SPEECH ENHANCEMENT USING AUDITORY SPECTRAL ATTENUATION

Novlene Zoghlami [1], Zied Lachiri [1, 2], and Noureddine Ellouze [1]

[1] Ecole Nationale d'Ingénieurs de Tunis / Département Génie Electrique, unité Signal, Image et Reconnaissance de formes
ENIT, BP.37, Le Belvédère, 1002, Tunis, Tunisia
Email: novlene_zoglami@yahoo.fr, N.Ellouze@enit.rnu.tn

[2] Institut National des Sciences Appliquées et de Technologie / Département Instrumentation et Mesures
INSAT, BP 676 centre urbain cedex, Tunis, Tunisia.
Email: zied.lachiri@enit.rnu.tn

ABSTRACT

This paper presents a new speech enhancement approach for noise reduction based on non-uniform multi-band analysis. The noisy signal is divided into a number of sub-bands using a gammatone filterbank with non-linear ERB resolution, and the sub-bands signals are individually weighted according the power spectral subtraction technique and the Ephraim and Malah spectral attenuation algorithm. Subjective evaluation tests demonstrate significant improvements results over classical subtractive type algorithms, when tested with speech signal corrupted a posterior by various noises at different signal to noise ratios.

1. INTRODUCTION

With the increasing attractiveness of automatic speech processing systems, a need to develop acoustic noise suppression rules for speech signal is imposed, since these systems are often used in environment where high ambient noise levels are present, so their performance degrades considerably. This problem has already received much attention in the literature, and many algorithms are developed in order to removing the background noise while retaining speech intelligibility. Actually, noise suppression algorithms are based on short-time spectral estimation. These methods are relied with a tradeoff between a minimum level of speech distortion introduced and efficient noise suppression. The spectral subtraction rule [1] is the most popular technique able to reduce the background noise using estimation of the short-time spectral magnitude of the speech signal, obtained by subtracting the noise estimation from the noisy speech. However, this method needs to be improved, since it introduces in the enhanced speech a perceptually annoying residual noise, called musical noise, and composed of tones and random frequencies. To overcome this problem, the Ephraim and Malah subtraction rule [2] exploits the average spectral estimation of the speech signal based on a prior knowledge of the noise variance, in the goal to mask and reduce the residual noise. Others methods [3], [4] exploit the proprieties of the human auditory system, especially the auditory masking to improve the quality and intelligibility of the speech signal without introducing speech distortion. The objective of this paper is to adapt the spectral subtraction technique and the Ephraim and Malah rule to a multi-bands analysis using non-linear frequency ERB resolution filterbank, in according with the human auditory system behavior. In section 2, we present the subtractive type algorithms. Section 3 shows the proposed noise reduction approach based on auditory spectral analysis and the section 4 exposes the results and subjective evaluation tests.

2. SPECTRAL ATTENUATION TECHNIQUES

The subtractive type algorithms present a family of reference algorithms for noise reduction, operating in the frequency domain based on spectral modifications. These methods are widely used for the enhancement of speech signals, which are corrupted by additive noise with constant or slowly varying spectral characteristics. The basic idea is to manipulate the magnitude of the noisy speech spectrum using fixed and uniform spaced frequency transformation. Widely applied examples of spectral attenuation technique are power spectral subtraction and Ephraim and Malah short time spectral amplitude estimator. These methods are based on applying a spectral gain to each frequency bin of the noisy speech signal. The spectral analysis and synthesis is usually performed by a Fast Fourier transform and its inverse with overlap-add technique. Assuming that $|X(m,f)|$, $|Y(m,f)|$ and $|\hat{N}(m,f)|$ are the FFT bin of the clean speech, noisy speech and noise spectrum estimate in a frame m at each frequency f respectively, spectral subtraction rule (SS) subtracts the estimated noise from the noisy speech signal in the power spectral density domain according to the following expression:

$$\left|\hat{X}(m,f)\right|^2 = \left|Y(m,f)\right|^2 - \left|\hat{N}(m,f)\right|^2 \qquad (1)$$

Where $\left|\hat{N}(m,f)\right|^2$ and $\left|\hat{X}(m,f)\right|^2$ are the noise power estimation calculated from averaging non speech segment and the power spectrum of the enhanced speech signal.

