Research Journal of Applied Sciences, Engineering and Technology 7(2): 296-302, 2014
DOI:10.19026/rjaset.7.254
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: April 05, 2013 Accepted: April 29, 2013 Published: January 10, 2014

Research Article

Speech Intelligibility Prediction Intended for State-of-the-Art Noise Estimation Algorithms

[1]Nasir Saleem, [2]Sher Ali, [1]Ehtasham Mustafa and [3]Usman Khan
[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 Campus, Kohat, KPK, Pakistan

Abstract: Noise estimation is critical factor of any speech enhancement system. In presence of additive non-stationary background noise, it is difficult to understand speech for normal hearing particularly for hearing impaired person. The background interfering noise reduces the intelligibility and perceptual quality of speech. Speech enhancement with various noise estimation techniques attempts to minimize the interfering components and enhance the intelligibility and perceptual aspects of damaged speech. This study addresses the selection of right noise estimation algorithm in speech enhancement system for intelligent hearing. A noisy environment of airport is considered. The clean speech is corrupted by noisy environment for different noise levels ranging from 0 to 15 dB. Six diverse noise estimation algorithms are selected to estimate the noise including Minimum Controlled Recursive Average (MCRA), MCRA-2, improved MCRA, Martin minimum tracking, continuous spectral minimum tracking, and weighted spectral average. Spectral subtraction algorithm is used for enhancing the noisy speech. The intelligibility of enhanced speech is assessed by the fractional Articulation Index (*f*AI) and SNR_{LOSS}.

Keywords: fAI, IMCRA, MCRA, MCRA-2, noise estimate, SNR_{LOSS}, spectral subtraction

INTRODUCTION

The requirement of speech signals enhancement rises in situations where speech is originated in the noisy environments. These additive components degrade the perceptual aspects and intelligibility of speech. In any speech enhancement and recognition system noise estimation is key component. The robustness of speech enhancement system is greatly affected under the conditions where the noise level is low and variation in noise levels. In this study particular emphasis is given on the importance of accurate noise estimation technique. Noise estimation technique that based on Voice Activity Detection (VAD) is restricted in absence of speech. The reliability of this method is severely declines in presence of weak speech signals and inputs SNR (Sohn *et al*., 1999; Meyer *et al*., 1997). Another method based on histogram in power spectral domain (Ris and Dupont, 2001) which is computationally more expensive and requires more memory. Also this method does not perform well in low noise level conditions. Here seven different noise estimation schemes are selected to estimate the noise spectrum. These schemes include including Minimum Controlled Recursive Average (MCRA), MCRA-2, improved MCRA, Martin minimum tracking, continuous spectral minimum tracking, weighted spectral average and connected frequency region. All these estimation methods are integrated with spectral subtraction of speech enhancement to increase the intelligibility and perceptual quality of noise affected speech. The main aim of this study is to select the appropriate noise estimation algorithm provides high intelligibility in noisy conditions.

METHOD: SPEECH ENHANCEMENT BY SPECTRAL SUBTRACTION

Spectral subtraction is one of the most widespread and modest technique of minimizing the background additive noise (Hu *et al*., 2001). In many applications where noise is accessible on dispersed channel, in this case it is possible to retrieve novel signal by subtracting the estimate of noise from noisy signal. However in some applications, like in receiver of noisy communication channel, only noisy signal is accessible. In this condition it is impossible to cancel out the random noise but there is possibility to reduce the average effect of noise on speech signal spectrum. The additive noise increases the variance and mean of the magnitude spectrum of speech signals. The increase in variance is because of noise fluctuations and cannot be cancelled out. The increase in mean can be eliminated by estimating the noise spectrum from noisy speech spectrum. The noisy signal can be modelled in time-domain as:

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

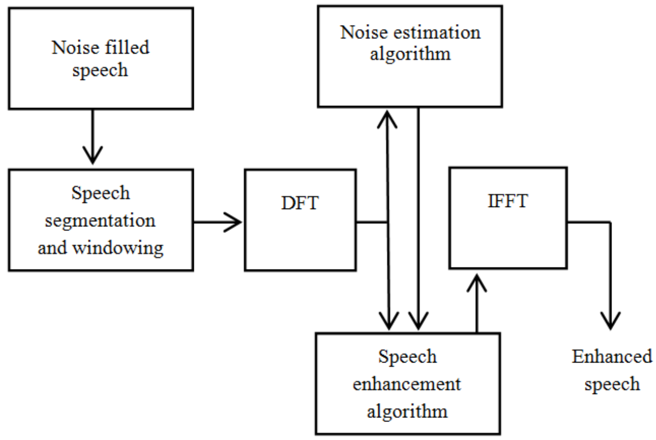

Fig. 1: Speech enhancement system

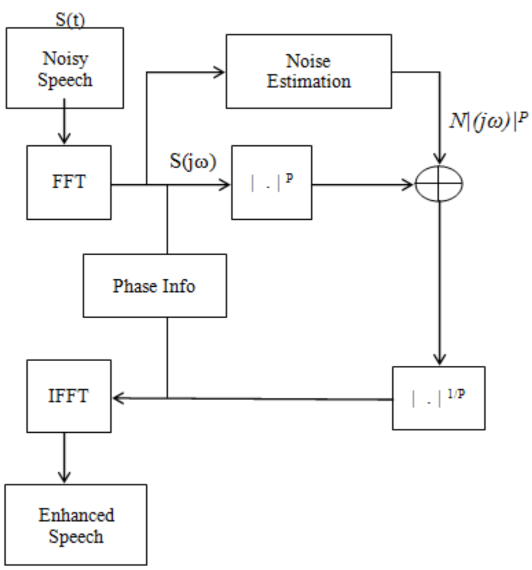

Fig. 2: Spectral subtraction algorithm

$$S(t) = x(t) + n(t) \tag{1}$$

x (t) = Clean speech
n (t) = Additive background noise

By taking the Fourier Transform:

$$S(j\omega) = X(j\omega) + N(j\omega) \tag{2}$$

In spectral subtraction technique of speech enhancement, the incoming speech signals *x* (*t*) are buffered and divided into small frames with *N* samples length. Every segment is windowed with either Hamming or Hanning window. This windowed segment is transformed to frequency-domain having *N* number of samples with Discrete Fourier Transform (DFT). The aim of the windowing is to minimize the effect of discontinuities at the endpoints of every segment. The windowed speech segments are given in time-domain as:

$$S_W(t) = W(t)\, S(t) \tag{3}$$

$$S_W(t) = W(t)\, [x(t) + n(t)] \tag{4}$$

$$S_W(t) = x_W(t) + n_W(t) \tag{5}$$

The windowing operation is expressed in frequency-domain as:

$$S_W(j\omega) = W(j\omega) * S(j\omega) \tag{6}$$

$$S_W(j\omega) = X_W(j\omega) + N_W(j\omega) \tag{7}$$

From Eq. (2), *S* (*jω*) can be expressed in magnitude and phase as:

$$S(j\omega) = S|(j\omega)|e^{j\phi} \tag{8}$$

The noise spectrum can be expressed as:

$$N(j\omega) = N|(j\omega)|e^{j\phi} \tag{9}$$

The enhanced speech estimation can be done by subtracting the noise spectrum from noisy spectrum as:

$$X_E(j\omega) = [S|(j\omega)| - N|(j\omega)|]\, e^{j\phi} \tag{10}$$

Finally the enhanced speech is obtained by taking the Inverse Fourier Transform. The block diagrams for speech enhancement and spectral subtraction way is sketched in Fig. 1 and 2, respectively.

Noise estimation: Noise estimation is the fundamental phase in speech enhancement algorithm.

Noise estimation is persistently executed in spectral domain or associated domain for many reasons. These reasons include:

- Speech and noise are moderately separated in spectral domain

- Spectral components of both noise and speech are partially de-correlated
- Psycho-acoustic representations are suitably applied in this domain

Because of these reasons following domains are used, all have diverse advantages:

- Complex spectral amplitudes
- Spectral magnitudes
- Spectral powers
- Log spectral powers
- Mel-spaced spectral amplitudes
- Mel-spaced log spectral powers
- Mel cepstral coefficients
- AR coefficients

In each of these domains, the coefficients are most regularly involved to be Gaussian or uncorrelated.

Classification of the noise estimation algorithms: The classification of the Noise estimation algorithms is:

- Minimal Tracking algorithms
- Time recursive algorithms
- Histogram based algorithms

Minimal tracking algorithm: This class of noise estimation is based on assumption that the power of noised masked speech in individual frequency bins often declines to the power level of noise even during speech activity (Loizou, 2007). Therefore; by tracking the minimum noisy speech of power in every frequency bin, the rough estimate of noise level in any bin can be discovered.

