Multizone Speech Enhancement using Adaptive Filter

Sruthi K Jagadesh
Electronics And Communication Dept.
College Of Engineering
Cherthala, Alapuzha.P.O
Kerala

***Abstract*—Speech reinforcement (near end listening enhancement) is required in practical scenarios such as mobile communication, railway announcement etc. Early speech enhancement techniques considered only a single zone scenario. However practical cases there exists multizone situation. In such a situation state, speech can be degraded by various factors like noise, crosstalk from other sources, reverberation.**

This if left unaltered affects effective communication. so here we designed a framework which is capable of cancelling crosstalk and reverberation. we then use a adaptive filter for noise cancellation. Simulations validate the optimality conditions and show clear benefit in multizone speech processing compared with unprocessed and single zone conditions.

Keywords—reinforcement; ; reverberation ; multizone ,crosstalk

I. INTRODUCTION

This speech reinforcement (near end listening enhancement) technology is gaining rapid interest in the research community .It is a requirement in practical scenarios such as in public address system, mobile communications, hearing aids and railway announcements etc.

An important goal in speech communication is to transmit a speech signal such that it is correctly understood by the receiver submerged in noisy environment. With a speech reinforcement system we are really trying to achieve is a high direct sound to reverberant sound ratio so that everyone hears direct sound at a significantly higher level than the reverberant sound level .when trying to achieve max speech intelligibility direct sound is important. Speech intelligibility which forms an important factor in the speech processing in order for the receiver to understand .unfortunately the speech intelligibility can be harmed by background noise. While a decrease in speech intelligibility can be annoying in a public address system it can be dangerous for voice alarms in a fire detection system. Therefore there arises a need of speech reinforcement.

The objective of speech enhancement algorithm is to improve one or more perceptual aspects of noisy speech , mostly quality and intelligibility .Improving quality however will not necessarily lead to improvement in intelligibility .In fact in some cases improvement in quality might be accompanied by a decrease in intelligibility .this is due to the distortion imparted on the clean speech signal resulting from excessive suppression of acoustic noise .In some applications the main goal of speech enhancement algorithm is to improve speech quality while preserving at the very least speech intelligibility .Hence much of the focus of various speech enhancement algorithm is to improve speech quality.

Speech enhancements aims to improve speech quality by using various algorithms .The objective of enhancement is improvement in intelligibility and /or overall perceptual quality of degraded speech signal using audio signal processing techniques .These speech enhancement techniques can be classified to time domain and spectral domain methods. Recent major speech enhancement techniques are of the spectral domain which is used in a cell phone. The algorithms of speech enhancement for noise reduction can be categorized into 3 fundamental classes :Filtering techniques-spectral subtraction method,. weinner filtering , signal subspace approach(SSA) Spectral restoration -minimum mean square error short time spectral amplitude estimator (MMSESTSA) Speech model based

II. DISTORTION MEASURE

A distortion measure is an assignment of a nonnegative number to an input /output pair of a system .The distortion between the input or original and output or reproduction represents the cost of distortion resulting when that input is reproduced by that output. such measures a wide variety of applications in the design and comparison of systems .A distortion generally possess the following properties It must be subjectively meaningful in the sense that small and large distortion correspond to good and bad subjective quality. It must be tractable in the sense that it is amenable to mathematical analysis and leads to practical design techniques. It must be computable in the sense that the actual distortion resulting in a real system can be efficiently computed. The most common distortion measure is the traditional squared error or error power or error energy. This is largely used because of its tractability and computability. several speech distortion methods are used and they are discussed later.

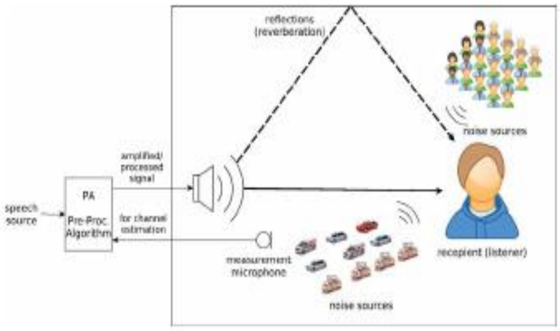
The objectives are to study various speech enhancement techniques, the role of distortion measures in speech processing and to implement a multizone speech reinforcement system.