The Ephraim and Malah subtractive rule (EM) is based on a

Bayesian estimation of the magnitude enhanced speech using a spectral weighting $G^{EM}(m,f)$ expressed according the a posterior signal to noise ratio R_{post} and the a prior signal to noise ratio R_{prio}:

$$G^{EM}(m,f)=$$

$$\left|\frac{\sqrt{\pi}}{2}\right|\sqrt{\left(\frac{1}{1+R_{post}}\right)\left(\frac{R_{priori}}{1+R_{post}}\right)}M\left[(1+R_{post})\left(\frac{R_{priori}}{1+R_{priori}}\right)\right] \quad (2)$$

(3)

The local level a posterior and the prior signal to noise ratio are expressed as:

$$R_{post}(m,f)=\frac{|Y(m,f)|^2}{|\hat{N}(m,f)|^2} \quad (3)$$

$$R_{priori}(m,f)=(1-\alpha)[R_{post}(m,f)-1]+\alpha\frac{|Y(m,f)|^2}{|\hat{N}(m,f)|^2} \quad (4)$$

3. AUDITORY SPECTRAL ATTENUATION

Commonly, the subtractive type algorithms for speech enhancement are based on short time spectral analysis according to a uniform decomposition of the noisy speech. Many aspects of audiology and psychoacoustics demonstrate that the human auditory system analysis sound in the time-frequency domain with a non linear frequency selectivity of the basilar membrane. Thus the human ear analysis can be conceptualized as an array of overlapping band-pass filters known as auditory filters. These filters occur along the basilar membrane and increase the frequency selectivity of the human ear, therefore the sensitivity to abrupt stimulus change and the transition component in speech perception. According these assumptions, an improvement speech perception in noise environment may be possible, since the speech component can be identified and the selectivity can be amplified. This suggests an approach to improving the quality and intelligibility of speech in background noise using spectral analysis according psychoacoustics aspects. In order to enhance the noisy speech, it is interesting to implement subtractive type methods according a non uniform filterbank analysis.

3.1 Auditory Filters modelling

The aim in auditory modeling is to find mathematical model which represents some physiological and perceptual aspects of the human auditory system. Auditory modeling is very useful, since the sound wave can be analyzed according the human ear comportment, with a good mode.
The simplest way to model the frequency resolution of the basilar membrane is to make analysis using filterbanks. The simplest and the most realistic model is the gammatone filterbanks [5], the impulsion response is based on psychoacoustics measurements, providing a more accurate approximation to the auditory frequency response, and it is represented by a gammatone function defined in the temporal model by the following expression:

$$g(t)=At^{n-1}\exp(-2\pi bB(f_c)t)\cos(2\pi f_c t+\varphi) \quad (5)$$

Where A defines the magnitude normalization parameter, n is the filter order, *fc* is the center frequency of filters, B is filters bandwidths and $bB(f_c)$ represents the filter envelop.

3.2 Choice of frequency scale

The frequency resolution of human hearing is a complex phenomenon which depends on many factors, such as frequency, signal bandwidth, and signal level. Despite of the fact that our ear is very accurate in single frequency analysis, broadband signals are analyzed using quite sparse frequency resolution. Critical bandwidth and the Equivalent Rectangular Bandwidth (ERB) scale are an accurate way to explain the frequency resolution of human hearing with broadband signals. The expression used to convert a frequency *f* in Hz in its value in ERB is:

$$ERB(f)=21.4\log\left(\frac{4.37f}{1000}+1\right) \quad (6)$$

3.3 Auditory spectral attenuation

The proposed speech enhancement method (figure1) is based on non-uniform decomposition of the input waveform. The processing is done by dividing the incoming noisy speech into separate frequency bands that could be individually manipulated using the spectral subtractive algorithms to achieve quality and intelligibility improvement of the overall signal.

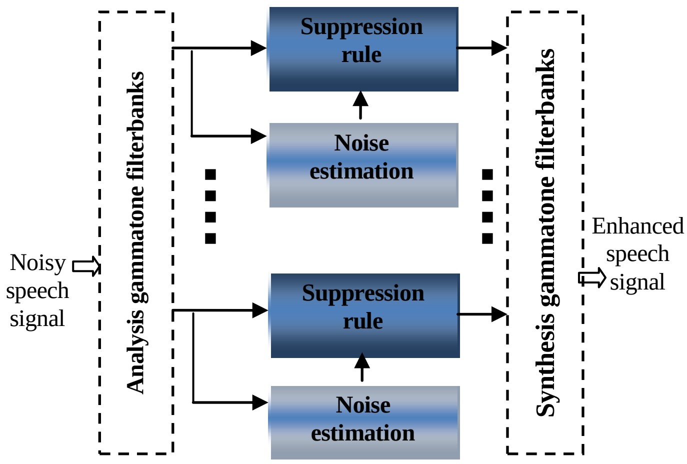

Figure 1 – Proposed speech enhancement method based on auditory spectral attenuation.

