Assessment of Speech Denoising Algorithms Using Cepstral Distance

R.Dhivya[1], S.Sathya[2], R.Dhanapal[3]

[1,2,3] Department of Biomedical Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore

Abstract- **Speech enhancement algorithms are assessed and their recitals are assessed using Cepstral Distance measure. The Spectral Subtractive algorithms, Statistical model based algorithm, Spectral Subspace algorithm and Wiener algorithm are compared with Neighshrink algorithm (a wavelet thresholding technique). An IEEE speech sentence is chosen for the study and the algorithm is tested with noisy speech signal produced by a prosthetic device for laryngectomy patients. Four real world noises at 5 dB, 10 dB & 15 dB noise levels are used for the assessment. From the results, it is observed that the performance of the wavelet thresholding technique is superior compared to the spatial domain techniques.**

Keywords: Cepstral Distance; Speech enhancement; Laryngectomy; Neighshrink

I. INTRODUCTION

Naturally noise signals are combined in ordinary speech signals which make the removal of background noise signals mandatory for the improvement of speech processing. Speech quality evaluation has paved a way in many domains like telecommunications, audio, video etc for which many speech algorithms are used. Precise and consistent evaluation of speech quality is being required for the clients who are more advantageous of using speech processing systems. The quality of a speech algorithm depends on how effectively it enhances the speech with desired feature [1]. Even if many algorithms with equal output ranges are available for speech processing, the user may identify the one algorithm producing usual, amusing and suitable speech of interest. Hence it is fixed to determine the trait of speech signals [2].

The consistency of speech quality is a very tedious function to achieve since it involves subjective evaluation. Quality evaluation may be done using examination of listeners or objective quality measures. Examination of listeners comes under Subjective evaluation, where many people with normal hearing capacity give their opinions based on the speech they hear. Their opinions can be categorized as Unsatisfactory-poor-fair-good-excellent and are rated as 1-2-3-4-5 respectively. These ratings after averaging give a mean score which cannot be taken as a benchmark since it is not consistent enough. Consequently, if objective measures with mathematical calculations are considered for the cepstral distance of speech can be determined without involving subjective assessment.

Enormous speech denoising algorithms have been projected to enhance the recital of devices related to speech processing without reducing the quality of speech among the noisy environment. Noise reduction methods such as Spectral Subspace algorithm, Statistical model based algorithm, Spectral Subtractive algorithm, Wiener algorithm and wavelet thresholding are considered for this study. The Cepstral distance values derived, assists to blotch the algorithm which has the finest excellence among these algorithms.

II. SPEECH ENHANCEMENT

For the reduction of background noises and enhancement of speech quality, Speech Enhancement Algorithms or noise suppression algorithms are used. The one parameter to be noted while enhancing speech quality is speech distortion which occurs in most of the algorithms. So the major dispute in designing an efficient algorithm is to restrain the noise without introducing any discernible distortion in the speech indicator.

A. Spectral Subspace Algorithm

The clean signal is limited to a subspace of the noisy Euclidean space and rancid the vector space of a noisy signal into a subspace that is engaged by the clean signal and the noisy signal. The clean signal is predictable by nulling the component of noisy vector that exists in the “noise subspace”. The orthogonal matrix factorization technique decomposes the vector space into “clean signal” and “noise signal”. Dendrinos et al [3] makes use of Singular Value Decomposition on a data matrix comprising time-domain amplitude values. Later on Ephraim and Mallah [4], projected the use of Eigen value decomposition of the signal covariance matrix [5] which gives the KLT based method to enhance signals ruined with noise.

B. Statistical Model Based Algorithm

The speech enhancement complexity is modeled in a statistical assessment framework. A linear estimatorof the constraint is found if the Fourier transform coefficient is a noisy signal and if it is a clean signal a nonlinear estimator of the parameter is found. This group includes Wiener algorithm and Minimum Mean Square Error (MMSE) algorithm. McAulay and Malpass [6], proposed a maximum-likelihood approach for evaluating the Fourier Transform coefficients of the clean signal. This was pursued by Ephraim et al [7], who predicts an MMSE estimator of the magnitude spectrum.

C. Spectral Subtraction Algorithm

Spectral-subtractive algorithm are based on the theory that, as the noise is additive, one can weigh up the noise spectrum when speech is not present and eliminate it from the noisy signal. Spectral-subtractive algorithms is proposed by Weiss et al [8] in the correlation domain progressed by Boll [9] in the Fourier transform domain.

