Improved multi-band spectral subtraction method for speech enhancement

Proceedings of the 6th IASTED International Conference
SIGNAL AND IMAGE PROCESSING
August 23-25, 2004, Honolulu, Hawaii, USA

IMPROVED MULTI-BAND SPECTRAL SUBTRACTION METHOD FOR SPEECH ENHANCEMENT

Yasser Ghanbari*, Mohammad Reza Karami-Mollaei*, Behnam Amelifard**
*Department of Electrical Engineering, University of Mazandaran, Babol, Iran
** Department of EE Systems, University of Southern California, Los Angeles, CA 90089
Ghanbari@ee.sharif.edu, mkaramifr@yahoo.fr, amelifar@usc.edu

ABSTRACT
In this paper, we propose a new approach to improve the performance of speech enhancement technique based on multi-band spectral subtraction for white Gaussian noise. First, the original power spectral subtraction and multi-band spectral subtraction methods are surveyed and implemented. Next, the generalization is applied on multi-band spectral subtraction. Finally, the flattened noise spectrum with the value of the average of spectrum components is used as the noise estimation and then a power factor is optimized in order to increase the enhancement performance. The results show a meaningful improvement. Comparisons of three methods are reported.

KEY WORDS
Speech processing, speech enhancement, spectral subtraction

1. Introduction

In many speech communication systems, background noise causes the quality of speech to degrade. In most of speech processing applications such as mobile communications, speech recognition and hearing aids [7], removing the background noise in a noisy environment is inevitable. So, speech enhancement as a necessity for related applications has been widely studied in recent years.

Although there are several techniques such as Wiener filtering [6], hidden Markov modeling [7], wavelet-based [8,11], adaptive filtering [9] and signal subspace methods [10], spectral subtraction is still a useful method [2,3,5]. In this method, we have to estimate the noise spectrum and subtract it from noisy speech spectrum. In this method, three following conditions are assumed: (1) the noise is additive (2) speech signal and noise are uncorrelated (3) one channel is available. Usually noise is estimated from non-speech frames. However by such a method, a new noise named musical noise is generated. To reduce the musical noise, Brouti [5] made a generalization on subtraction of estimated noise spectrum by over-subtraction and spectral-floor factors. Kamath & Loizu [2] have used this method in multi frequency bands and tried to reduce the noise by multi-band spectral subtraction (MBSS) method.

In this paper, we try to reduce the estimation error of noise spectrum for enhancement of corrupted speech with WGN by improvement of a multi-band spectral subtraction (MBSS) technique [2,4]. For improvement, we survey the variation effect of a generalized power factor and choose the best. Then we propose a method to reduce the difference between estimated noise spectrum and noise spectrum. At the end of the paper, we show that global SNR is considerably increased and the musical noise reduced.

2. Original Power Spectral Subtraction (PSS)

Because of the assumption that noise is additive, we can model the corrupted speech signal by following equation:

$$y(n) = s(n) + Gd(n) \tag{1}$$

Where s(n) is clean speech signal, d(n) is noise and G is the term for SNR control. We assume that the noise signal is uncorrelated, so:

$$r_d(\eta) = D_0\delta(\eta) \tag{2}$$

Where r_d is the autocorrelation function of noise signal and D_0 is a constant [1]. Because $d(n)$ is uncorrelated process, we can show:

$$\Gamma_y(\omega) = \Gamma_s(\omega) + \Gamma_d(\omega) \tag{3}$$

Where Γ is the power spectral density (PSD). So, if we can estimate $\Gamma_d(\omega)$, we will be able to estimate $\Gamma_s(\omega)$:

$$\hat{\Gamma}_s(\omega) = \Gamma_y(\omega) - \hat{\Gamma}_d(\omega) \tag{4}$$

Note that noise is estimated from silence frames. PSD is related to discrete-time fourier transform (DTFT) [1] as:

$$\Gamma_y(\omega) = \frac{Y(\omega).Y^*(\omega)}{N^2} = \frac{|Y(\omega)|^2}{N^2} \quad (5)$$

So we can conclude from (4) & (5):

$$\left|\hat{S}(\omega)\right|^2 = |Y(\omega)|^2 - \left|\hat{D}(\omega)\right|^2 \quad (6)$$