The analysis filterbank consists of 4th order gammatone filters that cover the frequency range of the signal. The filters bandwidth changes according the Equivalent rectangular bandwidth ERB scale. The output of the kth filter of the analysis filterbank can be expressed as:

$$y_k(n)=y(n)*g_k(n) \quad (7)$$

Where $g_k(n)$ is the impulse response of the kh, 4th-order gammatone filter.
Each sub-band of the noisy speech is manipulated using the spectral subtraction rule and the power spectral density of the noisy speech in each frequency sub-band *k* is given

according to the following expression:

$$\left|\hat{X}_k(m,f)\right|^2 = \left|Y_k(m,f)\right|^2 - \left|\hat{N}_k(m,f)\right|^2 \quad (8)$$

Using the Ephraim and Malah noise suppression algorithm, the spectral magnitude of the enhanced speech in each frequency band is estimated as:

$$\left|\hat{X}_k(m,f)\right| = G_k^{EM}(m,f)\cdot\left|Y_k(m,f)\right| \quad (9)$$

Where G_k^{EM} (k) is the spectral gain applied in each frequency using the relationship (2).

In the synthesis filterbank, the final enhanced output speech signal is obtained using the summation of the sub-band signals after processing.

The noise estimate has an important impact on the quality and intelligibility of the enhanced signal. If the noise estimate is too low, a residual noise will be audible, if the noise estimate is too high, speech will be distorted resulting in intelligibility loss. In the SS and the EM spectral attenuation algorithms, the noise spectrum is estimated by taking its average value computed during silent segment. Although this approach might give satisfactory result with stationary noise, it will not with more realistic environments where the spectral characteristics of the noise change constantly. Hence there is a need to update the noise spectrum continuously over time. Several noise-estimation algorithms have been proposed for speech enhancement applications [7]. In [8], the method for estimating the noise spectrum (Martin) is based on tracking the minimum of the noisy speech over a finite window based on the statistics of the minimum estimates. In [9], a minima controlled recursive algorithm (MCRA) is proposed; it updates the noise estimate by tracking the noise-only regions of the noisy speech spectrum. In the improved minima controlled recursive algorithm (IMCRA) approach [10], a different method was used to track the noise-only regions of the spectrum based on the estimated speech-presence probability. Recently a new noise estimation algorithm (MCRA2) was introduced [11], the noise estimate was updated in each frame based on voice activity detection based on the ratio of noise speech spectrum to its local minimum.

4. RESULTS AND EVALUATION

In our work, we evaluate the proposed auditory spectral attenuation for speech enhancement compared with the power spectral subtraction (SS) and the Ephraim and Malah (EM) subtractive rules under a noise environment. Speech signals are taken from the NOIZEUS speech corpus [12] sampled at 8 KHz and degraded by different noises, suburban train noise, multi-talker babble, car and street noise. The test signals include 30 speech utterances from 3 different speakers, female and male. To cover the frequency range of the signal, the analysis stage used in the auditory spectral subtraction (SS_GTFB) and the auditory Ephraim and Malah spectral attenuation (EM_GTFB) consists of 27- 4th order gammatone filters according to ERB scale. The background noise spectrum is continuously estimated using the MCRA, IMCRA, MCRA2 and Martin noise estimation algorithms. A part from noise reduction, naturalness and intelligibility of enhanced speech are important attributes of the performance of any speech enhancement system. Since achieving a high degree of noise suppression is often accompanied by speech signal distortion, it is important to evaluate both quality and intelligibility. The performance evaluation in our work includes a subjective test of perceptual evaluation of speech quality PESQ [6].

