ISSN (Online) 2249 - 054 X

International Journal of Computing and Corporate Research

Specialized and Refereed Journal for
Research Scholars, Academicians, Engineers and Scientists

http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

REVIEW OF SUBJECTIVE INTELLIGIBILITY COMPARISON AND EVALUATION OF SPEECH ENHANCEMENT ALGORITHMS

Abha Patyal

Final year student, M.Tech (E & C Deptt.)

MMU Mullana, Ambala, India.

Er. Anil Garg

Assistant Professor, E & C Deptt.

MMU Mullana, Ambala, India.

Abstract:

Speech enhancement algorithms consist of temporal dynamics, noise robustness, prior noise information and speaker characteristics. In this paper, analysis of all the four types of classes of algorithms has been described by a subjective framework. Different range of parameters show result as speech quality and speech intelligibility factor. In order to obtain speech only part from noisy speech signal, distortion producing processes are converted to controlling processes to make good intelligent quotient factor.

Keywords: Speech enhancement; noise removal; subjective framework; intelligent quotient; corpus speech; signal distortion; background intrusiveness.

1. Introduction:

Speech enhancement technology has been used as a platform to maximally extract the speech information from that of noisy speech signal. Four classes of algorithms are used to get corpus part but it is still vacillating as to which algorithm is more effective and in what aspect. A subjective framework is discussed here which will act as a platform for interpretation of all the algorithms in different types of noises.

The purpose is to observe a few of most speech enhancement techniques and secondly, to propose an alternative that can manage restricted scenario where processed speech get mixed with various terminals. In a noisy atmosphere, it is still difficult to check the noise priority of corpus sentence. And various approaches viz. subspace algorithm, spectral subtractive, statistical model - based algorithm and wiener filtering algorithm are considered for speech enhancement. These algorithms defined have been evaluated using a noisy speech corpus database AURORA suitable for the evaluation of the speech enhancement algorithms.

Noise Part: Subjective analysis consists of noise environments as babble (crowd of people) for 0 db & 5 db SNR.

Speech Part: 30 IEEE based sentences have been used as a base material for speech part.

Noisy Speech Part: Six persons (three men and three women) are considered as speakers for IEEE sentences comprising of different noise sources.

2. Classes of Speech Enhancement Algorithms: There are four classes of speech enhancement algorithms. Subspace algorithm, spectral subtractive,

statistical model - based algorithm and wiener filtering algorithm. All four classes' algorithms operate in the following fashion. The signal is firstly analyzed using a short time spectrum which is computed from short overlapping frames, typically of 20-30 msec. Windows with overlap (about 50%) between adjacent frames. Then analysis segment (several consecutive frames) is used in the noise spectrum computation. Typical time span of analysis segment may be 400 msec. to 1 sec. The analysis segment has to be necessarily long enough for encompassing the speech pauses and the low energy segments, but it also has to be necessarily short enough to track the fast changes in the noise level, hence the duration of the analysis segment result from track-off between these two types of restrictions. Now we will see the different classes of speech enhancement algorithms.

2.1 Subspace Algorithm:

Mittal and Phamdo proposed [1] Karhunen-Loeve transform based approach for the speech enhancement. The basic principle is decomposition of the vector space of noisy speech into speech-plus-noise subspace and a noise subspace. Enhancement is performed by removing the noise subspace and estimating the clean speech from the speech-plus-noise subspace [4]. The decomposition of noisy speech is performed by KLT. KLT & pKLT work well in both wide and narrow band signals as well as stationary and non-stationary input stochastic processes. These are defined for any finite time interval & they need high computational burden: no fast method. The KLT adapts itself to the shape of the input (signal + noise) by adopting as a reference frame.

The following steps will find the approximate KL basis [5].

- Expanding N vectors into a complete wavelet packet coefficients;

• Calculating the variance at each node and search this variance tree for the best basis;

• Sorting the best basis vector in decreasing order and select the top m best basis vectors to form a matrix U;

• Transforming N random vectors using the matrix U and diagonalizable the covariance matrix RN of these vectors to obtain the eigenvectors.

2.2 Spectral Subtractive Algorithm:

More papers describing variations of this algorithm than any other algorithm. Principle lies in assuming additive noise one can obtain an estimate of clean signal spectrum by subtracting an estimate of noise spectrum from the noisy speech spectrum. This category consists of low complexity and usually needs further enhancement. Spectral subtraction [4] is traditional method for enhancing speech degraded by additive stationary background noise in single channel system. The major drawback of this method is characteristic of the residual noise called *musical* noise. It comprises of tones of random frequencies.

$$Y(n) = x(n) + d(n)$$

$$Y(w) = X(w) + D(w)$$

$$X(w) = Y(w) - D(w)$$

OR

$$X|(w)|2=|Y(w)|2-|D(w)|2$$

Where Y(w) is the spectrum of noisy speech and D(w) is the estimated spectrum of the background noise signal.