In order to estimate the speech signal frame, the other necessary factor is $\hat{\varphi}_s(\omega)$ which is the estimated phase spectrum of speech frame. Boll has shown [3] that in practical applications, it is sufficient to use the noisy phase spectrum as an estimation of clean speech phase spectrum:

$$\hat{\varphi}_s(\omega) \approx \varphi_y(\omega) \quad (7)$$

Therefore from equations (6) & (7), we can obtain the estimated speech frame:

$$\hat{S}(\omega) = \left|\hat{S}(\omega)\right| e^{j\hat{\varphi}_s(\omega)} = \left[|Y(\omega)|^2 - \left|\hat{D}(\omega)\right|^2\right]^{\frac{1}{2}} . e^{j\varphi_y(\omega)} \quad (8)$$

And a generalization for (8) is [1]:

$$\hat{s}(n) = IDTFT\left\{\left[|Y(\omega)|^\gamma - \left|\hat{D}(\omega)\right|^\gamma\right]^{\frac{1}{\gamma}} . e^{j\varphi_y(\omega)}\right\} \quad (9)$$

Where the power exponent "γ" can be chosen by optimization.

The problem with this method is negative spectral components created after some subtractions. This problem is due to the error in estimation of the noise spectrum. So, these negative values must be modified. This problem can be eliminated by two methods [1]. One is half-wave rectification:

$$\left|\hat{S}(\omega)\right|^2 = \begin{cases} \left|\hat{S}(\omega)\right|^2 & if \left|\hat{S}(\omega)\right|^2 > 0 \\ 0 & else \end{cases} \quad (10)$$

and the other is full-wave rectification:

$$\left|\hat{S}(\omega)\right|^2 = abs\left(\left|\hat{S}(\omega)\right|^2\right) \quad (11)$$

These two modification methods introduce a new noise named "musical" noise which is the major limitation of spectral subtraction methods.

3. Multi-band Spectral Subtraction (MBSS)

Recent researches have shown that signal-to-noise ratio varies across the speech spectrum [4]. The fact that noise has non-uniform effect on different vowels and consonants [4] is one of justifications to work separately on multi-band spectral subtraction (MBSS). Figure 1 shows this fact [4].

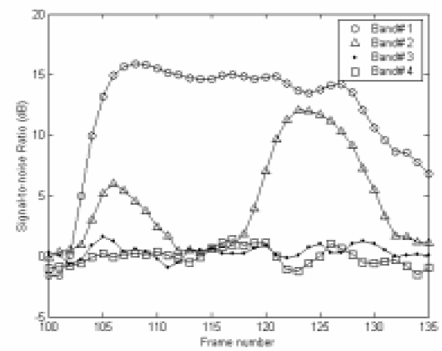

Figure 1: Segmental SNR of four (linearly-spaced) frequency bands of speech corrupted by WGN at 5dB SNR. (after [4])

As can be seen from Figure 1, in most of frames, SNR of frequency band 1 is greater than SNR of other frequency bands, also SNR of frequency band 2 is greater than SNR of frequency bands 3 and 4. So, Kamath & Loizou [2] used the spectral subtraction method proposed by Brouti [5] in several frequency bands for reduction of musical noise. This technique is expressed as:

$$\left|\hat{S}(\omega)\right|^2 = |Y(\omega)|^2 - \alpha\left|\hat{D}(\omega)\right|^2 \qquad \alpha \geq 1 \quad (12)$$

Where α is over subtraction factor and is a function of segmental SNR_i [5]:

$$\alpha = \begin{cases} \alpha_0 + \frac{3}{4} & SNR_i \leq -5db \\ \alpha_0 - \frac{3}{20} SNR_i & -5db \leq SNR_i \leq 20db \\ \alpha_0 - 3 & SNR_i \geq 20db \end{cases} \quad (13)$$

The segmental SNR_i of the i th frequency band is calculated as:

$$SNR_i = 10 \log \frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} \left|\hat{D}_i(k)\right|^2} \quad (14)$$

Where b_i and e_i are the beginning and ending frequency bins of the i th frequency band.

Brouti [5] proposed for α_0 to be between 3 & 6. Experimental results have shown that $\alpha_0 = 4$ is appropriate [2].