D.Wiener Algorithm

Signal processing using Fourier analysis is not suitable for the detection and classification, so Wiener type algorithm [10] determines the complex spectrum, to attain a clean signal from the noisy signal degraded by additive noise. Wiener filters works only when the signal characteristics are known in advance but will deform some of the desired signal when thresholding value is applied. The demerit of the Wiener filter is the fixed frequency response at all frequencies and the requirement is to approximate the power spectral density of the clean signal and noise to filtering.

E.Wavelet Thresholding

Wavelet shrinkage is functional to remove the noisy wavelet coefficients while hold the coefficients representing the signal features. Equation 1 characterizes the noisy signal.

$$y(t) = x(t) + n(t) \qquad (1)$$

where x(t) is the original signal and n(t) represents noise.

In wavelet thresholding techniques, the speech signal is first decomposed into approximation and detail sub band using DWT. The Neighshrink [11] utilizes the neighboring wavelet coefficients for thresholding. Threshold value for Neighshrink is estimated as given in equation 2.

$$\lambda = \frac{\sigma\sqrt{2\log(N)}}{1+\log(j)} \qquad (2)$$

where j is the number of levels of decomposition using DWT. The shrinkage function for Neighshrink is given in equation 3.

$$w_{j,k} = \begin{cases} w_{j,k}.\,(1-\lambda^2)/\,N_{j,k}^2, & N_{j,k}^2 \geq \lambda^2 \\ 0, & N_{j,k}^2 < \lambda^2 \end{cases} \qquad (3)$$

where $N_{j,k}^2 = w_{j,k-1}^2 + w_{j,k}^2 + w_{j,k+1}^2$, (the first and last term do not exist if k= 0 or k= N).

III. CEPSTRAL DISTANCE

If the Spectrum from a speech signal logarithmically undergoes Inverse Fourier Transform a cepstrum is derived. The Cepstrum Distance [12] is a measure of the log-spectrum distance among clean and distorted speech. Cepstrum is designed by taking the logarithm of the spectrum and adapting back to the time-domain. By doing so, we can divide the speech signal from the convolved vocal tract uniqueness. Cepstrum Distance can be intended as shown in equation 4.

$$d_{cep}(c_x, \bar{c}_{\hat{x}}) = \frac{10}{\log_e 10}\sqrt{2\sum_{k=1}^{p}\left[c_x(k) - c_{\hat{x}}(k)\right]^2} \qquad (4)$$

where Cx (k) and Cxˆ (k) are the cepstral coefficients of the clean and enhanced signals.

IV. RESULT

The IEEE sentence “Read verse out loud for pleasure” is taken for the study and the voice engendered by the speaker after implantation with Blom-Singer Duckbill Voice Prosthesis is documented using a unidirectional microphone in an anechoic room and is piled up on a computer. 8 male subjects were taken for the study. Spectral Subspace algorithm, Statistical model based algorithm, Spectral subtractive algorithm, Wiener algorithm and Neighshrink algorithms are applied and the cepstrum values are computed with reference to original signal for all the four real world noise signals such as babble, car, street, and train at three dB levels (5dB, 10dB and 15dB).

The typical resolute values of 8 subjects of cepstral distance with spectral subspace algorithm are given in Table 1 and the value ranges from 2.4017 to 5.5752. The typical resolute values of 8 subjects of cepstral distance scores with statistical model based algorithm are given in Table 2 and the value ranges from 1.5721 to 4.1846.Table 3 shows standard calculated cepstral distance values for 8 subjects with spectral subtractive algorithm and the value varies between 2.9852 and 5.1251.Table 4 and 5 shows the calculated values of cepstral distance scores of Wiener algorithm and Neighshrink algorithm with varied values 2.6890 to 6.1105 and 3.7066 to 6.8794.

The scores of cepstral distance specify that the excellence of sentences improved by Spectral Subspace algorithm, Spectral Subtractive algorithm and Statistical model based algorithm are insufficient and good. Compared to all algorithms, Neighshrink algorithm is fair. With 4 noises at 3 levels, bar charts of cepstral distance values of all algorithms are pointed out in Figure 1, 2, 3 and 4.

