Research Journal of Applied Sciences, Engineering and Technology 6(6): 1081-1087, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: October 31, 2012 Accepted: December 28, 2012 Published: June 30, 2013

Speech Enhancement with Geometric Advent of Spectral Subtraction using Connected Time-Frequency Regions Noise Estimation

[1]Nasir Saleem, [2]Sher Ali, [3]Usman Khan and [4]Farman Ullah
[1]Institute of Engineering and Technology, GU, D.I. Khan, KPK, Pakistan
[2]City University of Science and Technology, Peshawar, KPK, Pakistan
[3]University of Engineering and Technology, Kohat, KPK, Pakistan
[4]COMSATS Institute of IT Quaid Avenue, Wah Cantt, Punjab, Pakistan

Abstract: Speech enhancement with Geometric Advent of Spectral subtraction using connected time-frequency regions noise estimation aims to de-noise or reduce background noise from the noisy speech for better quality, pleasantness and improved intelligibility. Numerous enhancement methods are proposed including spectral subtraction, subspace, statistical with different noise estimations. The traditional spectral subtraction techniques are reasonably simple to implement and suffer from musical noise. This study addresses the new approach for speech enhancement which has minimized the insufficiencies in traditional spectral subtraction algorithms using MCRA. This approach with noise estimation has been evolved with PESQ, the ITU-T standard; Frequency weighted segmental SNR and weighted spectral slope. The analysis shows that Geometric approach with time-frequency connected regions has improved results than old-fashioned spectral subtraction algorithms. The normal hearing tests has suggested that new approach has lower audible musical noise.

Keywords: Frequency connected regions, FwSNRseg, MCRA, PESQ, speech enhancement, spectral subtraction, WSS

INTRODUCTION

The fundamental objective behind speech enhancement is to remove or reduce background noise. The background noise removal has a number of applications like using telephone in noisy environments including streets, public places etc. all these applications demand to reduce noise for normal hearing aids and improved quality. The spectral subtraction for speech enhancement with geometric approach (Yang and Philipos, 2008) is used with Time-Frequency connected regions (Karsten and Søren, 2005) noise estimation algorithm. Our aim is to test approach with different noise estimation algorithms (Martin, 2001) and compare results with other present methods to select appropriate estimation algorithms. The spectral subtraction technique (Loizou, 2007; Boll, 1979) works on very simple principle by assuming additive noise. The estimation algorithms estimate novel and Noisy speech spectra and subtract noise estimated spectrum form clean spectrum. The estimation of noise spectrum is computed in periods where signals are not present. If estimated signal spectrum is passed through inverse discrete Fourier transform which utilize phase of noisy signal, we obtain enhanced speech. The subtraction procedure has to be performed sensibly in order to sidestep the signal distortion. In case of over subtraction, major portion of speech is also subtracted and for under subtraction, small portion of noise still interfere the signal. The Fig. 1 shows the block diagram of spectral subtraction algorithm. Many algorithms are developed with different solutions, some suggests over subtraction (Berouti *et al*., 1979), some came with suggestion that speech spectrum is divided into

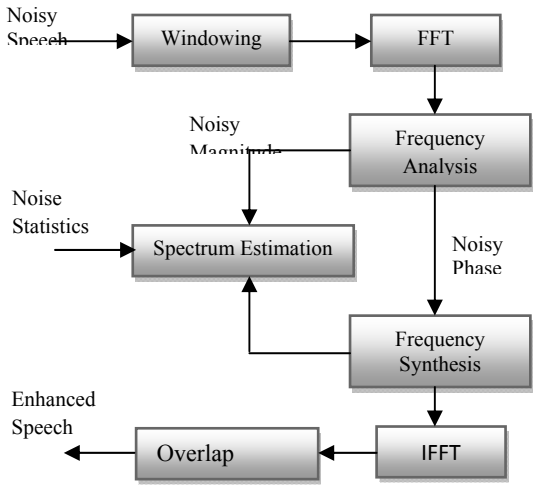

Fig. 1: Spectral subtraction algorithm block diagram

Corresponding Author: Nasir Saleem, Institute of Engineering and Technology, GU, D.I. Khan, KPK, Pakistan