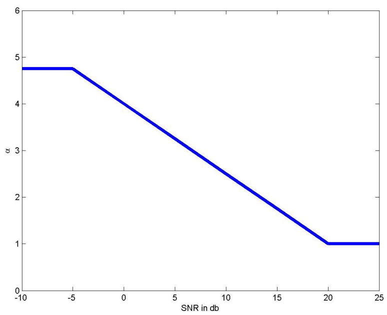

Figure 2: Over-subtraction factor α as a function of SNR with $\alpha_0 = 4$.

For negative spectral components, Brouti proposed using spectral floor factor [5] as:

$$\left|\hat{S}(\omega)\right|^2 = \begin{cases} \left|\hat{S}(\omega)\right|^2 & \text{if } \left|\hat{S}(\omega)\right|^2 > \beta\left|\hat{D}(\omega)\right|^2 \\ \beta\left|\hat{D}(\omega)\right|^2 & \text{else} \end{cases} \tag{15}$$

Where β is spectral floor factor and $\beta << 1$. From experiments, Kamath & Loizou [2] proposes $\beta = 0.002$. Therefore, the block diagram of MBSS method can be shown as the Figure 4.

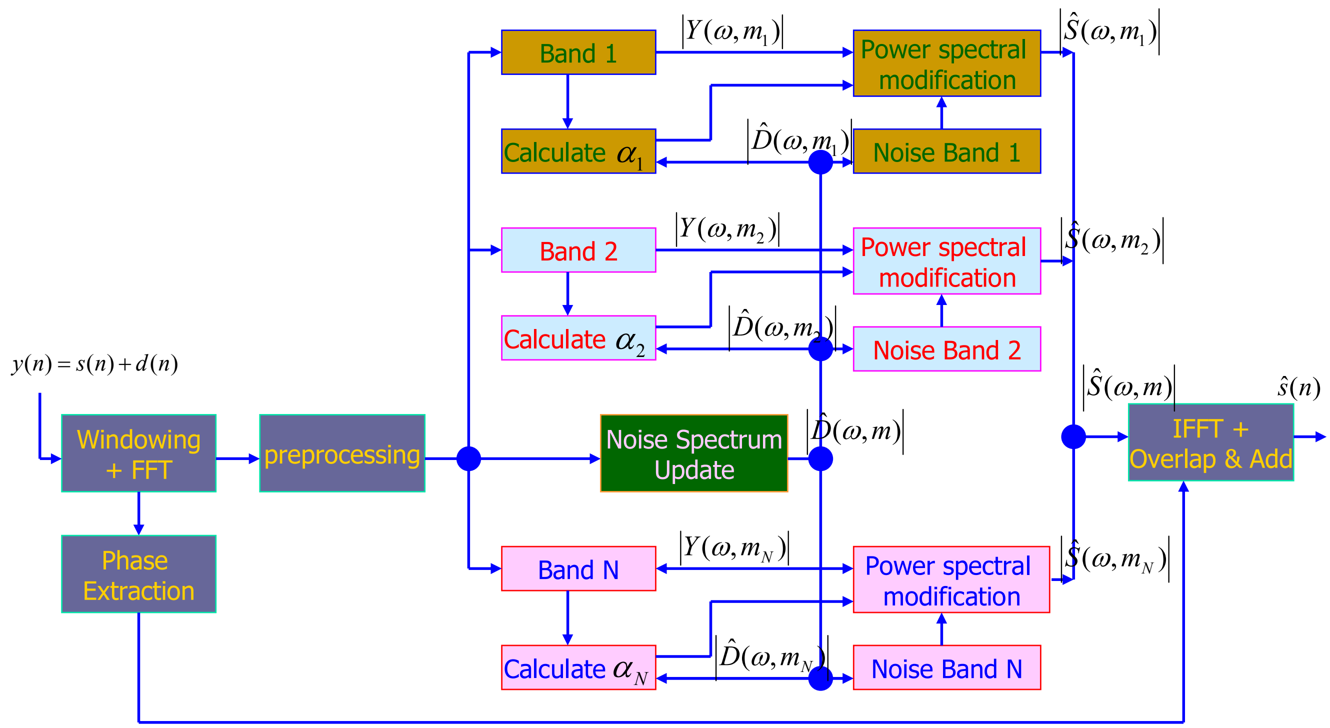

Figure 3: Block diagram for MBSS method (after [4])

4. Improvement of Multi-band Spectral Subtraction for WGN

As we know, the real-world noise is highly random in nature. So we try to improve the MBSS method [2] for reduction of WGN. Three factors that greatly affect the enhancement performance in spectral subtraction methods are: (1) correctness of noise estimation (2) processing of negative spectral components (3) the power factor "γ". So, we can improve the MBSS method by variation of power factor "γ" and choice of the best one.

