Performance Evaluation of Objective Quality Measures for Speech Enhancement

G. Amjad Khan*, Dr. K. E. Sreenivasa Murthy**
*Research Scholar [Regd No:PP.ECE.0019], Rayalaseema University,Kurnool,Andhra Pradesh, India
**Professor & HOD of ECE, G. Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India

Abstract—In this paper, we evaluate the performance of several objective measures in terms of predicting the quality of noisy speech enhanced by noise suppression algorithms. The objective measures considered a wide range of distortions introduced by four types of real-world noise at three signal-to-noise ratio levels by three classes of speech enhancement algorithms: optimally modified log spectral amplitude, Improved Minima controlled recursive averaging and sub band adaptive filtering algorithms. The objective quality ratings were obtained using the methodology designed to evaluate the quality of enhanced speech along three dimensions: signal distortion, noise distortion, and overall quality. This paper reports on the evaluation of correlations of several objective measures with these three subjective rating scales. Several new composite objective measures are also proposed by combining the individual objective measures using sub band adaptive filtering technique.

Keywords: sub band adaptive filtering, speech enhancement .overall quality

1. Introduction

In recent years, due to the rapid growth in the speech oriented applications including Automatic Speech Recognition (ASR), modern mobile communications, hands free telephony and human-computer interaction systems through voice communications, noise suppression for speech enhancement has become more essential component. All these speech based applications becomes ineffective in the presence of a speech signal with unnecessary disturbances. For example, in the hands free telephony speech communication systems, the microphones are typically placed at certain distant from the speaker's mouth. In such cases, various noise sources makes the speech signal corrupted, by which the speech oriented devices are not able to process the signal effectively. Generally the speech enhancement techniques consists of noise or disturbance estimation and filtering algorithms by which the speech quality and intelligibility increases significantly. However the complete elimination of noises form a noise contaminated speech signal results in the loss of speech related information also. Hence there is need to design an efficient noise suppression algorithm for speech enhancement approach with less information loss followed by greater noise removal.

Several approaches have been developed in earlier to perform noise suppression in speech signals and basically they are divided as two types, they are subspace methods or time domain methods and frequency domain methods. In the case of subspace methods, the noise suppression is applied directly over the speech signals whereas in the transform domain, the noise suppression is applied over the transformed values of speech signal. Both the methods have their own advantages and disadvantages. For instance the subspace methods are simple but the information loss is observed to be high due to the direct manipulations over the samples of speech signal. On the other hand, the transform domain approaches are more efficient but they consume high computational resources. Compared to the subspace methods, transform domain approaches are more efficient tin the preservation of speech related information followed by effective noise suppression. Since there are so many Transform techniques like Discrete Cosine Transform (DCT), Discrete Fourier transform (DFT), Short Time Fourier Transform (STFT), Discrete Wavelet Transform (DWT), selection of an appropriate transform technique is an important aspect to find the exact noise distribution in every sample of the speech signal. Compared to the DCT, DFT and STFT, DWT is more advantageous because the DWT represents the signal in the multi-resolution fashion by which there is a possibility to obtain a perfect discrimination between the noise and speech sample at each and every resolution. Since the pitch frequency varies from human to human and male to female and also depends on the age factor, the DWT is more appropriate in the provision of a clear distinction between the frequencies of different constraints.

Sub band Adaptive filtering (SAF) is a novel filtering concept for which there is so many applications, including noise suppression, feature extraction is data mining applications, signal quality preservation, etc. First, the sub band decomposition splits the full band signal into multiple sub band signals that allows for the processing of information in each sub band independently. In the case of SAF accomplishment over noise suppression in speech signals, initially the noisy speech signal is decomposed into several subband signals. Then the adaptive filtering can be applied on each sub band signal. The subband decomposition is achieved via a set of filters called filter bank. Since each subband contains a fraction of the original signal, the subband signal can be down sampled by the number of subbands while still preserving the information of original sample. For reconstruction of the original signals, the down sampled sub band signals are up sampled by the same rate as the decimation rate and are combined by a synthesis filter bank. Since the SAF tunes its parameters according to the characteristics of input signal, combining a SAF with noise suppression in speech signals can obtain an optimal quality in the speech signal after noise removal.

