SPEECH ENHANCEMENT WITH SIGNAL SUBSPACE FILTER BASED ON PERCEPTUAL POST FILTERING

K.Ramalakshmi
Assistant Professor, Dept of CSE
Sri Ramakrishna Institute of Technology, Coimbatore

R.N.Devendra Kumar
Assistant Professor, Dept of CSE
Sri Ramakrishna Institute of Technology, Coimbatore

Abstract-A novel technique is presented to design the signal subspace speech enhancement based on perceptual post filtering. Firstly, by subspace filter the noisy speech is enhanced. The underlying principle is to decompose the vector space of the noisy signal into a signal plus noise subspace and a noise subspace. The decomposition can theoretically be performed by applying the Karhunen-Loeve transform to the noisy signal. Then for reducing stationary noise added to speech in noise Environments spectral subtraction is used. Finally, by a perceptual filter based on hearing masking effect the enhanced speech is smoothed, the clean speech is gained.

Keywords-Karhunen loeve transform, Spectral subtraction, Perceptual filter.

I. INTRODUCTION

In most speech enhancement systems, *musical noise* can be attributed to errors in measuring noise statistics. This auditory annoyance resembles a sum of sinusoids of changing frequencies, turning “off” and “on” over successive frames. Signal subspace techniques eliminate musical noise originating from fluctuating energy estimates by averaging over long windows. However, other artefact sources exist. These include rapid changes of model order and subspace swapping. The latter condition refers to noise basis vectors being incorrectly employed to describe the signal subspace.

This paper presents a methodology to quell artefacts produced by signal subspace techniques. A perceptual post-filter is placed at the output of the signal subspace filters to smooth the enhanced signal. It will be shown that psychoacoustic knowledge can attenuate imperfections with minimal distortion to the speech signal being recovered. Perception has been employed to the speech enhancement problem on several occasions. In [1, 2, 3], it was shown that the utilization of properties of the human auditory system has the capability to attenuate noise without distortion. Rezayee and Gazor [5] incorporated coloured noise handling into their algorithm by diagonalizing the noise correlation matrix using the estimated eigenvalues of the clean speech and nulling any off-diagonal elements. In addition, they incorporated subspace using the projection approximation algorithm developed by Yang[6].Jabloun showed in [8] that knowledge of the ear can improve parameter estimates for signal subspace techniques. In this work, filter coefficients are derived using eigen values which are calculated by projecting the excitation pattern of the noisy signal onto the squared magnitude of the individual eigenvectors.. Limiting the attenuation in an enhancement scheme can decrease distortion.

In this application, the perceptual filter accomplishes this by attenuating artefacts until they lie close to the masking threshold. As such, some of the artefact which is imperceptible is retained. By attenuating less, it is expected that fewer disturbances will be produced. Spectral averaging increases the width of tones within the noise residual according to the resolution of the ear. Temporal averaging, by limiting magnitude changes of the noise residual over several frames, effectively attenuates musical noise. Rapid frame to- frame spectrum variations are with high probability, the product of noise. By considering human perception, artefacts can be smoothed without noticeably altering the underlying speech signal.

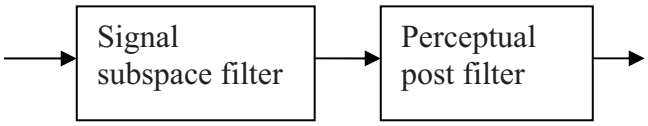

Fig 1.1 Proposed Hybrid System

The proposed hybrid system is illustrated by Fig. 1.1. It is a two-stage approach which is used to enhance the noisy speech in actual environment.
This paper will possess the following structure: Section 2 deals with the principle of the method, Section 3 deals with the Methodology of the work done, Section 4 with the Experimental results, Section 5 with the performance evaluation, and the conclusion is given in Section 6.

II. PRINCIPLE OF THE METHOD

2.1 SIGNAL SUBSPACE FILTER

Signal subspace based speech enhancement techniques decompose M-dimensional spaces into two subspaces: a signal subspace and a noise subspace. It is assumed that the speech signal can lie only within the signal subspace while the noise spans the entire space. Only the contents of the signal subspace are used to estimate the original speech signal. This chapter will describe the process of decomposing the complex space into orthogonal subspaces.

The speech enhancement problem will be described as a speech signal x being transmitted through a distortionless channel that is corrupted by additive noise w. The resulting noisy speech signal y can be expressed as

$$y = x + w$$

where x = [$x_1, x_2, \ldots, x_M]^T$, w = [$w_1, w_2, \ldots, w_M]^T$ and y = [$y_1, y_2, \ldots, y_M]^T$. The observation period has been denoted as M. Henceforth, the vectors w, x, y will be considered as part of C^M. The speech enhancement system will attempt to estimate the original signal using a single channel of received speech.