To reduce the noise estimation error and minimum the difference between noise spectrum and estimated noise spectrum, we can flatten the estimated noise spectrum and use the averaged value of spectral components of estimated noise for all bands.

5. Experimental Results

In this section, the enhancement performances of three methods are compared. For a good comparison, we have shown the temporal results and the spectrograms of clean, noisy and denoised speech. Clean speech sentence that is processed here, has been chosen from TIMIT database.

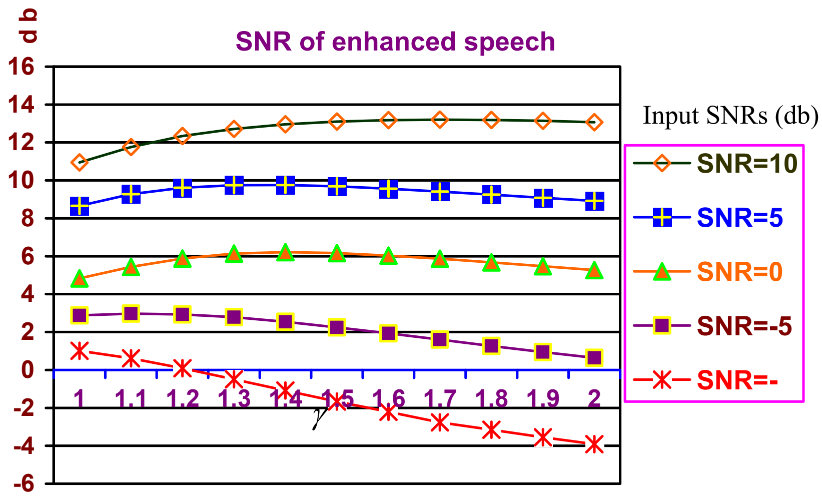

Figure 4: SNR of enhanced speech as a function of the power factor "γ".

The sentence is about 2msec with the sampling rate of 16 kHz and spoken by a male speaker. Figures 5(a), 5(b), 5(c) & 5(d) show the temporal results and Figures 6(a), 6(b), 6(c) & 6(d) show the spectrograms of clean, noisy and enhanced speech of two first methods.

For optimization of MBSS, first we varied the power factor "γ" from 1 to 2 for corrupted speech with input SNRs of -10db to 10db (with 5db steps) and the SNRs of enhanced speech is depicted in Figure 4.

The output enhanced speech was checked by hearing and observing the spectrogram and SNR improvement, then the best value for power factor "γ" was chosen $\gamma = 1.5$. Then we replaced the averaged value of estimated noise spectrum for all bands. Figures 5(e) and 6(e) show the enhanced speech.