Minimum statistic noise estimation: The MS algorithm was proposed by Martin R. and developed later on (Martin, 2001) by including bias compensation and improved smoothing factor. It is assumed that in any frequency band, there may condition when the speech signal energy will be little and this energy is controlled by the noise. So if it happens once per time period T, the noise power can be estimated as minimum power elevated in T typically within 0.5 to 1.5 sec. Similar methodology was adopted by Doblinger (1995) but instead of taking the minimum over the T, the noise spectrum is smoothed by using two diverse time constants, that is, short time constant and long-time constant. The short-time constant is used in condition when the energy in frequency band is decreasing to guarantee quick adaptation to new minimum whereas the long-time constant is used while power increases to avoid adaptation to speech energy. The use of the minimum makes the method extradelicateto outliers and investigated the use of other quantities, medians, gives better results (Stahl *et al*., 2000). Some people follow up this approach but concluded that it performs unwell in non-stationary noise (Manohar and Rao, 2006).

Continuous spectral minimum: In MS algorithm where the minimum tracking is deployed, this method has disadvantage, it is incapable to respond in conditions when noise spectrum is changing rapidly (Loizou, 2007). Instead of using a fixed window for tracing the minimum of noisy speech (Martin, 2001), the noise estimate is updated constantly by smoothing the noisy speech power spectra in each frequency band using a non-linear smoothing rule. In order to track minimum of noise spectrum, a short-time smoothed version of period gram of noisy speech is calculated before using the following equation:

$$P(n, j) = \gamma P(n-1, j) + (1 - \gamma)\,|S(n, j)|^2 \qquad (11)$$

where,

$$|S(n, j)|^2 = |X(n, j)|^2 + |N(n, j)|^2$$

The $|S(n, j)|^2$ is the period gram of noisy speech, $|X(n, j)|^2$ is period gram of clean speech and $|N(n, j)|^2$ is period gram of noise signal respectively. The *"γ"* represents the smoothing factor ranging from 0.7 to 0.9 and *"n"* shows the frame index while *"j"* represents the frequency band index. The non-linear rule for estimating the minimum of noise spectrum power $P_{min}(n, j)$ is used in every frequency band.

Time recursive averaging: The time recursive averaging algorithm explains the fact that noise has non-uniform effect on speech spectrum as different frequency band in spectrum have different effective SNR. For any category of noise, one can estimate and update distinct frequency bins of noise spectrum once speech presence probability is inattentive at certain frequency bin is high or the effective SNR at specific frequency band is really low. These observations commanded to the new class of noise estimation algorithms where the noise spectrum estimation is performed as weighted average of past noise estimate and present noisy speech spectrum. The weights change adaptively, depend either on effective SNR of every frequency bin or on speech presence probability. MCRA (Cohen, 2002), MCRA-2 (Loizou *et al*., 2004) improved MCRA (Loizou and Sundarajan, 2006) came under this class.

Histogram based noise estimation: Mathematically, the histogram is function βi which compute the number of observations that falls in each bin. Let Ω represents the total no. of observations and γ shows the no. of bins, the histogram will be:

$$\Omega = \sum_{i=1}^{\gamma} \beta i \qquad (12)$$

This class of noise estimation are based on most frequent values of energy in individual bands corresponds to noise level in specific frequency band that is histogram represents the noise levels observed. McAulay and Malpass (1980) proposed an idea, attributed to Roberts (1978) that is constructed on bimodality of the signal energy histogram reserved over 4-second window. The algorithm decides adaptive energy threshold which selects the existence of speech and also consist of upper and lower thresholds. The adaptive threshold is selected to lie at the 80th percentile of the histogram of energies that are underneath the upper fixed threshold. This approach is modified in (Compernolle, 1989). Similar approach is used by Hirsch and Ehrlicher (1995) where adaptive threshold is used in every frequency band to eliminate speech frames and the peak of histogram of recent noise frame is used as an estimate of noise power in that band. The accuracy of this approach is greater than VAD. The highest peak in histogram represents the noise where lower magnitudes in histogram show the speech.

