Variable Step Size NLMS Algorithm for Speech Dereverberation

P. Sony
Asst. professor
Vishnu Institute of Technology

K. Kiran
Asst. Professor
Vishnu Institute of Technology

Ch. Divya Susmitha
Asst. Professor
Guntur Engg College

ABSTRACT
In many speech communication applications, the recorded speech signal is subject to reflections on the room walls and other objects on its way from the source to the microphone. Then resulting signal is called reverberated, which decrease Automatic Speech Recognition (ASR) performance and loss of intelligibility to listeners. However, it is still a challenging problem because of the nature of common room impulse response (RIR). RIR is generated artificially based on parameters of the room and its intensity depends on the size, shape, dimensions and materials used in the construction of the room. Here proposed an LMS-like gradient adaptive maximizing algorithm that maximizes the kurtosis of the LP residuals of the speech signal to the clean speech. The method used in this is maximizes kurtosis of LP residual of speech to removing reverberation from degraded speech. The performances of these methods are analysed using Reverberation Index (RR) and Speech Distortion (SD) parameters.

Keywords
ASR, NLMS, RR, SD, RIR

1. INTRODUCTION
The last decade has seen the rapid development and pervasiveness of speech technologies,such as hands free(mobile) telephones, videoconferencing, and hearing aids. In the near future, we can expect to see a dramatic spread of human-machine communication systems, for example, voice-operated electrical appliances and intelligent communication robots, which have already been partially launched and are attracting attention in the market. The main user benefit of hands-free telephones is that they enable the user to walk around freely without wearing a headset or microphone, so they provide a natural communication style. Reverberant speech is generally assumed to consist of three parts: a direct-path response, early reflections,and late reverberation. Since late reverberation is known to be a major cause of ASR performance degradation and speech intelligibility loss. Over the years, several methods of speech dereverberation based on a simplified discreteModel of speech production have been proposed. The basic model consists of an excitation Source, and a time-varying vocal tract filter. Such as model can easily be modelled by an Auto-Regressive (AR)

2. PREVIOUS WORK
2.1 Reverberation
When speech signals are obtained in an enclosed space by one or more microphonespositioned at a distance from the talker, the observed signal consists of a superposition of many delayed and attenuated copies of the speech signal due to multiple reflections from the surrounding walls and other objects.

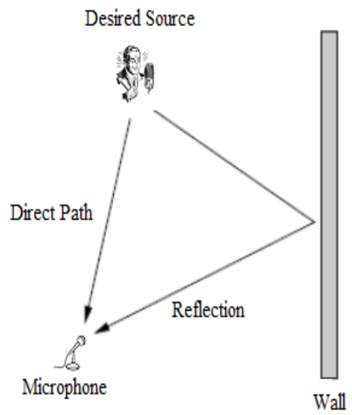

Fig. 1: The Simplest Case of One Reflective Surface

2.2 Sabines Equation
Sabine's reverberation equation was developed in the late 1890s in an empirical fashion. He established a relationship between the RT60 of a room, its volume, and its total absorption (in sabins).

This is given by the equation:

$$RT_{60} = \frac{4 ln 10^6}{C} \frac{V}{Sa} \approx 0.1611 m^{-1} \frac{V}{Sa} \quad (3.1)$$

Where **C** is a number that relates to the speed of sound in the room,

V is the volume of the room in m³,

S total surface area of room in m²,

A is the average absorption coefficient of room surfaces,

And the product **Sa** is the total absorption in sabins

The reverberation time RT_{60} and the volume V of the room have great influence on the

Critical distance d_c (conditional equation):

$$d_c \approx 0.057 \sqrt{\frac{V}{RT_{60}}} \quad (3.2)$$

Where critical distance d_c is measured in meters, volume V is measured in m³, and reverberation time$RT_{60}$is measured in seconds.

Early Reverberation: The sounds which have undergone reflection to one or more surfaces such as walls, floors, furniture are received after a short time. The reflected sounds are separated in both time and direct ion from the direct sound. These reflected sounds combine to form a sound component called as early reverberation. Early reverberation provides the details about the size and position of the source in space as it varies when the source or microphone moves in the space. As long as the delay of reflections doesn't exceed the limit 80 -100 ms approximately with respect to the arrival time of the direct sound, early reverberation is not perceived as separate sound. It is reinforced with respect to the speech intelligibility and also to enhance the direct sound known as precedence effect. This effect makes the conversations easier in small-room acoustics as the walls, ceiling and floor are very close. This reverberation also causes spectral distortion known as coloration.