continuous frequency bins and apply non-linear methods in bins (Kamath and Loizou, 2002) and some suggested psychoacoustical methods (Virag, 1999). The spectral subtraction algorithm is easy to implement for effective use to eliminate the background noise but still there are major weaknesses in this approach. Among those shortcomings one is the introduction of musical noise (Berouti *et al*., 1979). The estimated spectrum may contain some negative values which occur due to wrong estimation. One way is to use non-linear process to remove these errors by setting all those negative values to zero in order to guarantee the non-negative magnitude spectrum. But by doing so small random isolated peaks are generated in spectrum. These peaks sound like tones in time-domain which continuously changing frame wise. These newly generated tones are called musical noise. The spectral subtraction equations are derived on some norms which assume that cross terms are zero because of un-correlation nature of speech and interrupting noise. And this assumption is valid as speech and noise are statistically independent of each other means there is no correlation among them. As a result it is concluded that these equations are estimated not particular ones. But in Geometric approach with time-frequency connected region noise estimation, the equations for estimation of noise become non-negative and as result gain function will always be positive.

SPECTRAL SUBTRACTION MATHEMATICAL ANALYSIS

Consider s (n) is novel speech and e (n) is error signal (noise) and y (n) is noisy signal contains clean and error signal:

$$y(n) = s(n) + e(n) \quad (1)$$

By taking STFT of y (n), the resultant frequency-domain equation is:

$$Y(j\omega_n) = S(j\omega_n) + E(j\omega_n) \quad (2)$$

$\omega_n = 2\pi n/N$, where n = 0, 1, 2 and 3..... N-1 and N represents frame length. For short-term power spectrum of noisy speech computation, the $Y(j\omega_n)$ is multiplied with its conjugate that is $Y^*(j\omega_n)$:

$$|Y(j\omega_n)|^2 = |S(j\omega_n)|^2 + |E(j\omega_n)|^2 + S(j\omega_n).E^*(j\omega_n) + S^*(j\omega_n).E(j\omega_n) \quad (3)$$

$$|Y(j\omega_n)|^2 = |S(j\omega_n)|^2 + |E(j\omega_n)|^2 + 2|S(j\omega_n)|.|E(j\omega_n)|\cos(\theta_S(k) - \theta_E(k)) \quad (4)$$

The terms $|E(j\omega_n)|^2$, $S(j\omega_n).E^*(j\omega_n)$ and $S^*(j\omega_n).E(j\omega_n)$ are estimated with expectation operator $E\{.\}$as$E\{|E(j\omega_n)|^2\}$, $E\{S(j\omega_n).E^*(j\omega_n)\}$ and $E\{S^*(j\omega_n).E(j\omega_n)\}$. Now consider that e (n) is zero and there is no correlation with the novel signal s (n), the above terms will reduce to zero and equation for novel speech estimation will:

$$|S(j\omega_n)|^2 = |Y(j\omega_n)|^2 - |E(j\omega_n)|^2 \quad (5)$$

The gain or suppression function can be calculated from Eq. (4) as:

$$|S(j\omega_n)|^2/|Y(j\omega_n)|^2 = 1 - |E(j\omega_n)|^2/|Y(j\omega_n)|^2$$
$$|S(j\omega_n)|/|Y(j\omega_n)| = \sqrt{1 - |E(j\omega_n)|^2/|Y(j\omega_n)|^2}$$
$$|H(j\omega_n)|^2 = 1 - |E(j\omega_n)|^2/|Y(j\omega_n)|^2 \quad (6)$$

Equation (5) becomes:

$$|S(j\omega_n)|^2 = |Y(j\omega_n)|^2|H(j\omega_n)|^2 \quad (7)$$

By neglecting the cross terms in equation (4), $H(j\omega_n)$ will always be positive with range $0 \le H(j\omega_n) \le 1$. The cross terms can be computed from Eq. (4) as:

$$2|S(j\omega_n)|.|E(j\omega_n)|\cos(\theta_S(k) - \theta_E(k)) = \Delta Y(j\omega_n) \quad (8)$$

$$|Y(j\omega_n)|^2 = |S(j\omega_n)|^2 + |E(j\omega_n)|^2 + \Delta Y(j\omega_n) \quad (9)$$