Speech intelligibility: Speech intelligibility is specified by a number which shows how correctly a speech is understood by a listener in specific situation and expressed with simple mathematical expression as:

$$\text{Speech intelligibility (SI)} = 100R/T \qquad (13)$$

T = No. of speech units in test
R = No. of correct speech units

Many measures have been proposed in order to predict the intelligibility of speech in presence of background interfering noise. AI French and Steinberg (1947), Fletcher and Galt (1950), Kryter (1962a, b) and ANSI (1997) and STI (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985) are two speech intelligibility predicting tools that are used in noisy conditions. The AI is upgraded SII (ANSI, 1997). The SII effectively predicts the effects of linear filtering and additive noise on speech intelligibility but still have number of limitations. One of them is, the SII prediction is confirmed for typically stationary masking noise as it is based on long-term averaging spectra, proposed over 125 m sec pauses, of speech and masker signals, respectively. Therefore; cannot be applied to situation where the maskers are fluctuating that is competing talkers. Many attempts are carried out to predict the indelibility of speech with SII in fluctuating environments (Rhebergen and Versfeld, 2005; Rhebergen *et al*., 2006; Kates, 1992) where the speech and maskers signals are dividing into small intervals (9-20 m sec) and compute the AI in every band and then computing the overall AI by averaging the individual band AI to produce single AI value. Other additions to SII index were suggested in Kates and Arehart (2005) and Kates (1992) for predicting intelligibility of peak-clipping and center-clipping distortions in speech signal.

Experiment-1: speech intelligibility prediction using $\mathbf{SNR_{LOSS}}$: SNR_{LOSS} Jianfen and Philipos (2011) is critical-band spectral representation of clean and noise-suppressed speech signals that predict SNR_{LOSS} (intelligibility predicting measure) in every critical band. This technique provides diverse weight to the spectral amplification and attenuation distortion introduced by speech enhancement algorithm.

Speech processing: Let $S(n) = x(n) + n(n)$ shows the noisy signal where $x(n)$ represents the clean speech signal and $n(n)$ indicates the masker signal. The Hamming windowed signal with Hamming function $h(n)$, the STFT of noisy signal $S(n)$ is:

$$S(\omega_k, i) = \sum_{n=0}^{N-1} y(iR + n).h(n)\, e^{\omega kn} \qquad (14)$$

$\omega_k = 2\pi k/N$, where $k = 0, 1, 2\ldots.., N-1$, the frequency bands index, i is frame index, N is frame size (no. of samples in each frame) and R is update rate in samples. The critical-band spectrum of $S(n)$ is calculate by multiplying FFT magnitude spectra $|S(\omega_k, i)|$ by 25 overlapping Gaussian-shaped windows (Philipos, 2013) spaced in fraction to ear's critical bins and summing up powers within critical bin which results in critical-band spectra illustration of signal as:

$$S(j, i) = X(j, i) + N(j, i)\; j = 1, 2, .., K \qquad (15)$$

$X(j, i)$ is excitation spectrum of noise-free speech in j band and $N(j, i)$ is excitation spectrum of noise (masker) in j band. The SNR loss in j band and i frame can be computed as:

$$\text{Loss}(j, i) = \text{SNRX}(j, i) - \text{SNR}\hat{X}(j, i)$$

$$\text{Loss}(j, i) = 10\log_{10}\frac{X(j,i)2}{N(j,i)2} - 10\log_{10}\frac{\hat{X}(j,i)2}{N(j,i)2}$$

$$\text{Loss}(j, i) = 10\log_{10}\frac{X(j,i)2}{\hat{X}(j,i)2} \qquad (16)$$

$SNRX(j, i)$ is effective input speech SNR in j band while $SNR\,\hat{X}(j, i)$ is effective SNR of enhanced speech in j band respectively. And $\hat{X}(j,i)$ is excitation spectrum of the enhanced speech in j band.

RESULTS AND DISCUSSION

SNR_{LOSS} is computed when noisy signal passes through a noise-suppression system where noise spectrum is estimated using diverse noise estimation algorithms.

From equation 16, it can be concluded that if $X(j,i)$ becomes equal to $\hat{X}(j,i)$, value of SNR_{LOSS} will be zero which means there is no loss of speech contents and have high intelligibility. It is realistic that when SNR level is amplified i.e., $SNR \rightarrow \infty$, estimated spectra $\hat{X}(j,i)$, approachesthe cleanspectra $X(j,i)$ i.e., $\hat{X}(j,i) \rightarrow X(j,i)$, results in zero SNR_{LOSS}. The range for SNR_{LOSS} is; $0 \leq SNR_{LOSS} \leq 1$. This is tested in present study for measuring the speech intelligibility scores. Table 1 shows intelligibility scores for different noise estimation algorithms.

