car noise, airport noise and train noise) at 0 dB SNR. The noisy speech signal is 4-level decomposed into subband coefficients (real and imaginary). The MM-LSA estimator is employed for the enhancement of the noisy subband coefficients. The enhanced speech signal is obtained via the inverse DT-CWT as given in Figure 2. The original speech signal waveform and its spectrogram are given in Figure 6. The noisy speech signals and their spectrograms, at 0 (dB) SNR, are given in Figure 7 and Figure 8 respectively. Figure 9 and Figure 10 demonstrate the enhanced speech signals and corresponding spectrograms respectively.

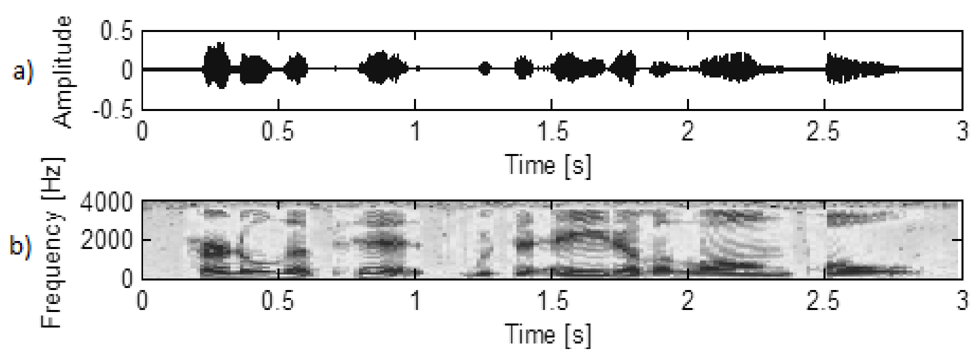


Figure 6. a) Original speech signal waveform, and b) Its spectrogram.

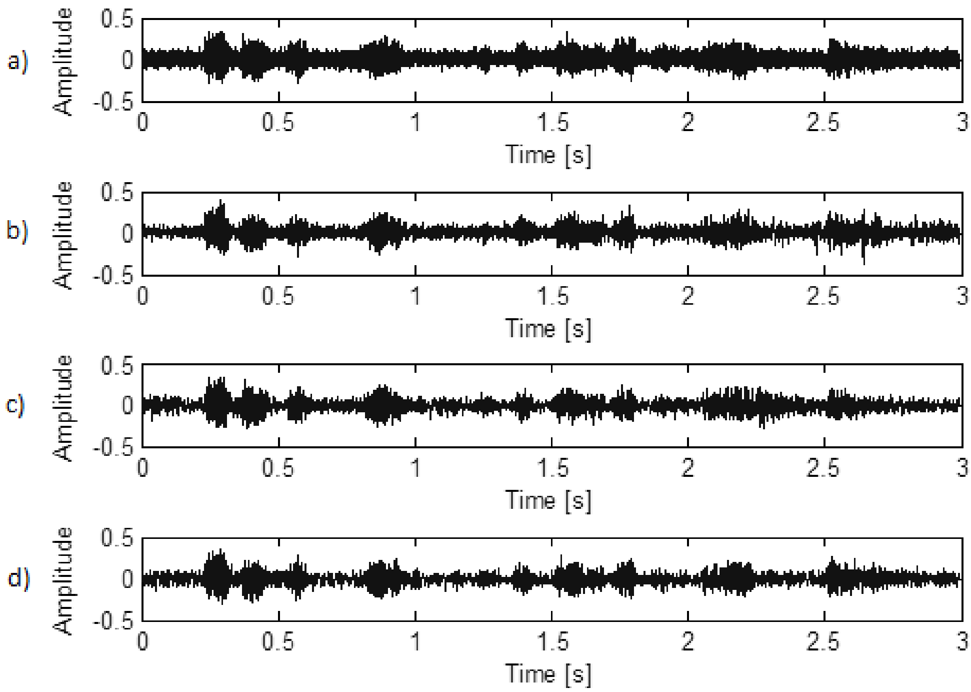


Figure 7. Noisy speech signal waveforms corrupted by various noise types: a) GWN, b) car noise, c) airport noise, and d) train noise, at 0 (dB) SNR.