The term $|S(j\omega_n)|^2 + |E(j\omega_n)|^2$ are replaced with $Y'(j\omega_n)$ and then equation becomes:

$$|Y(j\omega_n)|^2 = |Y'(j\omega_n)|^2 + \Delta Y(j\omega_n) \quad (10)$$

When the cross terms are neglected, the resultant error is:

$$\text{Error}(k) = |Y(j\omega_n)|^2 - |Y'(j\omega_n)|^2/|Y(j\omega_n)|^2$$
$$\text{Error}(k) = \Delta Y(j\omega_n)/|Y(j\omega_n)|^2 \quad (11)$$

The cross term error shows that actual noise spectrum estimation is not fulfilled that needs to be estimated which results in random tones.

GEOMETRIC SPECTRAL SUBTRACTION

The noisy spectrum $Y(j\omega_n)$ at frequency $\omega_n$is computed by the summation of two complex valued spectra. These spectra are now represented in complex geometrical plane where the $Y(j\omega_n)$ is the sum of two complex spectra $S(j\omega_n)$ and $E(j\omega_n)$ respectively. Representation of the complex values is sketched in Fig. 2.

In traditional spectral subtraction the cross terms are assumed to be zero for computing gain function but a new gain function is now computed without any assumption by transforming the Eq. (1) to polar notation:

$$a_Y e^{j\varphi Y} = a_S e^{j\varphi s} + a_E e^{j} \quad (12)$$

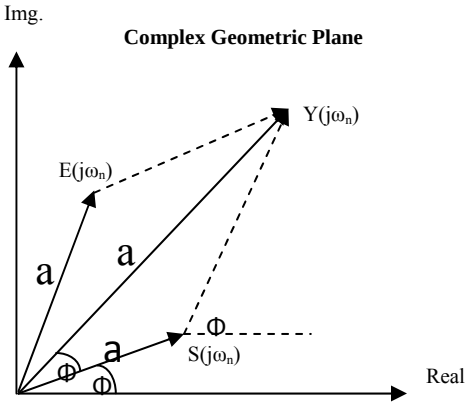

Fig. 2: Noise spectrum in complex geometric plane

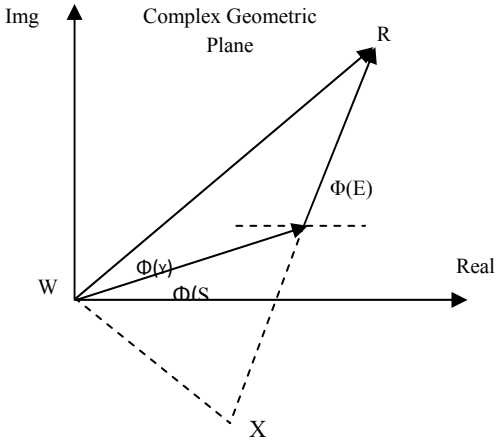

Fig. 3: Geometric relation between novel and noisy speech spectra

$\{a_Y a_S a_E\}$ are the magnitudes where $\{\varphi_Y\ \varphi_s\ \varphi_E\}$ are the phase angles for novel and noisy speech respectively. From Fig. 2, if we solve the triangle in Fig. 3 with the help of laws of sines, we can obtain a new suppression or gain function which will always be positive and real. By solving the above triangle WRX, following equations are computed which represents that WR is perpendicular to XR:

$$WX = a_Y \sin(\varphi_E - \varphi_Y) = a_S \sin(\varphi_E - \varphi_s) \quad (13)$$

$$|WX| = a^2_Y \sin^2(\varphi_E - \varphi_Y) = a^2_S \sin^2(\varphi_E - \varphi_s) \quad (14)$$

$$|WX| = a^2_Y (1 - \cos^2(\varphi_E - \varphi_Y)) \quad (15)$$

$$|WX| = a^2_S (1 - \cos^2(\varphi_E - \varphi s)) \quad (16)$$

The new gain function can be calculated from Fig. 3 as:

$$H^2_G = a^2_S/a^2 = H_G = \sqrt{a^2_S/a^2_Y} \quad (17)$$

$$H_G = \sqrt{1 - [\cos^2(\varphi_E - \varphi_Y)]/[1 - \cos^2(\varphi_E - \varphi s)]} \quad (18)$$