Experiment-2: speech intelligibility prediction using fractional articulation index: The conventional Articulation Index (AI) cannot be utilized in conditions where noise is additive and also the non-linear operations are taken in account. Because during non-linear operations, the speech and noise signals become ambiguous and as a result both speech and masker (noise) signals are affected. A new method of predicting the speech intelligibility is introduced, fractional articulation index (Philipos, 2011) where innovative effective SNR is achieved ensuing non-linear processing. The *f* AI scheme is used to specific band in situations where the non-linear processing mainly affects the target signal rather than masker signal. The *f* AI can be expressed as:

$$f_{AI} = \frac{1}{\sum_{i=1}^{N} Wi} \sum_{i=1}^{N} Wi \times fSNR_i \quad (17)$$

Wi represents the Weighting Functions (Band Importance Functions) applied to *i* band and *N* is the number of bands used. The $fSNR_i$ symbolizes portion of input SNR communicated by noise suppression algorithm and can be write as:

$$fSNR_i = \text{Min } \overline{SNR\imath}, SNR_i \quad (18)$$

if $SNR_i \geq SNR_L$

SNR_L shows the smallest allowable value of SNR and $\overline{SNR\imath}$ shows actual (true) or new SNR definition in *i* band and can be written as:

$$\overline{SNR\imath} = \hat{S}_{\cdot i}^{2} / N_i^2 \quad (19)$$

$\hat{S}_{\cdot i}^{2}$: Enhanced envelop
N_i^2 : Masker envelop

Signal processing: The speech signal is first fragmented with the help of 50 m sec Hamming window having 75% overlapping among the touching frames. The critical-band spectra of masker (before mixing), target and enhanced speech signals are calculated for each 50 m sec frame by multiplying FFT magnitude spectra by 25 overlapping Gaussian-shaped windows (Philipos, 2013) spaced in fraction to ear's critical bins and summing up powers within critical bin. Centre frequencies of bands are used to analyze *f*AI in each frame and by averaging individual bands we obtained the single *f*AI value.

Discussion: The *f*AI measure based on weighted average of fraction of input SNR sent by the speech enhancement algorithm in every band. In old-fashioned AI measure, the *fSNRi* are replaced with audibility functions that range from 0 to 1 represents the fraction of information present in speech sentence which is audible for the listener. Table 2 shows the *f*AI values computed over sentence processed by spectral subtraction algorithm using six different noise estimation algorithms. The speech sentence is corrupted by noise level ranging from 0 to 15 dB airport noise. From results it is deducted that with the increase in noise level, the speech intelligibility decreases and vice versa.

Table 1: SNR_{LOSS} intelligibility prediction values (round-off to three decimal points) for different noise estimation algorithm

		Noise estimation algorithm					
Noise environment	Noise levels (dB)	MARTIN	MCRA	MCRA 2	IMCRA	DOBLINGER	HIRSCH
Airport noise	0	0.921	0.893	0.854	0.895	0.900	0.892
	5	0.832	0.812	0.801	0.840	0.810	0.820
	10	0.723	0.691	0.702	0.702	0.705	0.711
	15	0.695	0.683	0.697	0.694	0.703	0.704

Table 2: *f*AI intelligibility prediction values (round-off to four decimal points) for different noise estimation algorithms

		Noise estimation algorithm					
Noise environment	Noise levels (dB)	MARTIN	MCRA	MCRA 2	IMCRA	DOBLINGER	HIRSCH
Airport noise	0	0.2177	0.1952	0.2098	0.1696	0.2139	0.2044
	5	0.4884	0.4758	0.4807	0.3438	0.4853	0.4850
	10	0.6416	0.6335	0.6344	0.6221	0.6140	0.6372
	15	0.6465	0.6432	0.6421	0.6442	0.6257	0.6482

CONCLUSION

In this study, we have addressed the significance of accurate noise estimation system in speech enhancement algorithms and also addressed the intelligibility prediction of noise estimation algorithms. Six diverse noise estimation algorithms are used in this study. The enhanced speech signals are evaluated for speech intelligibility using two modern intelligibility prediction methods, that is, SNR_{LOSS} which is the critical-band spectral representation of clean and noise-suppressed speech signals that predict SNR_{LOSS} intelligibility predicting measure in every critical band) and *f*AI where innovative effective SNR is achieved ensuing non-linear processing. The *f*AI scheme is used to specific band in situations where the non-linear processing mainly affects the target signal rather than masker signal. The results of both prediction methods indicate that if $S\underline{NR}\rightarrow\infty$, the estimated noise spectra approaches the clean spectra and the speech intelligibility increases.