TABLE-I CEPSTRAL DISTANCE SCORE- SPECTRAL SUBSPACE ALGORITHM

Noise in dB	*5*	*10*	*15*
Babble	5.5752	4.7722	3.9146
Car	3.7684	3.1331	2.4017
Street	4.9544	4.1190	3.1098

Train	4.0708	3.1440	2.6741

TABLE-II CEPSTRAL DISTANCE SCORE-STATISTICAL MODEL BASED ALGORITHM

Noise in dB	***5***	***10***	***15***
Babble	4.1761	3.7621	3.145
Car	3.4795	2.9827	2.2689
Street	4.1846	4.0222	2.5211
Train	3.2841	2.0718	1.5721

TABLE-III CEPSTRAL DISTANCE SCORE-SPECTRAL SUBTRACTIVE ALGORITHM

Noise in dB	***5***	***10***	***15***
Babble	5.1251	4.5306	3.7082
Car	3.9883	3.5317	3.1625
Street	5.108	4.1407	3.6034
Train	4.096	3.4855	2.9852

TABLE-IV CEPSTRAL DISTANCE SCORE- WIENER ALGORITHM

Noise in dB	***5***	***10***	***15***
Babble	6.1105	5.2844	4.6015
Car	4.7549	4.136	2.689
Street	5.8409	5.1042	4.5628
Train	4.9407	4.2444	2.8388

TABLE-V CEPSTRAL DISTANCE SCORE-NEIGHSHRINK ALGORITHM

Noise in dB	***5***	***10***	***15***
Babble	6.8794	5.553	4.708
Car	5.1268	3.6996	3.7066
Street	6.8405	5.8137	4.5337
Train	5.0473	3.845	3.8329

V. CONCLUSION

With the decreasing values of cepstral distance, Neighshrink shows the best performance among all other 5 algorithms. Speech quality and Intelligibility are two vital criteria in which voice communication rely on and the objective measure of cepstral distance predicts them. This work done can be extended to evaluate the performance of cepstral distance values over other types of objective measures like Composite measures, Bark Spectral Distortion, Modified Bark Spectral Distortion and Weighted Spectral Slope Distance.