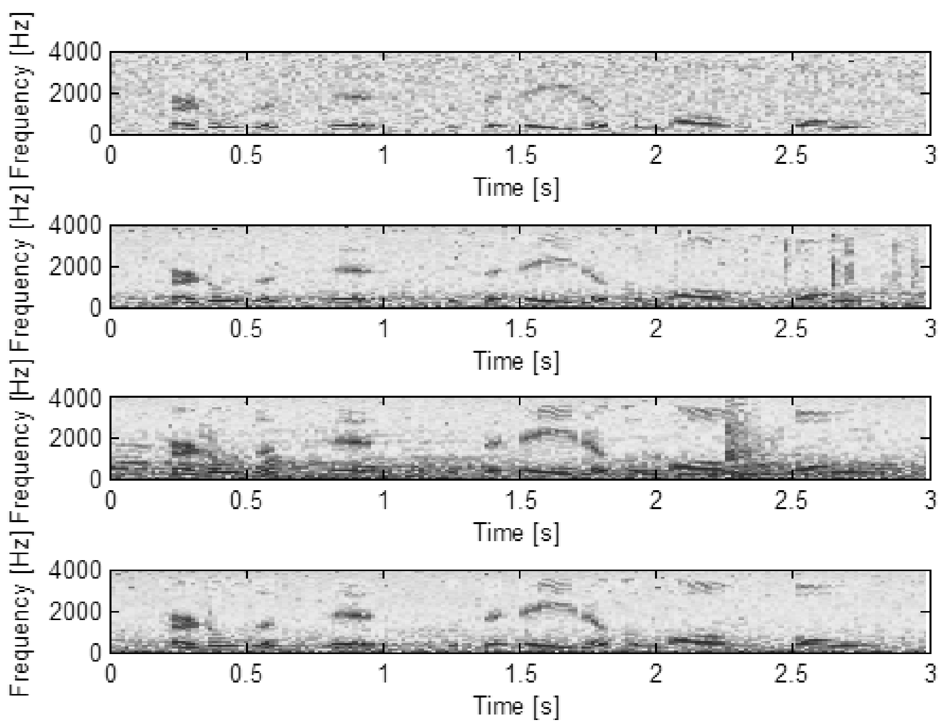


Figure 8. Noisy speech signal spectrograms corrupted by various noise types: a) GWN, b) car noise, c) airport noise, and d) train noise, at 0 (dB) SNR.

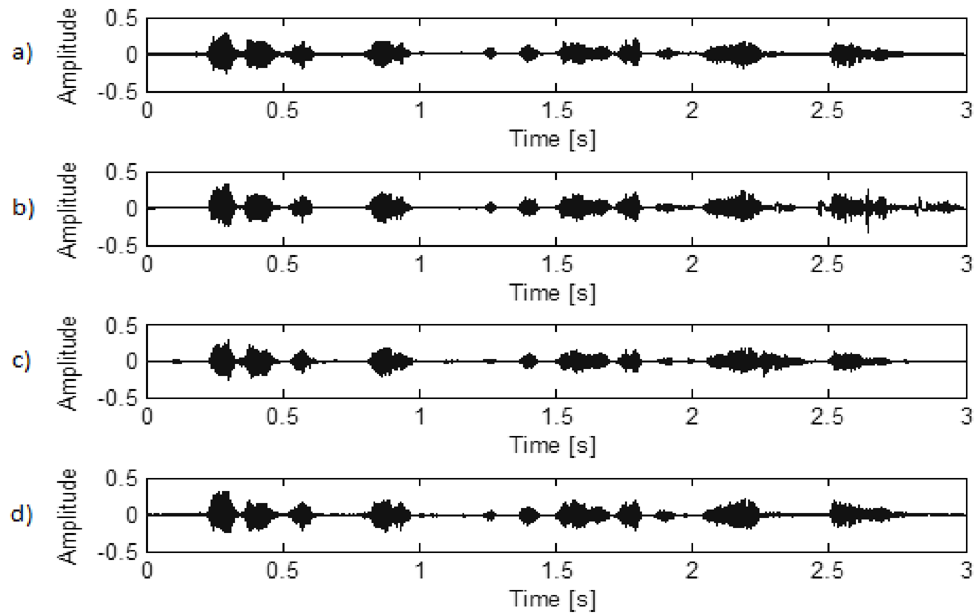


Figure 9. Enhanced speech signal waveforms by using the proposed speech enhancement method. The original signal is corrupted by a) GWN, b) car noise, c) airport noise, and d) train noise, at 0 (dB) SNR before enhancements.