Different types of algorithms comprising of the four classes are shown:

ALGORITHM CLASS ALGORITHM
Spectral subtractive SSUB
MBAND
RDC
Wiener-type Wiener-as
Wiener-wt
Statistical model based MMSE
MMSE-SPU
logMMSE
logMMSE-SPU-1
logMMSE-SPU-2
logMMSE-SPU-3

logMMSE-SPU-4
STSA-weuclid
STSA-wcosh
Subspace KLT
pKLT

2.**3 Statistical model based Algorithms:**

MMSE amplitude spectrum estimator; MMSE log-amplitude spectrum estimator; Non-Gaussian prior MMSE approaches being the dominant techniques because of better performance than the Spectral Subtraction methods. They need *a priori* info. of the speech and noise spectrum & derive the magnitude spectra by minimizing the mean squared error between the clean and estimated spectra (magnitude or power). This difference can be + or – . But MSE pays no attention to difference value. If + difference ~ signify attenuation distortion & if - difference ~ signify amplification distortion.

2.**4 Wiener –type Filtering Algorithm:**

Historically one of the first algorithms proposed for noise reduction. Principle lies to obtain an estimate of clean signal from that corrupted by additive noise. This estimate is obtained by minimizing the mean square error between the desired signal and the estimated signal. Drawback lies in fixed frequency response at all frequencies and requirement to estimate the power spectral density. The estimated speech signal mean mx and variance $\sigma2x$ are exploited. In case of additive type, it is assumed that the additive noise $v(n)$ is of zero mean and has a white nature with variance of $\sigma 2v$. Thus, the power spectrum $Pv(\omega)$can be approximated by:

$$Pv(\omega) = \sigma 2$$

Algorithm consists of following steps:

1) segment the noisy speech by using a 501 overlapping and a frame length of N=256 samples (32ms at *8lrHz* sampling frequency).
2) window every frame by Hanning windowing.
3) estimate the noise spectrum inside of non-speech frame**s** by means of a smoothing periodogram.
4) estimate the coefficients of the tenth-order *AR* modelling of the clean speech from the noisy speech signal.
5) design the non-causal Wiener filter from the above estimation of the speech and noise spectra.
6) filter the noisy speech frame through the previously designed Wiener filter. We consider a suitable FFT length in order to avoid aliasing effects caused by circular convolution (L=5 12 points FFr).
7) iterate until maximum number of iterations: GO TO step 4**,** by using the filtered speech signal instead of the noisy speech signal to estimate the clean speech spectrum.

3. TEST METHODOLOGY

This method instructs the listener to successively attend to and rate the enhanced speech signal on:

- SIG: the speech signal alone using a five-point scale of signal distortion (Table 1),

- BAK: the background noise alone using a five-point scale of background intrusiveness (BAK) (Table 2),

- OVRL : the overall effect using the scale of the Mean Opinion Score – [1 = bad, 2 = poor, 3 = fair, 4 = good, 5 = excellent].

Table 1 (SIG)	Table 2 (BAK)
5 – Very natural, no degradation	5 – Not noticeable
4 – Fairly natural, little degradation	4– Somewhat noticeable
3 – Somewhat natural, somewhat degraded	3 – Noticeable but not intrusive
2 – Fairly unnatural, fairly degraded	2–Fairly conspicuous, somewhat intrusive
1 – Very unnatural, very degraded	1 – Very conspicuous, very intrusive
(SPEECH SIGNAL)	(BACKGROUND NOISE)

4. STATISTICAL ANALYSIS

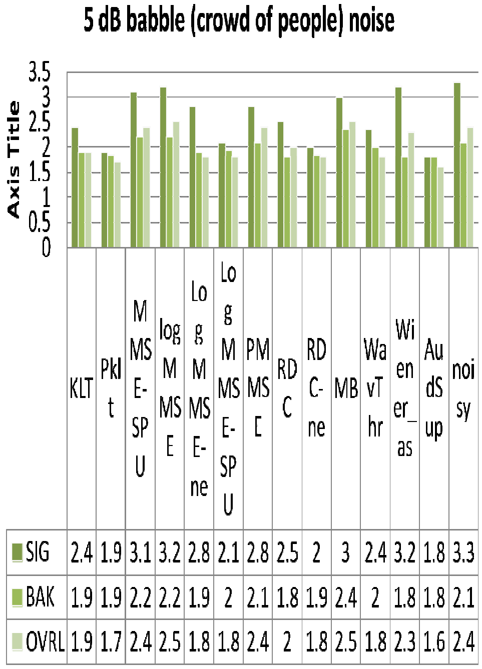
ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

Analysis of data is done in three steps:

Level 1: The performance of algorithms within each of four classes were compared to have significant difference.