This gain function is always positive, that is, $H_G \geq 0$. The block diagram of the Geometric approach of spectral subtraction for enhanced speech is sketched in Fig. 4.

BACKGROUND NOISE ESTIMATION

The key objective of speech enhancement is to eliminate or reduce the background noise by estimating noise. All speech enhancement algorithms normally use estimation methods for this purpose. If the background noise is progressing gently along with speech, its estimation is easy in pause periods of speech but if there is rapid noise growing, estimation becomes more difficult. Some of the estimation algorithms are discussed in this section including MCRA (Bernard *et al*., 2005) and frequency connected regions

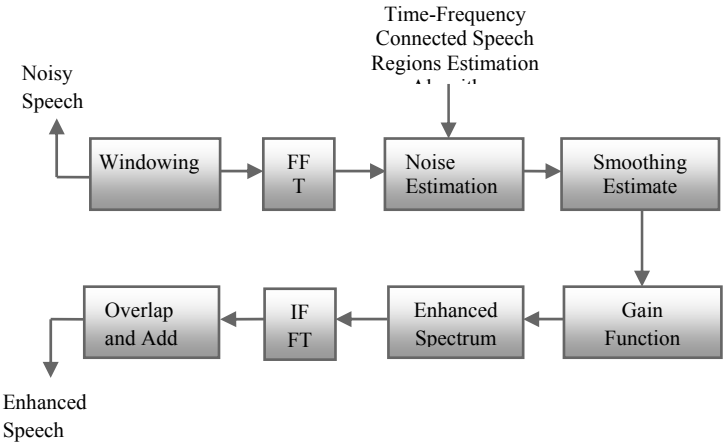

Fig. 4: Block diagram of spectral subtraction with geometric approach

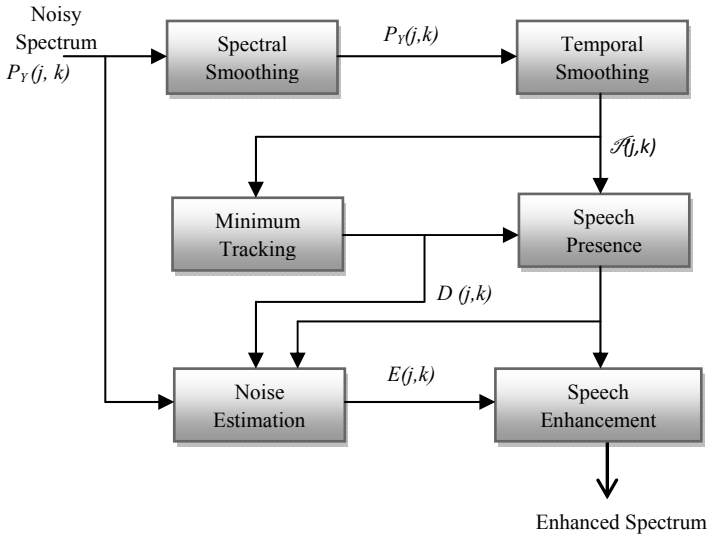

Fig. 5: Time-frequency connected speech regions block diagram

estimation (used estimation algorithm in GA spectral algorithm).