2.1.1 Karhunen-Lo`eve Expansion

It has been shown in many applications that the KL expansion is an excellent basis for dimensionality reduction.
The following definition is from Haykin: Definition 1 (Karhunen-Lo`eve Expansion) Let the M-by-1 vector u denote a data sequence drawn from a wide-sense stationary process of zero mean and correlation matrix R_u.

Let q1, q2, . . . , q_m be eigenvectors associated with the M eigenvalues of the matrix R_u. The vector u may be expanded as a linear combination of these eigenvectors as follows

$$u = \sum_{i=1}^{k} c_i q_i$$

The coefficients of the expansion are zero-mean, uncorrelated random variables defined by the inner product

$$c_i = q_i^H u$$

It can be shown that the KL expansion will always exist for a WSS random process using the spectral theorem.Clearly, as all WSS processes have Hermitian correlation matrices, they are diagonalizable. Even, if the correlation matrix is singular, the KL expansion will still exist. However, the column vectors of Q will not be linearly independent.

2.1.2 Subspace Decomposition Using Karhunen-Lo`eve Expansion

If an eigendecomposition is performed on the correlation matrix of the speech signal x, the following form is obtained

$$R_x = [Q_1 Q_2]\begin{bmatrix} \Lambda_{x1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} Q_1^H \\ Q_2^H \end{bmatrix}$$

The eigenvector matrix Q has been partitioned into two sub-matrices, Q1 and Q2. The matrix Q1 contains eigenvectors corresponding to non-zero eigenvalues. These eigenvectors form a basis for the signal subspace. Meanwhile, Q2 contains the eigenvectors which span the noise subspace. The matrix Q1QH1 is idempotent (P2 = P), Hermitian and span(Q1) = span(V). Thus, Q1QH1 is a projector onto the signal subspace. Similarly, Q2QH2 is the projector onto the noise subspace. As both subspaces complete CM, any input vector can be represented as

$$u = Q_1 Q_1^H u + Q_2 Q_2^H u$$

The expected power of a Karhunen-Lo˜eve coefficient can be shown to be equal to

$$E\{c_i^2\} = \lambda_i$$

As the eigenvectors which make up Q2 have null eigenvalues, they contribute no energy to the speech signal. As such, they can be omitted in a KL expansion without introducing error. The noise subspace eigenvectors, corresponding to a zero eigenvalue with multiplicity M−K, apart from being orthogonal to each other, are arbitrary.Thus, a reduced rank representation for the signal u will have the form

$$\bar{u} = \sum_{i=1}^{k} c_i q_i = Q_1 c$$

2.2 PERCEPTUAL POST FILTERING

The subspace filter described to be effective in improving the Signal-to-Noise Ratio (SNR) of an speech signal. Though, this method has also been found to introduce artefacts into the enhanced signal. These artefacts are known as musical noise and have often been evaluated as being more disturbing than the original corrupting noise. To remove these annoyances, a perceptual post-filter will be employed.

2.2.1 Spectral Subtraction

An estimate of the clean speech signal is required for an accurate masking threshold. This coarse approximation will be obtained from the generalized spectral subtraction algorithm.Spectral subtraction is based on the relationship for signals corrupted by uncorrelated noise

$$sy = sx + sw$$

Clearly, the magnitude response of the speech signal can be estimated from power subtraction. The noisy phase is retained in the enhancement system.

2.2.2 Masking threshold

Masking[10] is the phenomenon where the perception of one sound is obscured by the perception of another. A masker obscures a weaker signal known as the maskee. It is common to also refer to the maskee as the probe, target or signal. The threshold level above which a signal becomes audible in the presence of a masker is known as the masking threshold. Masking effects occur when two sounds occur at the same time or when separated by a small delay. The former is known as simultaneous masking while the latter is known as temporal masking. As the masking threshold is insensitive to phase, this approximation should not affect the performance of the perceptual post-filter. This system will smooth the output of the signal subspace filter and reduce the prominence of the musical noise. By utilizing properties of the human auditory system, the underlying speech signal should remain largely undistorted.

2.2.3 Psychoacoustic Filter

The psychoacoustic filter eliminates audible noise using a perceptual criterion. It is designed in the frequency domain to allow the vast sums of knowledge related to auditory perception to be applied. It will be shown that the incorporation of the principle of masking into an auditory post-filter will reduce these audible artefacts. Finally, an algorithm based on signal subspace methods utilizing an auditory post-filter will be outlined.

It is the goal of the perceptual post-filter to remove all traces of musical noise. Its strengths are two-fold: (1) distortion is minimized by attenuating only what is audible, and (2) peaks within the noise residual are smoothed by spectral and temporal averaging. However, the underlying speech should not be affected. Such systems have been used successfully in for speech enhancement. Limiting the attenuation in an enhancement scheme can decrease the production of artefacts.