Late Reverberation: Late reverberation results from reflections which arrive with larger delays after the arrival of the direct sound. They are perceived either as separate echoes, or as reverberation, and impair speech intelligibility. The acoustic channel between a source and a microphone can be described by an Acoustic Impulse Response (AIR). This is the signal that is measured at the microphone in response to a source that produces a „sound impulse". The AIR can be divided into three segments, the direct path, early reflections, and late reflections, as shown in figure2.2.1. The convolution of these segments with the desired signal results in direct sound, early reverberation and late reverberation respectively. In signal processing perception, early reflections materialize as separate delayed impulse in RIR whereas late reflections materialize continuously without any separation with the delayed impulses.

3. PROPOSED TECHNIQUE

3.1 Dereverberation

Speech dereverberation algorithms can be classified into one of the three main categories:

Beamforming– The observed signals received at the different microphones aredelayed, weighted and summed, so as to form a beam in the direction of the desired source and to attenuate sounds form other directions.

Speech Enhancement– The observed speech signals are modified so as to better represent some features of the clean speech signal according to a priorimodels of the speech waveform or spectrum.

Blind System Identification And Equalization–The acoustic impulse responses are identified blindly (using only the observed signals) and then used to design an equalization filter that compensates for the effect of the acoustic impulse responses

3.2 Linear Prediction of Speech

A Speech signal can be expressed as a linear combination of its past samples. Based on the source filter model the clean speech can be modelled as

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k)+u(n) \qquad (3.2.1)$$

Where a_k filter coefficient, u (n) glottal pulse excitation signal.

Let's say the predicted signal for the above speech be

$$\tilde{s}(n)=\sum_{k=1}^{p} b_k \; s(n-K) \qquad (3.2.2)$$

b_k are linear prediction coefficient

For $a_k=b_k$, error in prediction e(n)=s(n)-s`(n)=u(n),this error in prediction known as LP residual.

LP of the reverberant speech can be written as

$$X(n)=-\sum_{k=1}^{p} h_k x(n-k)+e_x(n) \qquad (3.2.3)$$

$e_x(n)$ is known as LP residual of reverberant speech.as reverberation mainly effects excitation signal ,it can be removed by modifying lp residual in manner to achieve $e_x(n)=u(n)$ and clean speech signal synthesized from cleansed signal.

3.3 Effects of Reverberation on The Prediction Residual

Consider a frequency domain formulation of the source-filter speech production model discussed. The speech signal is written as

$$S(e^{jw}) = E(e^{jw})V(e^{jw}), \qquad (3.3.1)$$

where $E(e^{jw})$ is the Fourier transform of the prediction residual and $V(e^{j\omega})$ is thetransfer function of the all-pole filter evaluated for $z = e^{j\omega}$.

Now consider the speech signal produced in a reverberant room as defined inwhich, in the frequency domain, leads to

$$X(e^{j\omega}) = S(e^{j\omega})H(e^{j\omega}) = E(e^{j\omega}) V(e^{j\omega})H_m(e^{j\omega}). \qquad (3.3.2)$$

An inverse filter, $B(e^{jw}) = 1+\sum_{k=1}^{p} b_k e^{jwk}$, can be obtained such that ε {B $(e^{jw}\} \cong A\left(e^{jw}\right)$Filtering the reverberant speech signal with this inverse filter, the coefficients of which are obtained from the reverberant speech signal, results in

$$E_m\left(e^{jw}\right) \cong E\left(e^{jw}\right)H_m\left(e^{jw}\right) \qquad (3.3.3)$$

Where $E(e^{jw}$ is the Fourier transform of the prediction residual $e_m(n)$ obtainedfrom the reverberant speech observation at the mth microphone. Thus, in the time domain, the prediction residual obtained from reverberant speech is approximately equal to the clean speech residual convolved with the room impulse response.

3.4 Maximum Kurtosis Based Derverberation

In this section the maximum kurtosis based blind dereverberation is discussed. The basic idea is to maximize the kurtosis of LP residual of received reverberant signal to achieve dereverberation.

- Speech Enhancement for human perception (useful for applications related to hearing aids).
- Single/Multi-channel implementation
- Exploits the properties of cumulants, and speech: Higherorder statistics of a Gaussian distribution are zero, henceremoves the Gaussian Noise.