MCRA: MCRA was introduced to approximate non-stationary background noise. The noise approximation in this algorithm is updated by utilizing averaging of previous spectral values of noisy spectrum which is measured by time and frequency dependent smoothing elements. The smoothing elements are computed on the basis of signal presence probability in frequency band and the probability is computed by utilizing the ratio of noisy speech spectrum to minimum evaluated over a fixed time. The estimation of noise spectrum from signal presence probability is computed on basis of following supposition:

$$S_A: \quad Y(j, k) = E(j, k) \tag{19}$$

$$S_P: \quad Y(j, k) = S(j, k) + E(j, k) \tag{20}$$

j represents frame No and k shows the frequency bin No. where S_A and S_P shows supposition of speech absence and presence respectively. This algorithm for noise estimation utilizes progressive recursive averaging given as:

$$S_A: \Psi(j,k+1) = \beta_n \Psi_n(j,k) + (1 + \beta_n) |E(j,k)|^2 \tag{21}$$

$$S_P: \Psi(j,k+1) = \Psi_n(j,k) \tag{22}$$

β_n is the smoothing element having range of $0 \leq \beta_n \leq 1$ and $\Psi_n(j, k)$ shows the amplitude power spectrum of noise computed by expectation operator $E\{.\}$. The speech presence probability can be computed from following equations:

$$P(j, k+1) = \beta_P p(j, k) + (1 + \beta_P) \quad \text{When } S(j,k)/S_{min}(j, k) > \zeta \tag{23}$$

$$P(j, k+1) = \beta_P p(j, k) \quad \text{When } S(j, k)/S_{min}(j, k) < \zeta \tag{24}$$

ζ represents threshold level of presence while $S(j, k)/S_{min}(j, k)$ shows ratio of noisy spectrum to its local minimum.

CONNECTED TIME-FREQUENCY REGIONS NOISE ESTIMATION

The block diagram for the connected time-frequency region noise estimation algorithm is shown in Fig. 5. After windowing speech, STFT is applied to compute periodogram of noisy speech, that is, $P_Y(j, k) = |Y(j, k)|^2$. After computing periodograms, they are under process of smoothing. The smoothed periodograms are temporally minimum tracked and are used for purpose of speech presence detection. This detection is utilized to attain low biased noise PSD estimates $P'_E(j, k)$ and for noise periodogram estimates $P_E(j, k)$ which is equal to $P_Y(j, k)$ in speech absence condition. But if speech is present, noise periodogram estimate is equal to noise PSD estimation. In later case, recursive smoothed bias compensation parameter is put on minimum tracked values. The bias compensation factor is updated during absence of speech in frames while remain unchanged during speech presence. The noise magnitude periodogram estimation $|E(j, k)|$ is computed from noise PSD estimation and on basis of these information, decision of speech presence is made and used in speech enhancement algorithm. The noisy speech periodograms $P_Y(j, k)$ are spectrally smoothed. The $P_Y(j, k)$ bands are composed of weighted sum of 2N+1 band. The spectral smoothing equation is:

$$P_Y(j,k) = \sum_{i=-N}^{+N} b(i)\, P_Y(j, (k-i)_K) \quad (25)$$

$(k-i)_K$ represents the modulus K and K shows complete spectrum length. The windowing function b (i) is used for spectral weighting which sums to 1, that is, $\sum_{i=N}^{+N} b(i) = 1$. The spectrally smoothed periodograms are temporally smoothed recursively with time-frequency changing smoothing factor ξ (j, k) to create the temporally spectrally smoothed periodogram P (j, k):

$$P(j,k) = \xi(j,k)\, P(j-1, k) + (1 - \xi(j,k))\, P_Y(j, k) \quad (26)$$

The temporal minimum values $P_{min}(j,k)$ are computed from P(j,k) by tracing within minimum search window have length W_{min}:

$$P_{min}(j, k) = \text{Min}\, [P(\varepsilon, k)]\ |j - W_{min} < \varepsilon \leq j| \quad (27)$$

The $P_{min}(j, k)$ tracks are utilized in speech presence. The speech presence results in increase of power in temporally smoothed spectrum because of the additive noise at particular time-frequency regions. As a result ratio of temporally smoothed spectrum to noise PSD estimate becomes more robust to estimate the SNR and noise-to-noise ratio at specific time-frequency regions. The smoothing phenomenon ensures the speech presence detection even in conditions where noisy speech power is unstable. As a result, connected speech presence and absence regions can be achieved.