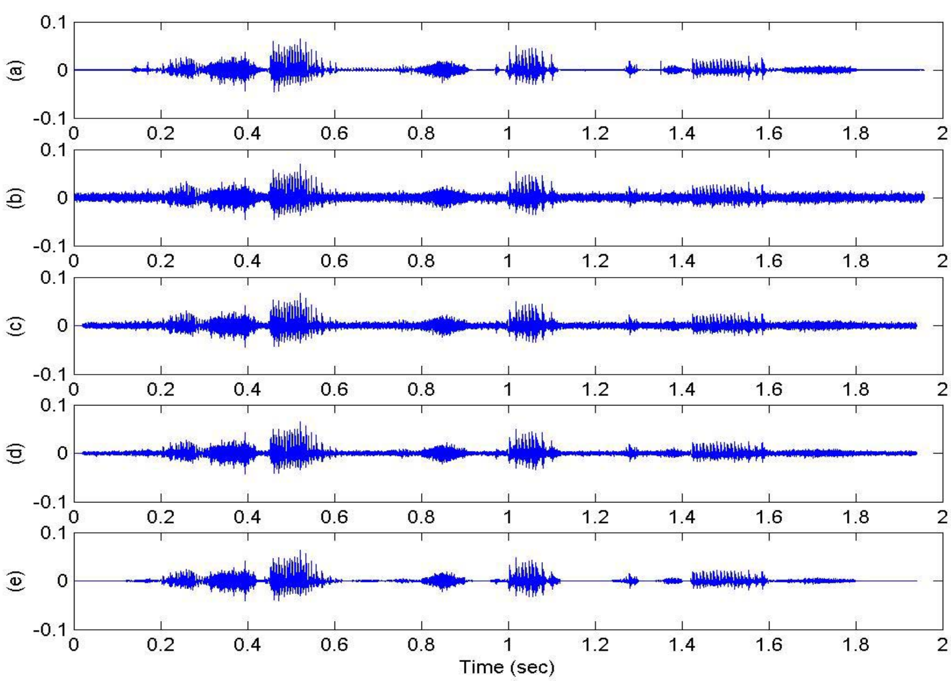

Figure 5: speech enhancement results:
(a) Clean speech
(b) Noisy speech (SNR=5db)
(c) Enhanced speech by PSS method (SNR=6.3db)
(d) Enhanced speech by MBSS method (SNR=8.9db)
(e) Enhanced speech by IMBSS method (SNR=11db)

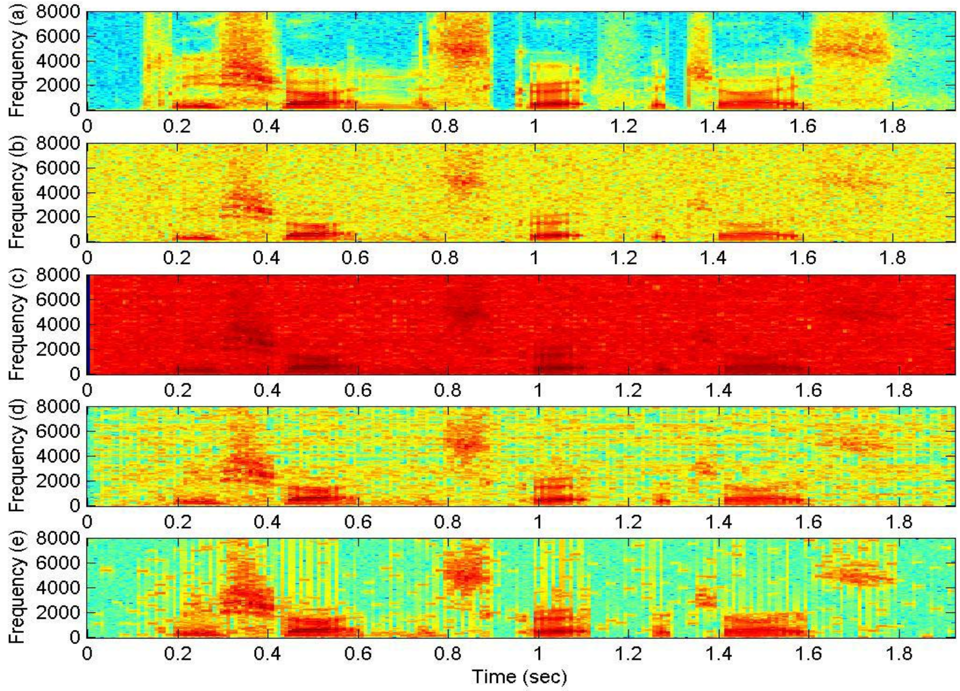

Figure 6: speech enhancement spectrograms:
(a) Clean speech
(b) Noisy speech (SNR=5db)
(c) Enhanced speech by PSS method (SNR=6.3db)
(d) Enhanced speech by MBSS method (SNR=8.9db)
(e) Enhanced speech by IMBSS method (SNR=11db)

The improved method (IMBSS) shows to be more efficient in comparison to the others. It not only develops the signal-to-noise ratio, but highly reduces the musical noise. All of these three methods were experimented for corrupted speech with SNRs of -10db to 10db and the SNRs of enhanced speech is reported in Figure 7.