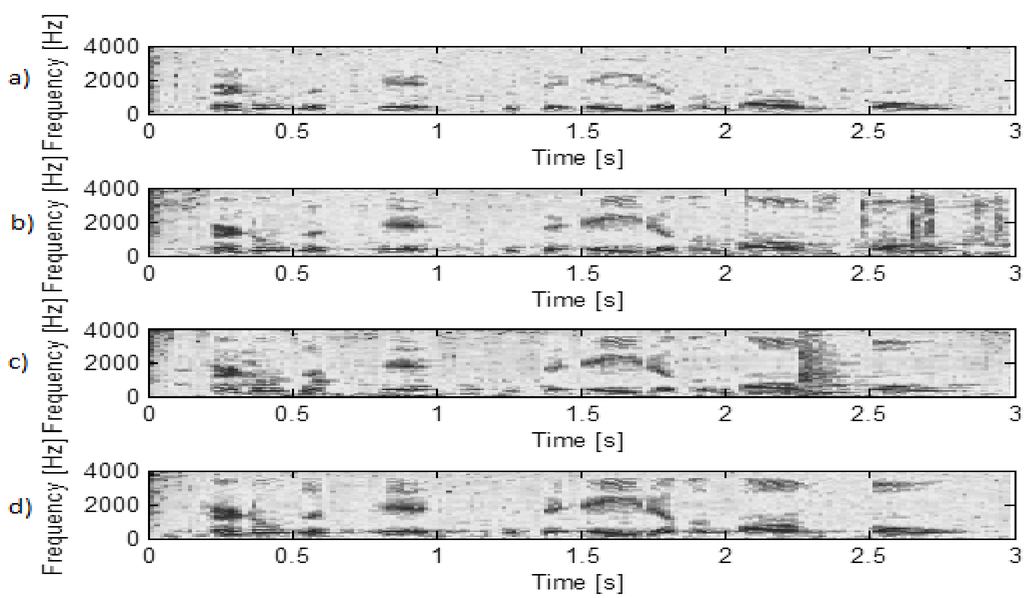


Figure 10. Enhanced speech signal spectrograms by using the proposed speech enhancement method. The original signal is corrupted by a) GWN, b) car noise, c) airport noise, and d) train noise, at 0 (dB) SNR before enhancements.

6. Results and Conclusions

In this paper, a new speech enhancement method based on DT-CWT is proposed. The noisy speech signal is decomposed via DT-CWT and the noisy subband coefficients (real and imaginary) are denoised using MM-LSA estimator. The enhanced speech signal is obtained via inverse DT-CWT as shown in Figure 2. In order to evaluate the performance of the proposed speech enhancement method the objective tests (SNR, segmental SNR and PESQ-MOS) are conducted. The results of the proposed method are compared with those of the other popular wavelet-based methods (DWT, WPT, SWT, and LWT). The proposed (DT-CWT based) speech enhancement method outperforms the other wavelet-based methods in almost all the tests except for Figure 5 d) where the SWT based method provides slightly better result. The SWT based method also provides successful results since it is shift-invariant, however it has high computational complexity. The objective test results and subjective (visual) results show that the DT-CWT is an efficient tool for speech enhancement with limited computational complexity. We attribute the success of the proposed method to the near shift-invariant nature of the DT-CWT implementation. Moreover, we think that it can be an efficient tool for the other 1-D signal processing applications.

References

[1] Martin, R., (2003), Statistical methods for the enhancement of noisy speech, *International Workshop on Acoustic Echo and Noise Control* (IWAENC 2003), 1-6, Kyoto, Japan, September 2003.

[2] Boll, S. F., (1979), Suppression of acoustic noise using spectral subtraction, *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 27(2), 113-120.

[3] Berouti, M., (1979), Schwartz, R., Makhoul, J., Enhancement of speech corrupted by acoustic noise, *ICASSP*, 208-211.

[4] McAulay, R. J. and Malpass, M. L., (1980), Speech enhancement using a soft-decision noise suppression filter, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(2), 137-145.

[5] Cappé, O., (1994), Elimination of musical noise phenomenon with the Ephraim and Malah noise suppressor, IEEE *Transactions on Speech and Audio Processing*, 2, 345-349.

[6] Scalart, P. & Filho, J. V., (1996), Speech enhancement based on a priori signal to noise estimation, *IEEE ICASSP'96,* 2, 629-632, May 7-10, 1996.

[7] Ephraim, Y. & Malah, D., (1984), Speech enhancement using a minimum mean square error short time spectral amplitude estimator, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-32(6), 1109-1121,

[8] Ephraim, Y. & Malah, D., (1985), Speech enhancement using a minimum mean square error log-spectral amplitude estimator, *IEEE Transactions on Acoustics, Speech, and Signal Processing (ASSP)*, 32(2), 449-445.

[9] Malah, D., Cox, R.V., Accardi, (1999), A. J., Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments, *Proceedings of IEEE ICASSP*, 789-792.