Perceptual filters accomplish this by suppressing until the residual noise lies below the masking threshold. As such, some noise which is imperceptible is retained. By attenuating less, it is expected that fewer disturbances will be produced. For the listener, there should not be a discernible increase in residual noise as compared with conventional algorithms.

III. METHODOLOGY

The speech signal is sampled at a rate of 8000HZ. The signal is decomposed into a fixed size frames. Each framed values are transformed using karhunen loeve transformation which decompose the subspace into signal subspace and noise subspace . A rectangular analysis window is applied to the data prior to signal subspace filtering. After application of the post-filter, a sine-squared synthesis window is utilized for reconstruction. The signal subspace is given to the signal subspace filter to suppress the further noise. The noise correlation matrix and the output of signal subspace filter is given as input to the perceptual post filter for attenuating the noise.

The signal subspace filter will be modified to suppress musical noise by appending a perceptual post-filter to the output of the signal subspace filter. It should be stressed that this filter does not significantly attenuate the noise. Rather, it smoothes its input in a manner that musical noise is diminished and speech is unaffected.

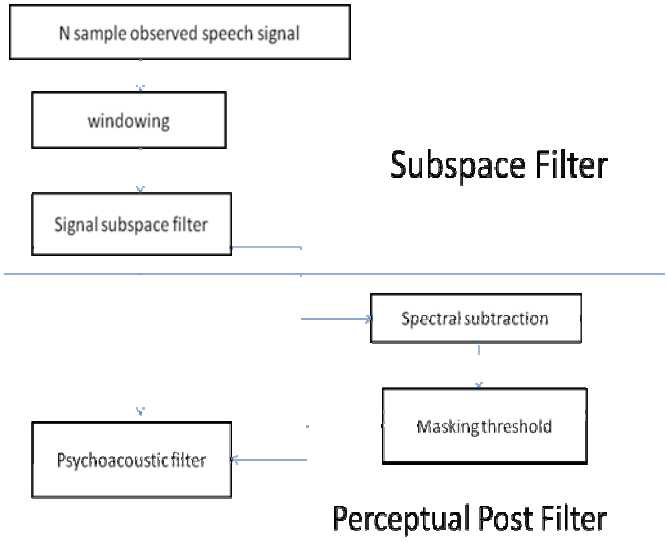

Fig.3.1 Block diagram of the hybrid system

A flow-chart describing the operation of the modified speech enhancement scheme can be found in Fig. 3.1. The signal subspace filter operates most effectively when utilizing very short frames (< 15 ms). The psychoacoustic filter attempts to conceal the salient noise using the perceptual properties of the ear while minimizing the distortion

to the underlying speech. This block is signal dependent, requiring an estimate of the noise correlation matrix and the masking threshold of the speech signal,to calculate an appropriate gain.

The input to the psychoacoustic filter is output frames from the signal subspace filter. The frames are combined by the overlap-add block which utilizes appropriate windows and overlap length.As the clean speech signal is unavailable, it is necessary to estimate the masking threshold of the speech signal from noisy data. Thus, the spectra of the clean speech is estimated using the spectral subtraction technique.

IV. EXPERIMENTAL RESULTS

The proposed algorithm for signal subspace speech enhancement is implemented and tested using speech files sampled at a frequency of 8KHZ at 16 bit rates. The speech wave file is converted into 16bit ASCII values. The raw values are applied to karhunen loeve transform to separate the speech and noise signal.The sample input signal with speech and noise is shown in Fig 4.1 and the sample output signal is shown in Fig 4.2