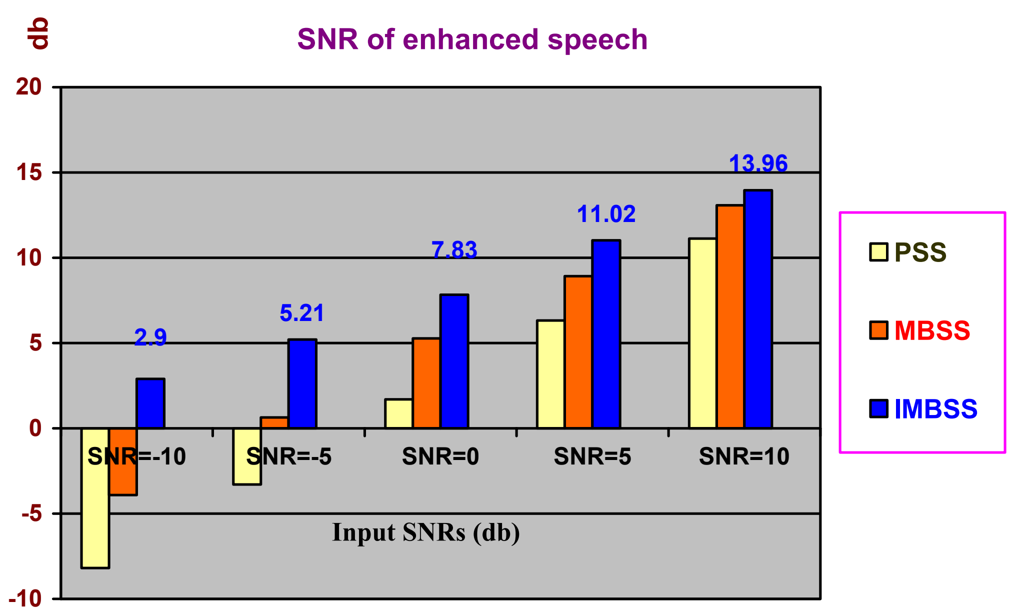

Figure 7: SNR comparison of three methods.

6. Conclusion

The IMBSS method has shown that it can considerably enhance the noisy speech corrupted by white gaussian noise and reduce the musical noise, also the enhanced speech signal have a good auditory quality. But the limitation to this method is the application for enhancement of the noisy speech corrupted by colored noise. Flattened colored noise spectrum may reduce the enhancement performance but the optimizated value for power factor "γ" is independent of the kind of noise spectrum. So this optimized value can be used for any applications.

References:

[1] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-time processing of speech signals*, (IEEE Press, 2000).

[2] S. Kamath, and P. Loizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, *Proceedings of ICASSP-2002*, Orlando, FL, May 2002.

[3] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. on Acoust. Speech & Signal Processing*, Vol. ASSP-27 , April 1979, 113-120.

[4] S. Kamath, A multi-band spectral subtraction method for speech enhancement, *MSc Thesis in Electrical Eng*., University of Texas at Dallas, December 2001.

[5] M. Berouti, R. Schwartz, and J. Makhoul, Enhancement of speech corrupted by acoustic noise, *Proc. IEEE ICASSP* , Washington DC, April 1979, 208-211.

[6] P. S. Whitehead, D. V. Anderson, M. A. Clements, Adaptive acoustic noise suppression for speech enhancement, *IEEE International Conference on Multimedia & Expo.*, July 2003.

[7] H. Sameti, H. Sheikhzadeh, Li Deng, R. L. Brennan, HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise, *IEEE Transactions on Speech and Audio Processing*, Vol. 6, No. 5, September 1998.

[8] D. L. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory*, Vol. 41, No. 3, May 1995, 613-627.

[9] K. Y. Lee, B. G. Lee, S. Ann, Adaptive filtering for speech enhancement in colored noise, *IEEE Trans. on Signal Processing Letters*, Vol. 4, October 1997, 277-279.

[10] M. Klein and P. Kabal, Signal subspace speech enhancement with perceptual post-filtering, *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, FL, May 2002, I-537-I-540.

[11] Ing Yann Soon, Soo Ngee Koh, Chai Kiat Yeo, Wavelet for Speech Denoising, *TENCON 97*, Brisbane, Australia, 1997, 479-482.