Table 1: *PESQ score for the proposed method (SS_GTFB) compared with the spectral subtraction (SS).*

Noises	SNR (dB)	SS	SS_GTFB			
			MCRA	IMCRA	MCRA2	Martin
Babble	0	1.78	1.93	1.89	1.96	1.84
	5	2,10	2.17	2.15	2.18	2.16
	10	2,40	2.51	2.45	2.52	2.43
	15	2,73	2.69	2.54	2.68	2.63
Car	0	1,83	1.81	1.79	1.80	1.75
	5	2,12	2.24	2.20	2.21	2.13
	10	2,46	2.33	2.22	2.32	2.25
	15	2,73	2.88	2.74	2.86	2.71
Street	0	1,73	1.86	1.80	1.83	1.82
	5	2,07	2.18	2.14	2.15	2.09
	10	2,40	2.51	2.47	2.49	2.43
	15	2,65	2.81	2.75	2.80	2.71
train	0	1,81	1.87	1.84	1.87	1.76
	5	2,17	2.16	2.12	2.14	2.02
	10	2,45	2.48	2.43	2.46	2.32
	15	2,72	2.86	2.81	2.85	2.68

Table 2: *PESQ score for the proposed method (EM_GTFB) compared with the Ephraim and Malah rule (EM).*

Noises	SNR (dB)	EM	EM_GTFB			
			MCRA	IMCRA	MCRA2	Martin
Babble	0	1.82	1.87	1.87	1.86	1.78
	5	2.15	2.20	2.18	2.19	2.09
	10	2.45	2.55	2.51	2.53	2.40
	15	2.79	2.89	2.85	2.86	2.75
Car	0	1.81	1.92	1.88	1.94	1.74
	5	2.09	2.23	2.18	2.21	2.01
	10	2.43	2.57	2.48	2.56	2.34
	15	2.77	2.91	2.86	2.87	2.67
Street	0	1.70	1.86	1.78	1.86	1.73
	5	2.02	2.18	2.14	2.18	2.02
	10	2.38	2.53	2.46	2.52	2.38
	15	2.66	2.80	2.75	2.79	2.67
train	0	1.80	1.87	1.85	1.85	1.70
	5	2.05	2.16	2.13	2.13	1.96
	10	2.35	2.46	2.42	2.44	2.26
	15	2.70	2.83	2.78	2.77	2.61

In fact, significant gains in noise reduction are accompanied by a decrease in speech intelligibility. Formal subjective listening test are the best indicates of achieved of overall quality. So the subjective listening test [13] used instructs the listener (32 listeners) to successively attend and rate the enhanced speech signal on : the speech signal alone using five-point scale of signal distortion (SIG) [5= very natural, no degraded, 4= Fairly natural, little degradation, 3= Somewhat natural, somewhat degraded, 2= Fairly unnatural, fairly degraded, 1=very unnatural, very degraded]. The background noise alone using a five-point scale of background intrusiveness (BAK) [5= Not noticeable, 4= somewhat noticeable, 3= noticeable but not intrusive, 2=

Fairly conspicuous, 1= Very conspicuous, very intrusive somewhat intrusive] and the overall effect using the scale of the mean opinion score (OVRL) [1=bad, 2=poor, 3=fair, 4=good, 5=excellent]. This process is designed to integrate the effects of both the signal and the background in making the rating of overall quality. Table 1 and table 2 list the PESQ score obtained after processing, we observe that the PESQ score is consistent with the subjectively perceived trend of an improvement in speech quality with the proposed speech enhancement approach over that the spectral subtractive algorithms alone. This improvement is particularly significant in the case of car noise at 15 dB, we register a score of 2,91 for the proposed EM_GTFB in spite of 2,77 for the EM alone, the PESQ improvement is also observed using the SS_GTFB at 0 dB (1,96) for babble noise continuously estimated with the MCRA2, contrary in the SS (1,78).Table 3 and table 4 list at different signal to noise ratio the subjective overall quality the OVRL measure that includes the naturalness of speech (SIG) and intrusiveness of background noise (BAK). We notice that the proposed auditory spectral attenuation using different continuous noise estimation algorithms performed significantly better than the classic subtractive attenuation. Lower signal distortion (higher SIG score) is observed with the proposed approach in most condition with significant differences at 10dB for car noise: a SIG score of 3,09 given by the SS, and improved by the SS_GTFB to 3,59 using Martin noise estimation and a score of 3,64 registered by the proposed EM_GTFB with the MCRA noise estimation. This demonstrates the performance of our approach to reduce the noticeable of the background noise and minimize the signal distortion.

We notice, also that incorporating continuous noise estimation in particularly the IMCRA and the MCRA2 continuous noise estimation in the auditory spectral attenuation approach performed better than the power spectral subtraction and the Ephraim and Malah rules in overall quality. This indicates that the proposed auditory spectral attenuation for speech enhancement is sensitive to the noise spectrum estimate. The results obtained show that the proposed speech enhancement method using different continuous noise estimation performed, in most condition, better than the classic spectral attenuation algorithms in terms of perceptual improvement, overall quality and low signal distortion. The auditory spectral analysis contributed significantly to the speech enhancement and to the improvement of the voice quality at different signal to noise ratio and practically for all the types of noise. Indeed the decomposition in filterbank and the continuous noise estimation given the best subjective results with regard to the subtractive noise reduction method based on uniform decomposing using Fourier transform and a simple approach to estimate the noise during the silent moment.