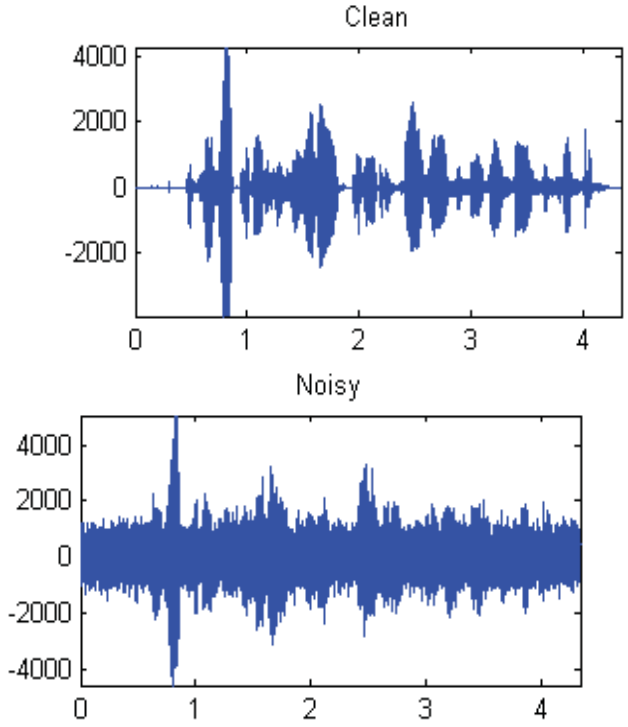

Fig 4.1 Clean and Noisy Speech Signal

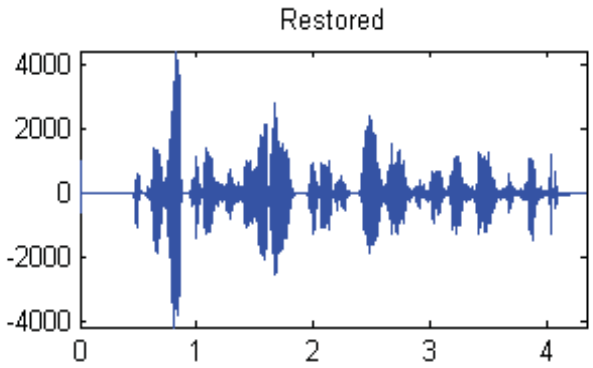

Fig 4.2 Output Signal

V. PERFORMANCE EVALUATION

As an objective measure, segmental signal-to-noise ratio (SNRseg) and weighted spectral slope (WSS) are used in the evaluation.

The weighted spectral slope measure is calculated,using the formula

$$d_{wss\,(j)} == K_{spl}\left(k - \hat{k}\right) + \sum_{K=1}^{35} w_a(k)\left(S(k) - \hat{s}(k)\right)^2$$

where k and $\hat{k}$ are related to overall sound pressure level of the original and enhanced utterances, and K_{spl} is a parameter which can be varied to increase overall performance.

Signal-to-noise ratio is used for evaluation of the Quality of random signal transmission. signal-to-noise ratio in decibels can be calculated, using the formula.

$$p_L = 10log\frac{\sum x(n)^2}{\sum[x(n)-y(n)]^2}$$

Where x (n) and y (n) are speech signals in discrete time.
Both, SVD based signal subspace and spectral subtraction noise reduction schemes were tested and compared in enhancing speech signals , which have been degraded by computer generated additive white Gaussian noise at different SNR Table 5.1 levels.

Table 5.1 signal to noise ratio calculation

SNR(db)	KLT based Signal Subspace	Spectral subtraction(SS)
0	10.44	10.4
5	8.66	7.91
10	6.95	5.86
15	5.32	4.20

VI. CONCLUSION

In this work, a frame-work to attenuate musical noise produced by signal subspace speech enhancement methods was presented. This speech restoration system incorporates the auditory concept of masking to smooth spectral parameters. Through informal listening tests, it has been shown that this algorithm is effective at attenuating musical noise while leaving speech relatively undistorted.
It has been further ascertained that the speech enhancement algorithm is well suited for many adverse noise environments. Their performance is evaluated using measures segmental signal-to-noise ratio (SNRseg) and weighted spectral slope (WSS).

REFERENCES

[1] M. Dendrinos, S. Bakamidis, and G. Carayannis, “Speech enhancement from noise: A regenerative approach,” Speech Communication, vol. 10, pp. 45–57, Feb. 1991.
[2] Y. Ephraim and H. L. V. Trees, “A signal subspace approach for speech enhancement,” IEEE Trans. Speech and Audio Processing, vol. 3, pp. 251–266, July 1995.

[3] J. Huang and Y. Zhao, "An energy-constrained signal subspace method for speech enhancement and recognition in colored noise," in Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing, vol. 1, (Seattle, WA), pp. 377–380, May 1998.
[4] J. Huang and Y. Zhao, "A DCT-based fast signal subspace technique for robust speech recognition," IEEE Trans. Speech and Audio Processing, vol. 8, pp. 747–751, Nov. 2000. References 87
[5] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," IEEE Trans. Speech and Audio Processing, vol. 9, pp. 87–95, Feb. 2001.
[6] B. Yang, "Projection approximation subspace tracking," IEEE Trans. Signal Processing, vol. 43, pp. 95–107, Jan. 1995.
[7] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," IEEE Trans. Speech and Audio Processing, vol. 8, pp. 159–167, Mar. 2000.
[8] F. Jabloun and B. Champagne, "On the use of masking properties of the human ear in the signal subspace speech enhancement approach," in Int. Workshop on Acoustic Echo and Noise Control, (Darmstadt, Germany), Sept. 2001.
[9] G. A. Soulodre, Camera Noise from Film Soundtracks. Ph.D. thesis, McGill University, Department of Electrical Engineering, Nov. 1998.
[10] N. Virag, "Signal channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Processing, vol. 7, pp. 126–137, Mar. 1999.