2. Literature Survey

Compared to the speech coding literature [1], only a small number of studies examined the correlation between objective measures and the subjective quality of noise-suppressed speech [12]–[17]. Salmela and Mattila [13] evaluated the correlation of a composite measure with the subjective (overall) quality of noise-suppressed speech. The composite measure consisted of 16 different objective measures which included, among others, spectral distance measures, LPC measures (e.g., Itakura–Saito) and time-domain measures [e.g., segmental signal-to-noise ratio (SNR)]. The noisy speech samples were not processed by real enhancement algorithms, but rather by ideal noise-suppression algorithms designed to provide controlled attenuation to the background alone or to both background and speech signals. The resulting composite measure produced a high correlation of 0.95 with overall quality. Rohdenburg et al. [12] evaluated the correlation of several objective measures including LPC-based measures [e.g., log-area ratio (LAR)] and the perceptual evaluation of speech quality (PESQ) measure with speech enhanced by a single algorithm. The subjective listening tests were done according to the ITU-T P.835 methodology specifically designed to evaluate the distortions and overall quality of noise suppression algorithms. Correlations ranging from 0.7 to 0.81 were obtained with ratings of background distortion, signal distortion, and overall quality. Turbin and Fluchier [14] proposed a new objective measure for predicting the background intrusiveness rating scores obtained from ITU-T P.835-based listening tests. High correlation was found with the background noise ratings using a measure that was based on loudness density comparisons and coefficient of tonality. Turbin and Fluchier later extended their work in [18] and proposed an objective measure to estimate signal distortion (but not overall quality).

To distinguish speech from noise, both the estimation of speech and the estimation of noise power spectrum have fully considered the differences between speech and noise. For example, the estimation of noise PSD is generally based on the assumption that the noise power is varying more slowly than the speech power. Besides, the differences are also made between different types of noise in the characteristic statistical properties. However, the design of the speech enhancement algorithms usually does not take the differences into consideration, which causes these algorithms to be not always optimal for various noise environments. As to improve the performance of speech enhancement, these algorithms should be adjusted to adapt to different types of noise and deal with them respectively. In other words, we can improve the performance of these speech enhancement algorithms by incorporating the noise classification into them.

3. Objectives

To overcome the above mentioned problems, this research work focused to develop a new noise suppression framework to suppress different types of noises added during the speech signal recoding. The complete objectives of this research work are outlines as follows;To improve the quality of a speech signal (means improving SNR, SegSNR, and PESQ), this work proposes to design an adaptive regularized filtering algorithm based on SAF which discriminates the noise and speech samples based on their spectral characteristics. Compared to the normal characteristics, the spectral characteristics gives more clear idea about the noise distribution over every sample in the noisy speech signal. To improve the convergence speed in the proposed SAF, this work proposes to design a Linear Correlated Band-dependent Wight factors which depends on the correlation between the bands in the hierarchical level. Since there is a possibility to exist a correlation in the frequencies of multiple bands, the correlation base weight factors will reduce the number of bands required to process. Due a single weight factor for multiple bands, the computational complexity also reduces more effectively followed by faster convergence. To achieve the robustness, the two objectives are allowed to apply over different types of noise contaminated to a single speech signal and for every case, the quality will be observed.

4. Methodology

The complete methodology of this research work is accomplished through the collection of literature survey, theoretical study and implementation. Initially the literature survey is carried out over all the earlier approaches and the problems with all those approaches are derived. Based on the observed problems, new methods are derived theoretically for noise suppression. Further the proposed methods are processed for implementation and tested over standard speech database like TIMIT, NOISEX.
According to the SSAF reported in [16], the weight vector of the original SSAF is obtained as Consider a desired signal $d(n)$ derived from the unknown noise suppression system is given by

$$d(n) = v(n) + \boldsymbol{u}^T(n)\boldsymbol{w}_0 \qquad (1)$$

Where $\boldsymbol{w}_0$ denotes a weight vector which we needs to be estimated through the adaptive filter, $\boldsymbol{u}(n)$ is an input signal vector, T denotes the transpose and $v(n)$ denotes the additive noise.
In this noise suppression system, the desired signal $d(n)$ and input signal $u(n)$ are partitioned into N sub band signals, $d_i(n)$ and $u_i(n)$, $i = 0,1,2, \dots, N-1$, through the analysis filter bank $\{H_i(z), i = 0,1,2, \dots, N-1\}$. Then the sub band signals $y_i(n)$ and $d_i(n)$ are decimated such that the decimated sub band signals are

$$w(k+1) = w(k) + \mu \frac{\boldsymbol{U}(k)sgn(\boldsymbol{e_D}(k))}{\sqrt{\sum_{i=0}^{N-1} \boldsymbol{u}_i^T(k)\boldsymbol{u_i}(k)+\epsilon}} \qquad (2)$$

Further, unlike the conventional approaches which uses a constant step size, this method switches to variable step size and the step size is varied according got the following criterion.

$$\mu(k) = \mu(k-1) + \varphi e(k)e(k-1)x^T(k-1)x(k) \qquad (3)$$

where φ is a small arbitrary constant which controls the adaption in the step size.