Table 3: *SIG- BAK -OVRL scores for the proposed speech enhancement compared to the subtractive algorithms at 0 dB and 5 dB.*

		Babble			**Car**			**Street**			**train**		
SNR= 0dB		**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**
SS		2,08	1,46	1,68	2,29	1,63	1,85	2,00	1,52	1,63	2,04	1,62	1,69
SS_GTFB	MCRA	2,22	1,19	1,84	2,46	1,31	2,04	2,31	1,25	1,90	2,36	1,33	1,97
	MCRA2	2,16	1,20	1,81	2,36	1,30	1,97	2,24	1,24	1,85	2,32	1,33	1,94
	IMCRA	2,30	1,22	1,89	2,46	1,31	2,03	2,34	1,27	1,92	2,40	1,35	2,00
	MARTIN	2,48	1,31	2,00	2,59	1,38	2,09	2,50	1,34	2,02	2,45	1,37	1,99
EM		2,44	1,68	1,97	2,43	1,74	1,97	2,42	1,68	1,95	2,31	1,78	1,92
EM_GTFB	MCRA	2,45	1,32	2,01	2,64	1,41	2,16	2,51	1,37	2,06	2,54	1,44	2,10
	MCRA2	2,37	1,32	1,97	2,62	1,44	2,16	2,47	1,38	2,04	2,51	1,44	2,08
	IMCRA	2,39	1,29	1,97	2,43	1,32	2,01	2,36	1,29	1,92	2,40	1,37	1,99
	MARTIN	2,54	1,34	2,03	2,54	1,36	2,03	2,53	1,36	2,02	2,42	1,38	1,96
		Babble			**Car**			**Street**			**train**		
SNR= 5dB		**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**
SS		2,59	1,84	2,12	2,78	1,96	2,26	2,63	1,91	2,15	2,53	1,96	2,11
SS_GTFB	MCRA	2,83	1,47	2,36	3,05	1,59	2,54	2,91	1,55	2,42	2,88	1,58	2,42
	MCRA2	2,71	1,45	2,29	2,94	1,56	2,46	2,84	1,53	2,37	2,78	1,56	2,34
	IMCRA	2,87	1,49	2,39	3,06	1,59	2,54	2,93	1,55	2,43	2,90	1,59	2,42
	MARTIN	3,02	1,56	2,46	3,10	1,61	2,52	3,03	1,60	2,47	2,93	1,60	2,40
EM		2,96	2,09	2,44	3,02	2,16	2,46	3,04	2,14	2,48	2,86	2,17	2,37
EM_GTFB	MCRA	3,05	1,60	2,52	3,17	1,67	2,61	3,08	1,64	2,54	3,04	1,67	2,53
	MCRA2	2,95	1,59	2,46	3,13	1,68	2,59	2,96	1,58	2,47	2,95	1,67	2,47
	IMCRA	2,93	1,54	2,43	2,98	1,57	2,47	2,99	1,58	2,47	2,92	1,60	2,44
	MARTIN	3,07	1,59	2,49	3,04	1,59	2,45	3,04	1,60	2,46	2,88	1,60	2,36

Table 4 *SIG- BAK -OVRL scores for the proposed speech enhancement compared to the subtractive algorithms at 10 dB and 15 dB.*