In section 3, the problem is formulated .In section 4 , a detailed literature survey will be provided where the details of different speech enhancement techniques will be investigated. In section 5, provides solution for the problem with respect to the work done so far.

III. PROBLEM FORMULATION

The quality and intelligibility of speech are an important factor in our day to day life such as in mobile communication ,public addressing system and hearing aids. A practical scenarios consists of several zones. in such a framework signals from one zone can leak into other zones (crosstalk), causing intelligibility or quality degradation. Current source based systems only consider a single playback region where some kind of speech reinforcement is provided. In order to perform such a task the system should be capable of first estimating the noises and other disturbances which can interfere with our clean speech and reinforce it accordingly.

Therefore we have to build an optimization framework. This optimization framework accounts into noise, reverberation and zone crosstalk simultaneously .we then use a general adaptive filter for the noise reduction. Thus we aim to preprocess all speech signals from different zones directly. The noise reduction filter is followed by a amplitude and frequency shaper.



The above figure shows our problem diagrammatically which we face in our daily scenarios. Our objective is to design a method capable of overcoming this problem.

Before we begin to look into the paper, we will first provide a brief introduction regarding the various speech enhancement methods developed. Since speech enhancement techniques are many in number we provide only such methods which are related with our method which we are going to propose .

IV. RELATED WORKS

Most methods were based on SNR recovery. A speech signal present in a noisy environment is less intelligible than the same signal present in a quiet environment due to the fact that spectral distance between the speech signal and noise signal is reduced .The fundamental idea is to amplification of the speech signal to re-establish the distance between the average measured speech spectrum and the average measured noise spectrum to recover a certain signal to noise ratio. First approach is to amplify the speech signal in time domain with respect to a gain subjected to a constraint. This method is capable of giving good results in case of white noise. However increasing speech power can often cause discomfort and listening fatigue to the listener when the speech power has to be raised to a favourable SNR in presence of heavy noise.

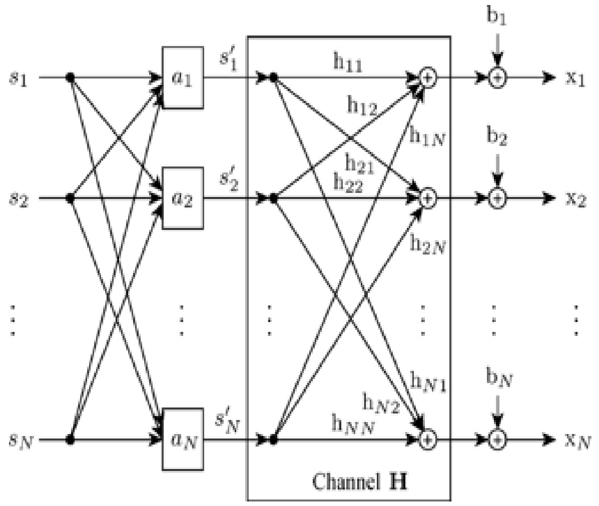
Another method is using tunable equalization filter . In this method consonants of the speech signal are enhanced by processing the input signal using a tunable band pass shelving filter whose cut off frequency can be dynamically adjusted. The proposed method is capable of preserving speech clarity well without introducing audible distortions. Cut off frequency of shelving filter adjusted such that level of output speech is approximately equal to the level of input speech. During transition from consonant to vowel cut off frequency of filter is ususally set high and output speech estimates are lower compared to the input speech level estimates since vowels carry low frequency and cut off frequency of filter is high. Therefore cut off frequency is shifted to a low frequency range and the shift is made proportional to the difference of input speech level estimates and the output speech level estimates. to maintain equality.

The spectral subtraction is based on the principle that the enhanced speech can be obtained by subtracting the estimated spectral components of the noise from the spectrum of the input noisy signal. The noise spectrum cannot be calculated accurately, but can be estimated during time when no speech is present in the input signal. Most of the single channel spectral subtraction methods use a voice activity detector (VAD) to determine when there is silence in order to get an accurate noise estimate. The noise is assumed to be short-term stationary, so that the noise from the silent frames can be used to remove noise from each speech frames.