5. Simulation Results:

Extensive computer simulations were carried out with 10 male and 10 female utterances, randomly selected from the TIMIT database to illustrate the effectiveness of the proposed speech enhancement algorithm. Four types of noises i.e., F-16,factory-1,Jet-1,jet-2 noises are considered form the NOISEX-92 database. These noises are mixed with the clean speech signals received from the TIMIT database at various SNRs such as 0dB, 5dB and 10 dB. The acquired all clean speeches considered here is elapsed a time of 2sec. The clean speech samples are considered here for performance evaluation and are composed of 23000 samples on an average. After this, the proposed subband adaptive filtering algorithm is accomplished to perform noise suppression and finally the obtained noise free speech is processed for performance evaluation. The quality of enhanced speech using composite measures is evaluated in terms of overall quality C_{ovl}. Table I gives the comparison of composite measure for C_{ovl} of proposed method with other methods.

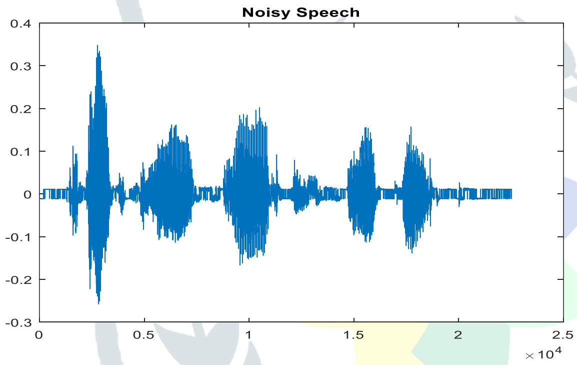

Fig. 1(a) shows the input noisy speech degraded by factory1 noise at a SNR of 0 dB

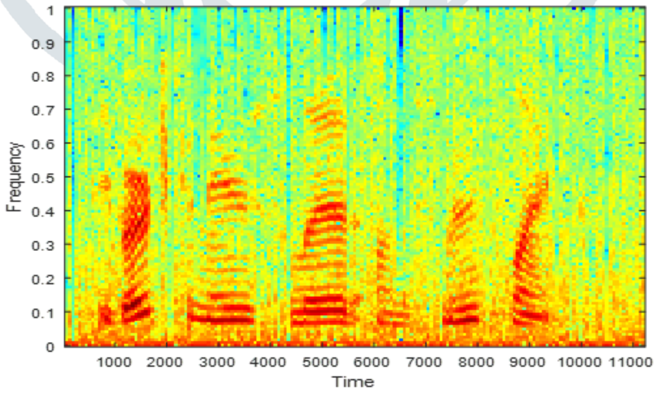

Fig. 1(b) shows the spectrogram of Fig. 1(a).