		Babble			**Car**			**Street**			**train**		
SNR= 10dB		**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**
SS		2,99	2,16	2,49	3,09	2,23	2,57	3,01	2,20	2,49	2,93	2,21	2,45
SS_GTFB	MCRA	3,45	1,78	2,89	3,52	1,84	2,98	3,45	1,81	2,89	3,40	1,82	2,86
	MCRA2	3,28	1,73	2,77	3,40	1,81	2,89	3,32	1,77	2,81	3,27	1,78	2,77
	IMCRA	3,47	1,79	2,90	3,55	1,85	2,98	3,46	1,,81	2,90	3,42	1,83	2,88
	MARTIN	3,55	1,82	2,92	3,59	1,85	2,96	3,54	1,84	2,93	3,44	1,82	2,84
EM		3,53	2,53	2,92	3,45	2,54	2,87	3,54	2,55	2,92	3,39	2,55	2,82
EM_GTFB	MCRA	3,59	1,88	3,00	3,64	1,92	3,05	3,56	1,89	2,99	3,51	1,89	2,94
	MCRA2	3,49	1,87	2,95	3,60	1,92	3,03	3,47	1,87	2,93	3,41	1,87	2,88
	IMCRA	3,52	1,83	2,94	3,47	1,82	2,90	3,44	1,82	2,88	3,40	1,83	2,85
	MARTIN	3,57	1,83	2,93	3,54	1,82	2,89	3,53	1,85	2,91	3,38	1,81	2,78
		Babble			**Car**			**Street**			**train**		
SNR= 15dB		**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**	**SIG**	**BAK**	**OVRL**
SS		3,33	2,41	2,79	3,41	2,44	2,83	3,28	2,39	2,73	3,29	2,44	2,76
SS_GTFB	MCRA	3,93	2,05	3,34	4,04	2,10	3,44	3,83	2,01	3,26	3,92	2,09	3,35
	MCRA2	3,76	2,00	3,22	3,89	2,05	3,33	3,32	1,77	2,81	3,77	2,04	3,24
	IMCRA	3,95	2,05	2,90	4,08	2,10	3,46	3,88	2,02	3,28	3,94	2,09	2,88
	MARTIN	4,00	2,06	3,35	4,05	2,08	3,37	3,95	2,05	3,30	3,93	2,06	3,28
E**M**		3,98	2,94	3,34	4,01	2,98	3,35	3,98	2,94	3,33	3,87	2,96	3,26
EM_GTFB	MCRA	4,03	2,12	3,42	4,10	2,15	3,48	3,93	2,07	3,33	3,98	2,12	3,38
	MCRA2	3,93	2,11	3,36	4,00	2,14	3,41	3,85	2,06	3,28	3,84	2,09	3,28
	IMCRA	3,98	2,08	3,37	4,01	2,10	3,40	4,00	2,07	3,38	3,86	2,03	3,26
	MARTIN	4,03	2,07	3,36	3,99	2,05	3,31	3,94	2,04	3,28	3,88	2,04	3,23

5. CONCLUSION

In this paper, we proposed a new noise reduction method which consists in integrating an auditory analysis in the subtractive process of the noise. We noticed that the use of a frequency resolution according to the critical bands behavior, in particular the ERB scale, allowed to obtain, from the perceptive point of view and from the vocal quality, better results than those supplied by the classic spectral subtraction rules in improving the quality and intelligibility of the enhanced speech signal.

REFERENCES

[1]S.F. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech", in *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing, Washington*, DC, pp. 200-203, Apr. 1979.

[2]Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", *IEEE transaction on acoustics speech and signal processing*, vol.assp-33,no 2, December 1985.

[3]D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression", IEEE Trans. Speech and Audio Processing, vol. 5, pp. 497 514, Nov. 1997.

[4] N. Virag, "Single channel speech enhancement based on masking proprieties of the human auditory system",*IEEE Trans. Speech and Audio Processing, vol. 7, pp. 126 137, Mar. 1999.*

[5] V. Hohmann, "Frequency analysis and synthesis using a Gammatone filterbank", *Acta Acustica united with Acoustica, 88(3), 433-442.(2002).*

[6] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", *International Telecommunication Union*, Geneva, 2001.

[7] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL.

[8] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", *IEEE Trans. on Speech and Audio Processing*, vol. 9,no. 5, pp. 504-512, July 2001.

[9] I. Cohen, and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement", *IEEE Signal Proc. Letters*, vol.9, no. 1, pp. 12-15, January 2002.

[10] I. Cohen, "Noise Spectrum estimation in adverse environnements: improved minima controlled recursive averaging". IEEE Trans. Speech Audio Process. 11 (5), 466–475, 2003.

[11] S. Rangachari, P. C Loizou and Y. Hu, "A noise estimation algorithm with rapid adaptation for highly non-stationary environments", in *IEEE Int. Conf. on Acoustics, Speech, signal processing,* pp. I-305-I-308, May 17-21 2004.

[12] Y. Hu and P. Loizou, " Subjective comparison of speech enhancement algorithms", *Proc. IEEE Int. conf. Acoust. Speech Signal Process*., I,pp.153-156. 2006.

[13] Y. Hu and P. Loizou, "Evaluation of objective measures for speech enhancement, *Proc. Interspeech*, pp. 1447-1450. 2006.