The filter was preceded by a logarithmic and followed by an exponential characteristic. This is known as a logarithmic filter. This method was only applied to speech before adding noise, and not only with any reverberation. By a filter characteristic << 1(0:05) at very low 0 and unity greater than 2 Hz, a strong compression of the slow level changes is achieved without degrading the perceptually important modulation frequencies. The scaling factor in the logarithm defines the average level for compression, which should lie above the noise level. The latter has to be known for each critical band. 240 monosyllabic nouns were presented to 10 subjects with a p SNR of about 0, 5 and 10 dB. The dynamic compression referred to the masking level of the 0 dB white noise. This time, the pre processed words reached intelligibility some 40 percentage above those of the unprocessed words.

Tests were also carried out with dynamic expansion of higher frequencies (e.g at 4 Hz) .This, however, leads to annoying power peaks, e.g.in explosives, and suppression of short low-power segments.

V.PROPOSED METHOD



we consider a speech reinforcement scenario working across multiple zones, say N zones. We consider frame-based signal representations in the discrete Fourier transform (DFT) domain. These could come, e.g., in the context of a ubiquitous DFT-based speech processing scheme. The signal processing flow we consider is shown in above Fig. speech is segmented into overlapping time frames of about length N (shift size R < N), windowed, processed, and combined via weighted overlap-add (WOLA).Consider a (clean) speech frame s(t) in the time domain, where N denotes the frame number.

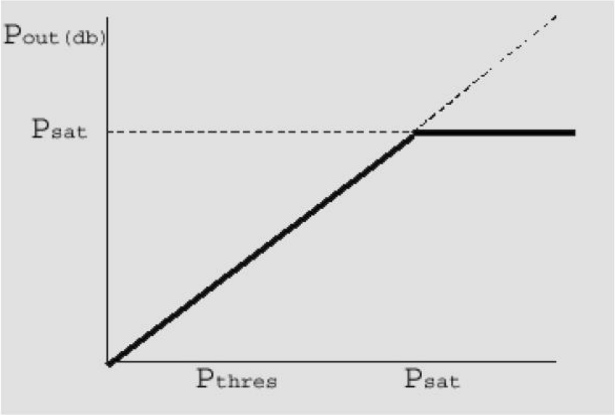
We also note that this model supports a source signal which is the same for all zones which is also known as single source broadcast in which we have for some single-zone source speech DFT coefficient . The sources packed in the vector S gets jointly processed by pre-processing functions (a1, a2, …,aN) obtained from overlap adding , thereby producing pre-processed signals. A decomposition into many frequency bands and an envelope filtering for each of them is carried out; the results are summed up. In practice this could be done by a filter bank, but in order to achieve more flexibility for our experiments, we employed the ”overlap add” (OLA) method. Overlapping segments of the signal are Hamming weighted each, and after appending zeroes symmetrically, are transformed by the FFT. The frame rate depends on the bandwidth of the window spectrum according to the sampling theorem.

The signal is then passed through the channel . It is here after passing through the channel noise comes in contact with the signal. Therefore it will be meaningful to add noise after dividing into frames. Simultaneously noise should also be divided into zones and added with speech accordingly. In most cases we are adding white Gaussian noise for testing and evaluation purpose.

WGN has a continuous and uniform frequency spectrum over a specified frequency band and has equal power per Hertz of this band. It consists of most of all frequencies at equal intensities and has a normal (Gaussian) probability density function. For example, a hissing sound or the sound of many people talking as a group can be modeled as WGN. Since we know that white Gaussian noise is random, we can generate it using MATLAB using the random number generator function known as random. we used the command wdencmp ,which performs noise reduction/compression using wavelets. It then returns a de-noised version of the input signal using wavelet coefficients thresholding. We also utilized the MATLAB command ddencmp.