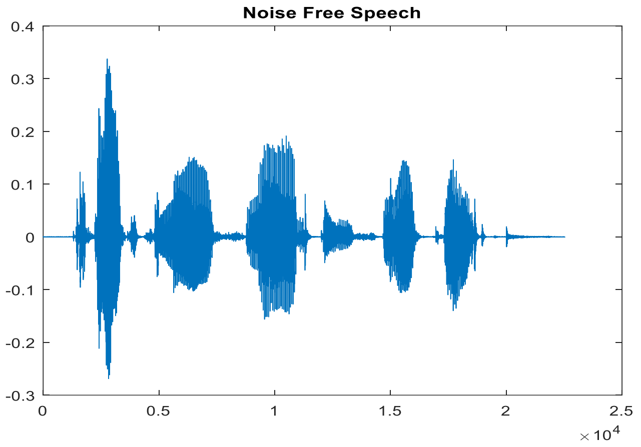

Fig. 1(c) shows the output enhanced speech at 0 dB

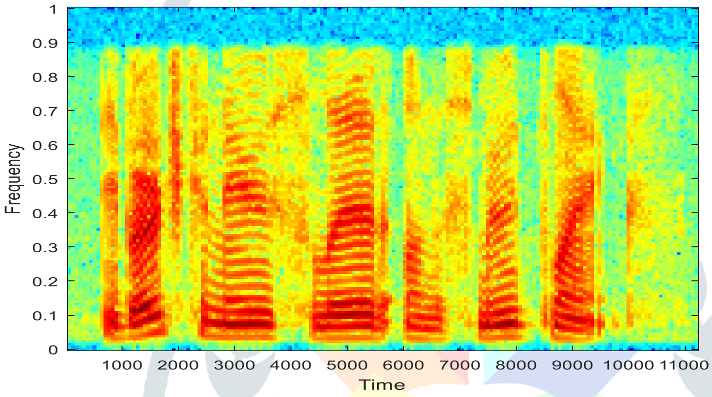

Fig. 1(d) shows the spectrogram of Fig. 1(c).

Table.I Results of composite measure for Overall Quality C_{ovl}

Noise	SNR	Unprocessed	OM-LSA	MMSE-BC	Proposed
F-16	0	1.86	2.47	2.30	2.87
	5	2.29	2.94	2.87	3.19
	10	2.79	3.50	3.36	3.68
Factory1	0	1.84	2.10	1.94	2.29
	5	2.33	2.68	2.54	2.85
	10	2.81	3.29	3.08	3.45
Jet1	0	1.71	2.22	1.79	2.44
	5	2.17	2.76	2.37	2.85
	10	2.64	3.21	2.93	3.39
Jet 2	0	1.56	1.96	1.83	2.25
	5	2.02	2.55	2.40	2.79
	10	2.49	3.05	2.98	3.30

6. REFERENCES:

[1] Lu C-T. Enhancement of single channel speech using perceptual-decision directed approach. Speech Commun 2011;53(4):495–507.

[2] Boll S. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process 1979;27(2):113–20.

[3] Cohen I, Berdugo B. Speech enhancement for non-stationary noise environments. Signal Process 2001;81(11):2403–18.

[4] Choi J-H, Chang J-H. On using acoustic environment classification for statistical model-based speech enhancement. Speech Commun 2012;54(3):477–90.

[5] Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process 1984;32(6):1109–21.

[6] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process 1985;33(2):443–5.

[7] Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans Speech Audio Process 2001;9(5):504–12.

[8] Cohen I, Berdugo B. Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Process Lett 2002;9(1):12–5.

[9] Cohen I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans Speech Audio Process 2003;11(5):466–75.

[10] Rangachari S, Loizou PC. A noise-estimation algorithm for highly nonstationary environments. Speech Commun 2006;48(2):220–31.

[11] Hendriks R, Heusdens R, Jensen J. MMSE based noise psd tracking with low complexity. In: Proc. IEEE ICASSP; 2010. p. 4266–9.

[12] Taghia J, Taghia J, Mohammadiha N, Sang J, Bouse V, Martin R. An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments. In: Proc. IEEE ICASSP; 2011. p. 4640–3.

[13] Gerkmann T, Hendriks R. Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. IEEE Trans Audio Speech Lang Process 2012;20(4):1383–93.

[15] Kates JM. Classification of background noises for hearing-aid applications. J Acoust Soc Am 1995;97(1):461–70.

[16] Alexandre E, Cuadra L, Álvarez L, Rosa-Zurera M, López-Ferreras F. Automatic sound classification for improving speech intelligibility in hearing aids using a layered structure. In: Intelligent data engineering and automated learning–IDEAL 2006. Springer; 2006. p. 306–13.

[17] Büchler M, Allegro S, Launer S, Dillier N. Sound classification in hearing aids inspired by auditory scene analysis. EURASIP J Appl Signal Process 2005;2005:2991–3002.

[18] Xiang J-J, McKinney M, Fitz K, Zhang T. Evaluation of sound classification algorithms for hearing aid applications. In: 2010 IEEE international conference on acoustics speech and signal processing (ICASSP); 2010. p. 185–8.

[19] Ma L, Milner B, Smith D. Acoustic environment classification. ACM Trans Speech Lang Process 2006;3(2):1–22.

[20] J.H. Kim, J. Kim, J.H. Jeon, and S.W. Nam, *Member, IEEE*Delayless individual-weighting-factors sign sub band adaptive filter with band-dependent variable step-sizes IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING 2016;20(4):1383–93.