Here we have computed two different noise estimations; one is noise PSD estimation and second is noise spectrum estimation. The PSD estimation is used in speech enhancement algorithm while noise spectrum estimation shows the properties of residual noise from speech enhancement algorithm. The speech enhancement algorithm for this noise estimation is spectral subtraction with geometric approach.

EXPERIMENTAL SETUP AND EVALUATIONS

The NOIZEUS (Hu and Loizou, 2007) a noisy speech database is developed to evaluate speech enhancement algorithms. Three real world noise environments including Airport, exhibition hall and street are considered at different noise levels ranging from 0dB to 15dB. The speech enhancement algorithm with noise estimation is evaluated with PESQ (ITU, 2000; Bernard *et al*., 2005; Jianfen *et al*., 2009), FwSNRseg (Bernard *et al*., 2005; Jianfen *et al*., 2009) and WSS (Bernard *et al*., 2005; Jianfen *et al*., 2009). PESQ is ITU-T standard to evaluate the perceptual quality of enhanced speech. Basically the PESQ estimates the Mean Opinion Score (Loizou, 2007) from novel and degraded speech signals. PESQ (Table 1 and 2) is rated with following five-point scale.

Frequency weighted segmental SNR is one of variant of SNR which is weighted SNRseg within the particular frequency bin which is related to the critical bin. Noise in certain frequency bins is less harmful than in other bins of input speech signal. Higher the FwSNRseg (Table 3) value better is quality. Weighted Spectral slope (Table 4) measures the distance in

Table 1: PESQ rating with MOS scale

Rating	Quality	Degradation
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible, slightly annoying
2	Poor	Annoying
1	Unsatisfactory	Very annoying

Table 2: Perceptual Evaluation of Speech Quality (PESQ) experimental results

	Noise level	TFCR noise estimation	MCRA noise estimation
Airport noisy environment	0dB	1.31	1.13
	5dB	1.92	1.85
	10dB	2.38	2.33
	15dB	2.86	2.64
Exhibition noisy environment	0dB	1.37	1.22
	5dB	1.75	1.72
	10dB	2.24	2.14
	15dB	2.59	2.59
Street noisy environment	0dB	1.80	1.76
	5dB	2.10	2.10
	10dB	2.53	2.30
	15dB	2.84	2.43

Table 3: Frequency weighted segmental SNR (FwSNRseg) experimental results

	Noise level	TFCR noise estimation	MCRA noise estimation
Airport noisy environment	0dB	4.57	5.51
	5dB	6.45	6.05
	10dB	6.95	7.02
	15dB	8.65	7.18
Exhibition noisy environment	0dB	4.66	4.92
	5dB	5.46	5.17
	10dB	6.44	5.89
	15dB	7.73	6.50
Street noisy environment	0dB	4.50	5.07
	5dB	6.41	5.47
	10dB	7.01	6.71
	15dB	8.72	7.63

Table 4: Weighted Spectral Slope (WSS) measure experimental results

	Noise level	TFCR noise estimation	MCRA noise estimation
Airport noisy environment	0dB	66.74	67.50
	5dB	61.04	65.81
	10dB	45.53	51.32
	15dB	30.17	43.53
Exhibition noisy environment	0dB	65.63	70.55
	5dB	61.62	69.75
	10dB	50.09	52.10
	15dB	42.78	45.32
Street noisy environment	0dB	60.31	65.58
	5dB	51.97	61.85
	10dB	48.25	53.72
	15dB	44.37	48.62