Wavelets are nonlinear functions and do not remove noise by low-pass filtering like many traditional methods. Low-pass filtering methodes, which are linear and time invariant, can blur the sharp features of a signal and sometimes it will be difficult to separate noise from the signal where their Fourier spectra overlaps. For wavelets their amplitude unlike instead of the location of the Fourier spectra, differs from the noise. This allows for thresholding of the wavelet coefficients to remove the noise. If a signal has its energy concentrated in small number of wavelet coefficients, then their values will be somewhat large in comparison to the noise that has its energy spread over a large number of coefficients. This localizing properties of a wavelet transform allows the filtering of noise from a signal to be very much effective. Since most of the linear methods trade-off suppression of noise for broadening of the signal features, noise reduction using wavelets maintains features in the original signal that will remain sharp. A problem with such a wavelet denoising is the lack of shift invariance, which means the wavelet coefficients cannot move by the same amount that that the signal is shifted. This can be be overcomed by averaging the denoised result over all possible shifts of the signal. This works very well and even overcomes pseudo-Gibbs phenomena that is often seen due to lack of shift invariance. A frequency shaper is used followed by this noise reduction filter. The frequency shaper applies gain < 1 for hard-to-hear frequencies. Modifies gain for other specified ranges .The frequency shaper is designed to correct for loss of hearing at certain frequencies. The filter applies a gain greater than one to the frequencies that the user has difficulty hearing. As one of its parameters, the filter uses a vector of frequencies, that will define the user’s hearing characteristics. For each frequency range, the frequency shaper will apply a certain gain based on the user’s specific hearing . Thus , our frequency shaper is completely configurable to any user. After passing through frequency shaper it is then applied to a amplitude shaper.Once the signal is passed through the above mentioned Noise Reduction Filter and the Frequency Shaper, it will be then allowed to pass through our Amplitude Shaper. The dynamic range of hearing is measured in terms of sound pressure ranging in decibels. A normal hearing usually ranges from approximately 0 dB to 120 dB, where 0 dB is the Threshold of Hearing and 120 dB is the Threshold of Pain. Discomfort usually begins to occur around a saturation level of about 90 dB of sound.

We assume that the Frequency Shaper will raise the frequencies that the user has difficulty hearing to and sound pressure levels within his dynamic range of hearing. Therefore, our Amplitude Shaper only has to do is a check, bit by bit, that output power does not exceed a given saturation level, Psat. Since noise is concentrated to the low power levels as well, the filter also removes a significant amount of noise. Output power is nearly zero for levels below Psat.



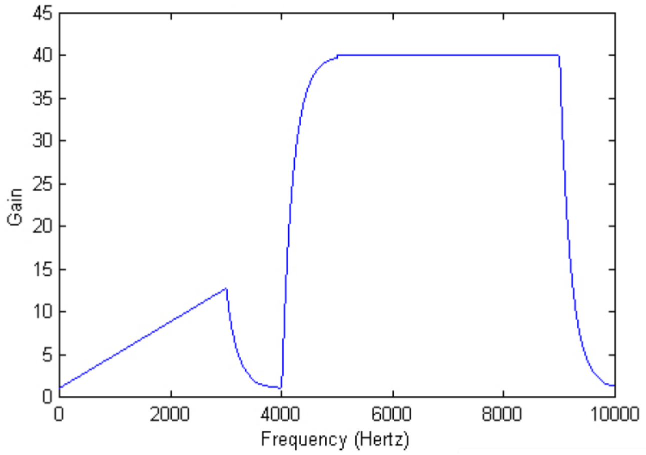
VI.RESULTS

In order to create the gain filter , we used the concatenation of piecewise functions that change at f = 3000,4000,5000, and 9000 in hertz.