[10] Cohen, A. & Kovacevic, J., (1996), Wavelets: The mathematical background, *Proceedings of the IEEE*, 84(4), 514-522.

[11] Vetterli, M. & Herley, C., (1992), Wavelets and filter banks: Theory and design, *IEEE Transactions on Signal Processing*, 40(9), 2207-2232.

[12] Polikar, R., (1999), The story of wavelets, *Proceedings of IMACS IEEE CSCC'99*, 5481-5486.

[13] Coifman R. R. & Wickerhauser, M. V., (1992), Entropy-based algorithms for best basis selection, *IEEE Transactions on Information Theory*, 38(2), 713-718.

[14] Coifman, R. R., Meyer, Y. & Wickerhauser, V., (1992), Adapted wave form analysis, wavelet-packets and applications, *Proceedings of the Second International Conference on Industrial and Applied Mathematics*, Philadelphia, PA, USA, 41-50.

[15] Rajpoot, N. M., Wilson, F. G., Meyer R. G. & Coifman, R. R., (2003), Adaptive wavelet packet basis selection for zero tree image coding, *IEEE Transactions on Image Processing*, 12(12), 1460-1472.

[16] Taşmaz, H. & Erçelebi, E., (2008), Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments, *Digital Signal Processing*, 18(5), 797-812.

[17] Cohen, I., (2001), Enhancement of speech using bark scaled wavelet package decomposition, *EUROSPEECH-2001*, 1933-1936, Denmark, September 2001.

[18] Sweldens, W., (1995), The lifting scheme: A new philosophy in biorthogonal wavelet constructions, *Wavelet Application in Signal and Image Processing III*, 2569, 68-79.

[19] Sweldens, W., (1996), The lifting scheme: A custom design construction of biorthogonal wavelets, *Applied and Computational Harmonic Analysis*, 3(2), 186-200.

[20] Taşmaz H. & Erçelebi, E., (2010), Image enhancement via space-adaptive lifting scheme exploiting subband dependency, *Digital Signal Processing*, 20(6), 1645-1655.

[21] Mahidar P, Rajeswara, Prasad, N. V. G. & Sastry, V. L. N., (2011), A novel noisy image fusion technique using dual-tree complex wavelet transform, *International Journal of Engineering Science and Technology (IJEST)*, 3(6), 4523-4528.

[22] James, E. F., (2005), The redundant discrete wavelet transform and additive noise, IEEE *Signal Processing Letters*, 12, 629-632.

[23] Shensa, Mark J., (1992), The discrete wavelet transform: Wedding the A Trous and Mallat algorithms, *IEEE Transactions on Signal Processing*, 40(10), 2464-2482.

[24] Kingsbury, N. G., (1999), Image processing with complex wavelets, *Philosophical Transactions of the Royal Society of London*, 357, 2543-2560, September 1999.

[25] Kingsbury, N. G., (1999), A dual-tree complex wavelet transform with improved orthogonality and symmetry properties, *IEEE Signal Processing Magazine*, 22(6), 123-151.

[26] Kingsbury, N. G., (2011), Complex wavelets for shift invariant analysis and filtering of signals, *Journal of Applied and Computational Harmonic Analysis*, 10(3), 234-253.

[27] Kingsbury, N. G., (2003), Design of Q-shift complex wavelets for image processing using frequency domain energy minimization, *IEEE International Conference on Image Processing (ICIP'03)*, 1013-1016, September 2003.

[28] Selesnick, I. W., Baraniuk, R.G. & Kingsbury, N. G., (2005), The dual-tree complex wavelet transform, *IEEE Signal Processing Magazine*, 22(6), 123-151.

[29] http://taco.poly.edu/WaveletSoftware/

[30] Chen, G. Y. & Xie, W. F., (2007), Pattern recognition with SVM and dual-tree complex wavelets, *Image and Vision Computing*, 25(6), 960-966.

[31] Demirel, H. & Anbarjafari, G., (2010), Satellite image resolution enhancement using complex wavelet transform, *IEEE Geoscience and Remote Sensing Letters*, 7(1), 123-126.

[32] Hansen, J. H. L. & Pellom, B.L., (1998), An effective quality evaluation protocol for speech enhancement algorithms, ICSLP-98: *International Conference on Spoken Language Processing, 7, Sydney, Australia*, 2819-2822.

[33] Rix, A. W. J., Beerends, G., Hollier M. P. & Hekstra, A. P., (2000), PESQ-the new ITU standard for end-to-end speech quality assessment, *AES 109$_{th}$ convention*, Los Angeles, 1-18, September 22-25, 2000.