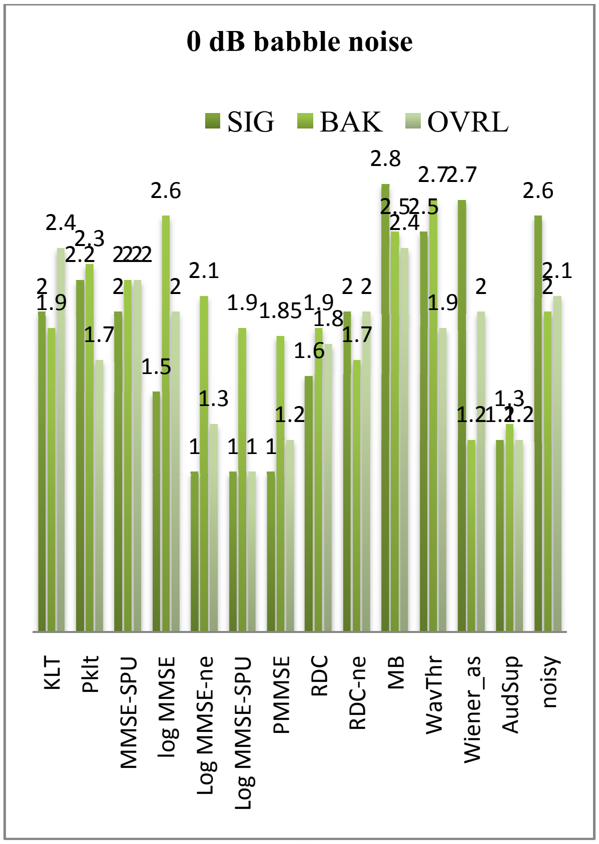
Level 2: The performance of the various algorithms across all classes aiming to find the algorithm(s) that performed the best across all noise.

Level 3: The performance of all algorithms in reference to the noisy speech (unprocessed).



5 dB babble (crowd of people) noise

	KLT	Pklt	MMSE-SPU	logMMSE	LogMMSE-ne	LogMMSE-SPU	PMMSE	RDC	RDC-ne	MB	WavThr	Wiener_as	AudSup	noisy
SIG	2.4	1.9	3.1	3.2	2.8	2.1	2.8	2.5	2	3	2.4	3.2	1.8	3.3
BAK	1.9	1.9	2.2	2.2	1.9	2	2.1	1.8	1.9	2.4	2	1.8	1.8	2.1
OVRL	1.9	1.7	2.4	2.5	1.8	1.8	2.4	2	1.8	2.5	1.8	2.3	1.6	2.4



5. CONCLUSION:

- Subjective evaluation of speech enhancement algorithms can be done for separating free-text mixed sentences spoken by different speakers by using speaker diarization to corpus sentences.
- In terms of overall quality and distortion of speech , the algorithms performed the best are: MMSE-SPU, logMMSE, logMMSE-ne, pMMSE and MB. The subspace algorithms performed poorly.

- Proper selection of algorithm for particular process can be found after observing at different SNR behaviour.

6. REFERENCES:

[1] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," IEEE Trans. Speech Audio Processing, vol. 8, pp. 159-167, Mar. 2000.

[2]Cohen, I., 2003. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans. Speech Audio Proc., 466–475.

[3] Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. ASSP-32, 1109–1121.

[4] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement;' IEEE Trans. Speech Audio Processing, vol. 3, pp. 251-266, Jul. 1995.

[5] M. V. Wickerhauser. Adapted Wavelet Analysis from Theory to Software. A K Peters Press,Wellesley, MA,1994.

[6] F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113–120, Apr. 1979.

[7] Y. Hu and P. C. Loizou, “Subjective comparison of speech enhancement algorithms,” in *Proc. IEEE Int.Conf. Acoust., Speech, Signal Processing*, 2006, pp. 153–156.

[8] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.

[9] Ephraim, Y., Malah, D. 1985. Speech enhancement using a minimum mean square error log-spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, Signal Processing, vol. ASSP-33, pp. 443-445, Apr. 1985.

[10] Loizou, P. 2005. Speech enhancement based on perceptually motivated Bayesian estimators of the speech magnitude spectrum. IEEE Transactions on Speech and Audio Processing, vol. 13, no. 5, pp. 857-869, Sept. 2005.

[11] Hu, Y., Loizou, P. 2004. Speech enhancement by wavelet thresholding the multitaper spectrum. IEEE Transactions on Speech and Audio Processing, vol. 12, no. 1, pp. 59-67, Jan. 2004.

[12] Hu, Y., Loizou, P. 2003. A generalized subspace approach for enhancing speech corrupted with colored noise. IEEE Transactions on Speech and Audio Processing, vol. 11, no. 4, pp. 334-341, July 2003.

[13] Jabloun, F., Champagne, B. 2003. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. IEEE Transactions on Speech and Audio Processing, vol. 11, no. 6, pp. 700-708, Nov 2003.

[14] Gannot, S., Burshtein, D., Weinstein, E. 1998. Iterative and sequential Kalman filter-based speech enhancement algorithms. IEEE Transactions on Speech and Audio Processing, vol. 6, no. 4, pp. 373-385, July 1998.

[15] ITU, ITU-T Rec. P. 862. Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, 2000.

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