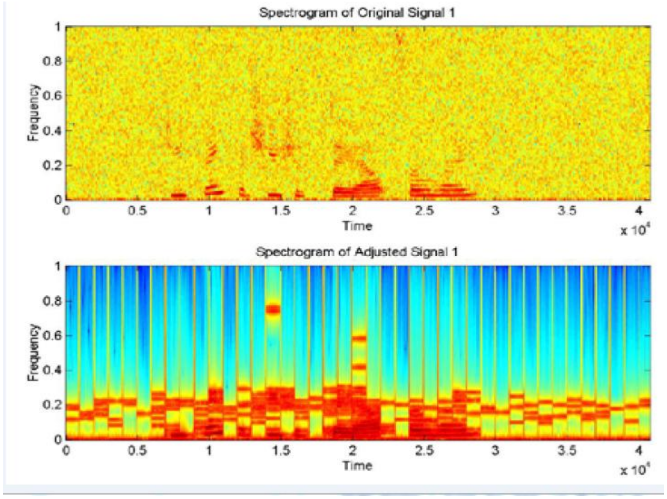
In the actual Matlab function, f was mapped to the specific

DFT coefficient that it corresponds to.

The frequency shaper transfer function is given by



The graphs below show what we hear. (Blue is before filtering, Red is after filtering)



VII.CONCLUSION

Comparing the graphs of the original signal and the filtered signal, we saw that the amplitude of the noise in the signal was noticeably reduced. We also compared the spectrograms of the two signals and saw that the speech signal was stronger and more recognizable.

VIII. REFERENCES

(1) B. Sauert and P. Vary, Near end listening enhancement: Speech intelligibility improvement in noisy environments, in Proc. IEEE Int. Conf. Acoust.Speech, Signal Process. (ICASSP), May 2006, vol. I, pp. 493496.

(2) C. H. Taal, R. C. Hendriks, and R. Heusdens, A speech preprocessing strategy for intelligibility improvement in noise based on a perceptual distortion measure, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Mar. 2012, pp. 40614064.

(3) C. H. Taal, R. C. Hendriks, and R. Heusdens, A low-complexity spectro-temporal distortion measure for audio processing applications, IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 5, pp. 15531564, Jul. 2012.

(4) T. Dau, D. Pschel, and A. Kohlrausch, A quantitative model of the effective signal processing in the auditory system. I. Model structure, J. Acoust. Soc. Amer., vol. 99, no. 6, pp. 36153622, Jun. 1996.

(5) C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, An algorithm for intelligibility prediction of time-frequency weighted noisy speech, IEEE Trans. Audio, Speech, Lang.Process., vol. 19, no. 7 , pp. 21252136, Sept. 2011.

(6) C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech, J. Acoust. Soc. Amer., vol. 130, no. 5, pp. 30133027, 2011.

(7) R. Niederjohn and J. Grotelueschen, The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression, IEEE Trans. Acoustics, Speech, Signal Process., vol. ASSP-24, no. 4, pp. 277282, Aug. 1976.

(8) I. B. Thomas, The second formant and speech intelligibility, in Proc. Nat. Electron. Conf., 1967, vol. 23, pp. 544548.

(9) I. B. Thomas and R. J. Niederjohn, The intelligibility of filtered-clipped speech in noise,J. Aud. Eng. Soc., vol. 18, no. 3, pp. 299303, Jun. 1970.

(10) P. S. Chanda and S. Park, Speech intelligibility enhancement using tunable equalization filter, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Apr. 2007,vol. 4, pp. IV-613IV-616.

(11) ANSI S3.5-1997. Methods for Calculation of the Speech Intelligibility Index ANSI. NewYork, NY, USA, 1997.

(12) T. Langhans and H. Strube, Speech enhancement by nonlinear multiband envelope filtering,in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), May 1982,

(13) J. Crespo and R. Hendriks, Multizone near-end speech enhancement under optimal second-order magnitude distortion, in Proc. IEEE Workshop Applicat. Signal Process.Audio Acoust. (WASPAA), Oct. 2013.

(14) R. Gray, A. Buzo, J. Gray, A. , and Y. Matsuyama, Distortion measures for speech processing, IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-28, no. 4, pp.367376, Aug. 1980.

(15) SEPARATION OF SPEECH FROM SIMULTANEOUS TALKER R.I. Damper, J.R.Thorpe and C.H. Shadle , Department of Electronics and Computer Science,University of Southampton Southampton SO17 1BJ, UK

(16) SPEECH SEPARATION BY KURTOSIS MAXIMIZATION, James P. LeBlanc PhillipL. De Le on Klipsch School of ECE,New Mexico State University Las Cruces, NM USA