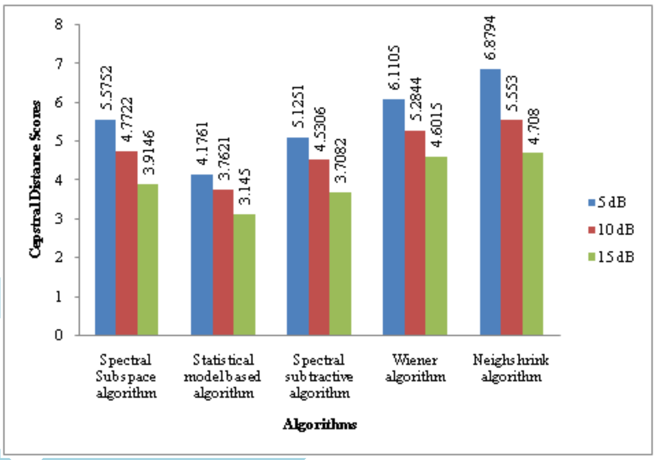

Fig-1 Cepstral Distance Score-Babble Noise

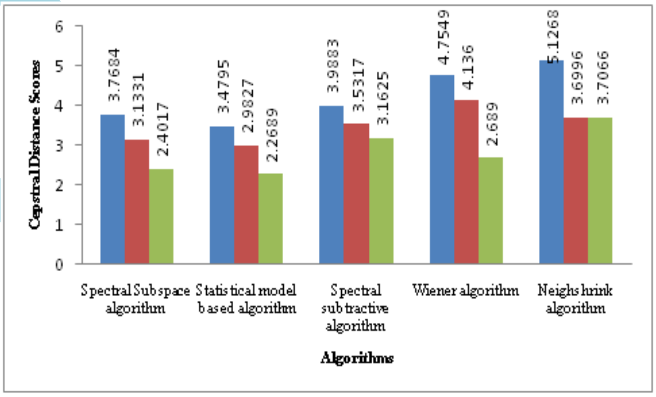

Fig-2 Cepstral Distance Score-Car Noise

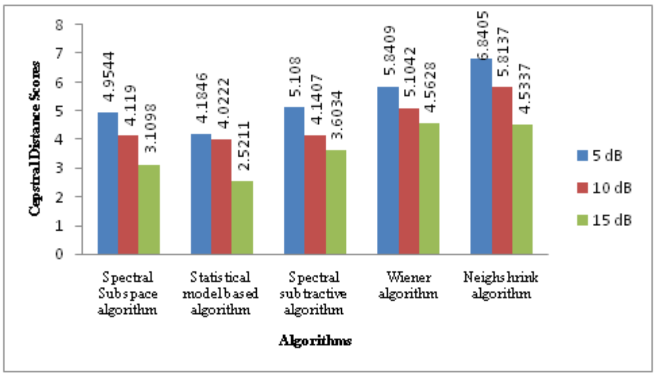

Fig-3 Cepstral Distance Score-Street Noise

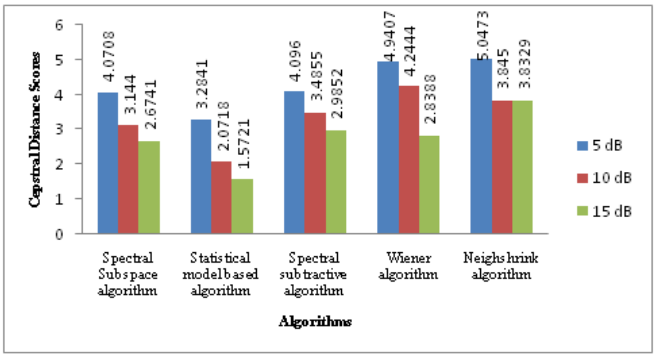

Fig-4 Cepstral Distance Score-Train Noise

REFERENCES

[1] Ma, J., Hu, Y. and Loizou, P., "Objective measures for predicting speech intelligibility in noisy conditions based on new band- importance functions", Journal of the Acoustical Society of America, 125(5), pp 3387-3405 (2009)

[2] Philipos C loizou," Speech enhancemnt :theory and practice", CRC press, Boca Raton FL (2007)

[3] Dendrinos, M.Bakamides, S., and Carayannis, G, "Speech enhancement from noise: a regenerative approach", Speech Communication., 10, pp 45-57(1991)

[4] Ephraim. andMalah, "A signal subspace approach for speech enhancement", Proc., IEEE Int, Conf,Acoustic. Speech Signal Process., 11, pp.355-358(1993)

[5] Mittal, U., Phamdo, N., "Signal/noise KLT based approach for enhancing speech degraded by colored noise", IEEE Trans. Speech Audio Proc. 8, pp159–167 (2000)

[6] McAulay, R.J. and Malpass, M.L, "Speech enhancement using soft decision noise suppression filter",IEEE Trans. Acoust. Speech Signal Process, 28, pp 137-145 (1980)

[7] Ephraim. andMalah, "A signal subspace approach for speech enhancement", Proc., IEEE Int, Conf,Acoustic. Speech Signal Process., 11, pp.355-358(1993)

[8] Weiss.M, Aschkenasy, E., and Parsons.T, "Study and development of the INTEL technique for improving speech intelligibility", Technical report NSC-FR/4023,Nicolet scientific corporation (1974)

[9] Boll, S.F, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoustic. Speech Signal Process. 27(2),pp113-120 (1979)

[10] Y.Hu and P. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," IEEE Trans. on Speech and Audio Processing, vol.12(1), pp.59-67, (2004)

[11] T.T. Cai and B.W. Silverman, "Incorporating information on neighbouring coefficients into wavelet estimation", Sankhya, Series A, (2001)

[12] Hu Y, Loizou PC (2008) Evaluation of objective quality measures for speech enhancement, IEEE Trans Audio Speech Lang Process 16(1

[13] N.Levine., Generalised closed sets in topology, Rend. Circ.Math.Palermo,19(2) (1970), 89-96.

[14] Levine N.Strong continuity in topological spaces,Amer.Math.Monthly 67(1960)269.

[15] H. Maki,R.Devi and K.Balachandran., Generalized α-closed sets intopology, Bull. Fukuoka Univ. Ed. Part III, 42(1993), 13-21.

[16] A.S.Mashhour, I.A.Hasanein and S.N.E.L Deeb, α-Continuous and α-open mappings, Acta Math.Hung., Vol.41,(1983), 213-218.

[17] Njastad, O., On Some Classes of Nearly Open Sets, Pacific J. Math. 15(1965) No.(3),961-970.

[18] S.Pious Missier and P.Anbarasi Rodrigo,On α*continuous functions in Topological spaces, OUTREACH Volume VII 2014 ,148 – 154.