REFERENCES

ANSI, 1997. Methods for calculation of the speech intelligibility index. Technical Report No, S3.5-1997, American National Standards Institute.

Cohen, I., 2002. Noise estimation by minima controlledrecursive averaging for robust speech enhancement. IEEE Signal Proc. Lett., 9(1): 12-15.

Compernolle, D.V., 1989. Noise adaptation in hidden markov model speech recognition. Syst. Comput. Speech Language, 3: 151-167.

Doblinger, G., 1995. Computationally efficient speech enhancement by spectral minimatracking in subbands. Proc. Euro Speech, 2: 1513-1516.

Fletcher, H. and R.H. Galt, 1950. The perception of Speech and its relation totelephony. J. Acoust. Soc. Amer., 22: 89-151.

French, N.R. and J.C. Steinberg, 1947. Factors governing the intelligibility of speech sounds. J. Acoustic. Soc. Amer., 19: 90-119.

Hirsch, H.G. and C. Ehrlicher, 1995. Noise estimation techniques for robust speech recognition. Processing of the International Conference on Acoustics, Speech and Signal ICASSP-95, 1: 153-156.

Houtgast, T. and H. Steeneken, 1985. A review of the MTF concept in roomacoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Amer., 77: 1069-1077.

Hu, Y., M. Bhatnager and P. Loizou, 2001. A cross correlation technique for enhancing speech corrupted with correlated noise. Proceedings of the (ICASSP '01) IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 673-676.

Jianfen, M. and C.L. Philipos, 2011. SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech. Speech Commun., 53(3): 340-354.

Kates, J., 1992. On using coherence to Measure distortion in hearing aids. J. Acoust. Soc. Amer., 91: 2236-2244.

Kates, J. and K. Arehart, 2005. Coherence and the speech intelligibility index. J. Acoust. Soc. Amer., 117: 2224-2237.

Kryter, K., 1962a. Methods for calculation and use of the articulationIndex. J. Acoust. Soc. Amer., 34(11): 1689-1697.

Kryter, K., 1962b. Validation of the articulation index. J. Acoust. Soc. Amer., 34: 1698-1706.

Loizou, P.C., 2007. Speech Enhancement: Theory and Practice. 1st Edn., FL: CRC Press, Boca Raton.

Loizou, P. and R. Sundarajan, 2006. A noise estimation algorithm for highly non-stationary environments. Speech Commun. Sci. Direct, 48: 220-231.

Loizou, P., R. Sundarajan and Y. Hu, 2004. Noise estimation algorithm with rapid adaption for highly non-stationary environments. Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP '04), pp: 1-305-8.

Manohar, K. and P. Rao, 2006. Speech enhancement in nonstationary noise environments using noise properties. Speech Commun., 48: 96-109.

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

McAulay, R. and M. Malpass, 1980. IEEE transactions acoustics, speech and signal processing. IEEE J. Mag., 28(2): 137-145.

Meyer, J., K.U. Simmer and K.D. Kammeyer, 1997. Comparison of one- and two-channel noiseestimation techniques. Proceeding of the 5th International Workshop on Acoustic Echo and Noise Control, IWAENC-97, London, UK, pp: 137-145.

Philipos, C.L., 2011. Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms. J. Acoust. Soc. Am. 130 (2): 986-995.

Rhebergen, K.S. and N.J. Versfeld, 2005. A speech intelligibility index based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. J. Acoust. Soc. Am., 117: 2181-2192.

Rhebergen, K.S., N.J. Versfeld and W. Dreschler, 2006. Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. J. Acoust. Soc. Am., 120: 3988-3997.

Ris, C. and S. Dupont, 2001. Assessing local noise level estimation methods: Application to noise robust ASR. Speech Commun., 34(1): 141-158.

Roberts, J., 1978. Modification to Piecewise LPC-10E. Retrieved form: http://www.clear-labs.com/tutorial/refer01.php.

Sohn, J., N.S. Kim and W. Sung, 1999. A Statistical model-based voice activity detector. IEEE Signal Process. Lett., 6(1): 1-3.

Stahl, V., A. Fischer and R. Bippus, 2000. Quantile based noise estimation for spectral subtraction and wiener filtering. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '00, 3: 1875-1878.

Steeneken, H. and T. Houtgast, 1980. Physical method for measuring speech transmission quality. J. Acoust. Soc. Am., 67: 318-326.