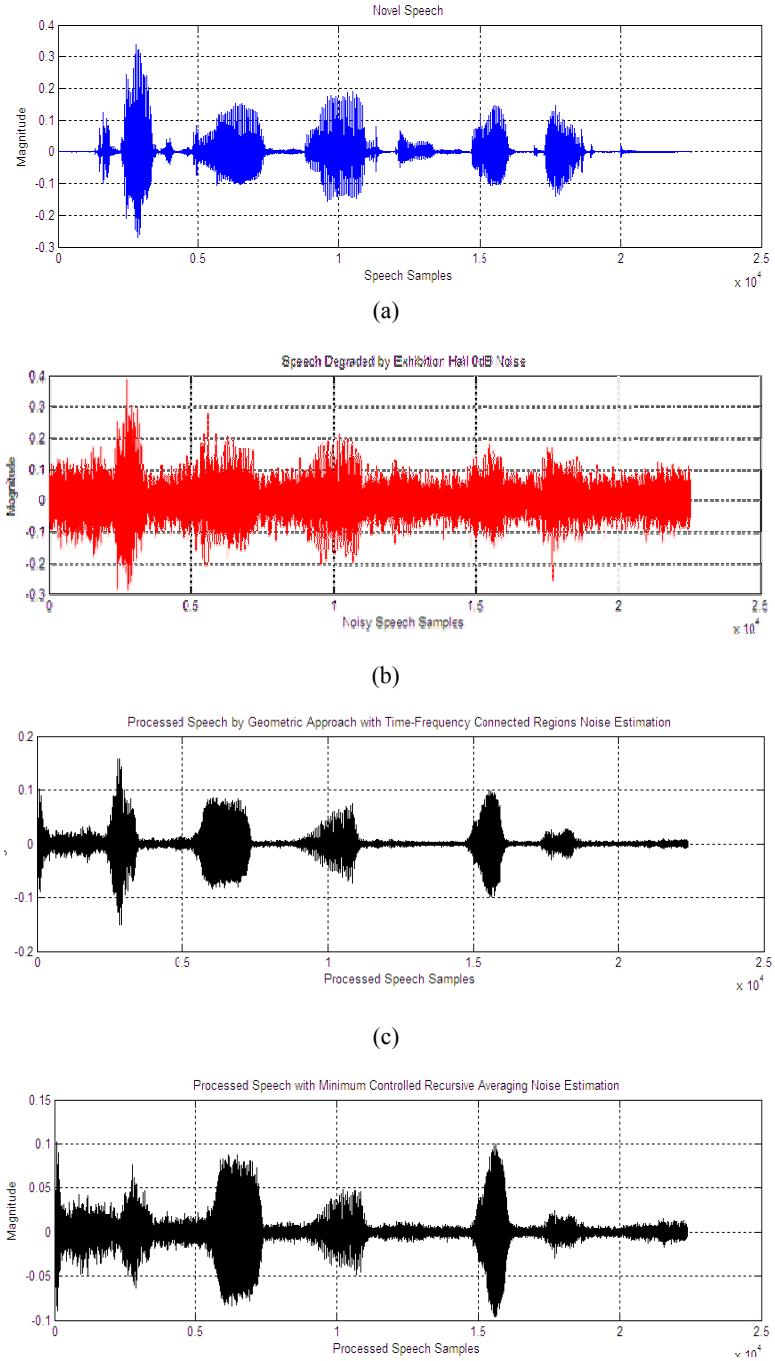

Fig. 6: Time-domain waveforms for clean speech (BLUE), exhibition Hall 0dB noisy speech (RED), TFCR noise estimation enhanced speech and MCRA noise estimation enhanced speech

spectral domain. It is based on the comparison of smoothed and distorted spectra of novel and degraded speech signal respectively. The smaller distance measured means better quality of speech and vice versa. Figure 6 shows the time-domain waveform analysis of novel, noisy speech and enhanced speech with two different noise estimations. Similarly the Fig. 7 shows the spectrum of clean, noisy and enhanced speech signals.

CONCLUSION

In present study Geometric approach of spectral subtraction is implemented with time-frequency

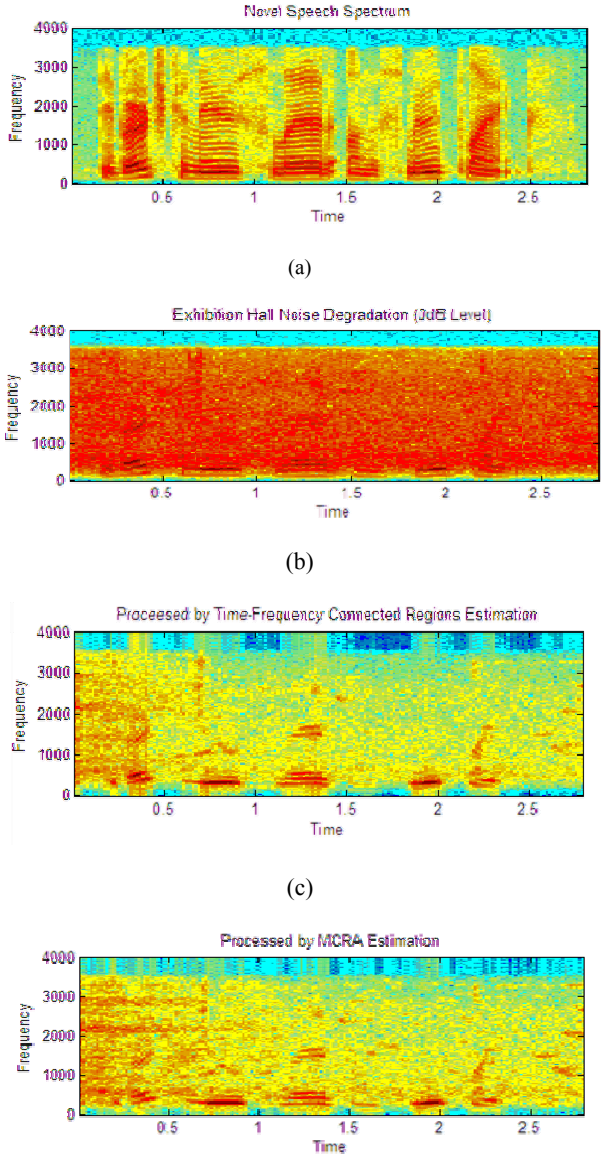

Fig. 7: Spectra of novel, noisy and processed speech by TFCR and MCRA

connected speech region noise estimation. In Contrast with conventional power spectral subtraction method, where the cross terms are assumed to be zero which results in cross terms errors, a new gain function in Geometrical approach has been developed which will always be positive and real. This GA method is supported by TFCR noise estimation for noise spectrum estimation purpose, whose temporal minimum values are computed from smoothed periodogram which are further utilized in speech presence activity. This combination of spectral subtraction with noise estimation is evaluated for quality in three different real world noisy environments with PESQ, FwSNRseg and WSS. The experimental results show that time-frequency noise estimation when combined with GA approach performs better than minimum controlled recursive averaging noise estimation.

REFERENCES

Berouti, M., M. Schwartz and J. Makhoul, 1979. Enhancement of speech corrupted by acoustic noise. Proceeding IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 208-211.

Bernard, G., L. Johan, B. Radu and R. Justinian, 2005. Performance Assessment Method for Speech Enhancement Systems. SPS-DARTS 2005, Antwerp, Belgium.

Boll, S.F., 1979. Suppression of acoustic Noise in speech using spectral subtraction. IEEE T. Acoustic. Speech Signal Process., 27(2): 113-120.

Hu, Y. and P. Loizou, 2007. Subjective Evaluation and comparison of speech enhancement algorithms. Speech Commun., 49: 588-601.

ITU, 2000. Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs. ITU-T Recommendation, pp: 862.

Jianfen, M., H. Yi ans C.L. Philipos, 2009. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Am., 125(5): 3387-3405.

Kamath, S. and P. Loizou, 2002. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing.

Karsten, V.S. and V.A. Søren, 2005. Speech enhancement with natural sounding residual noise based on connected time-frequency speech presence regions. EURASIP J. Appl. Signal Proc., 18: 2954-2964.

Loizou, P., 2007. Speech Enhancement: Theory and Practice. CRC Press LLC, Boca Raton, FL.

Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE T. Speech Audio Process., 9(5): 504-512.

Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. IEEE T. Speech Audio Process., 7(3): 126-137.

Yang, L. and C.L. Philipos, 2008. A geometric approach to spectral subtraction. Speech Commun., 50: 453-466.