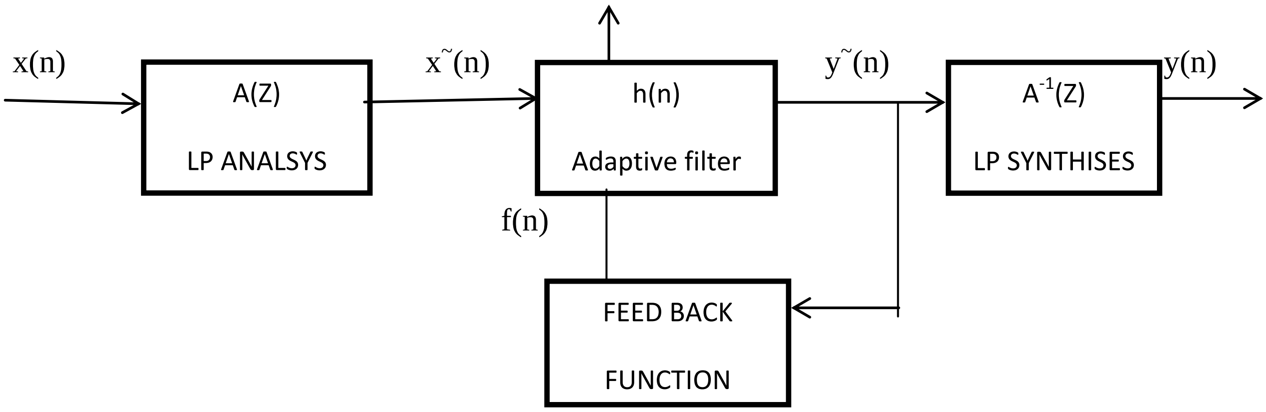

Fig. 2: Steepest Ascent Algorithm Used For Speech Dereverberation

4. RESULTS

A clean speech of 8000 Hz sampling frequency was convolved with the RIR. The kurtosis of the LP residual of this output was calculated. The distance between the source and mic increases, the reverberation in the received speech increases, and the kurtosis of its residual decreases.

In this we perform dereverberation experiments using maximising kurtosis LP residual algorithm with variable step size parameter µ from NLMS algorithm and with fixed step size parameter. The experiments performed in this are

Exp1.in this speech source position src=[3 2 1], microphone position mic=[7 6 1],n=10, room size is rm=[18 17 20] and r=0.7

Here we taken two setups for reflection coefficients r=0.7 as setup 1, r=0.9 as setup 2.

For Kurtosis with VSS

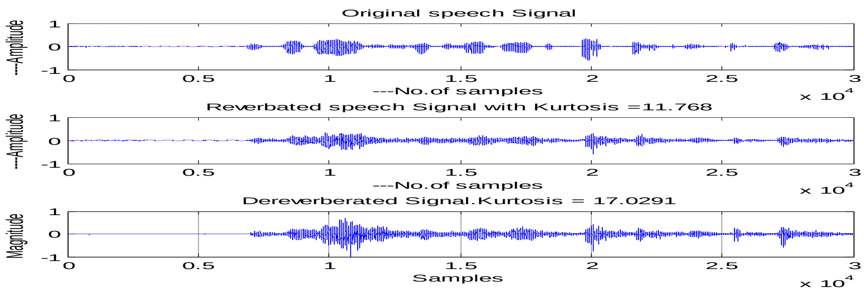

Fig. 4.1 Waveforms of clean, reverberant and processed speech – exp1

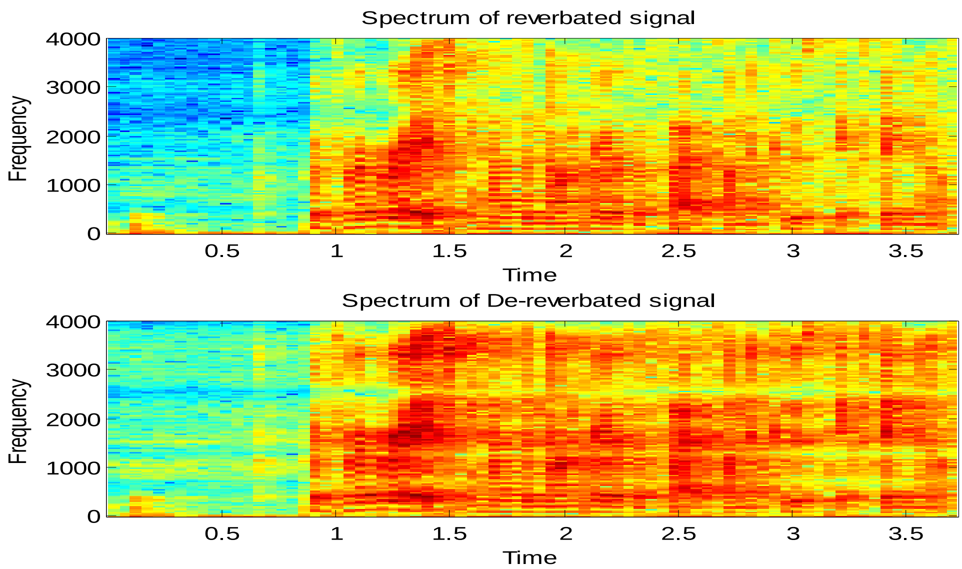

Fig 4.2 spectrograms of reverberant, dereverberant signals-exp1

It is apparent from the plots that reverberation spreads the signal energy in time making the LP residual less peaky, the kurtosis values of different LP residuals also align with the results. It is also evident that the LP residual of the filtered output is much closer to that of the clean speech, and the kurtosis value has also increased.

For Kurtosis With Out Vss

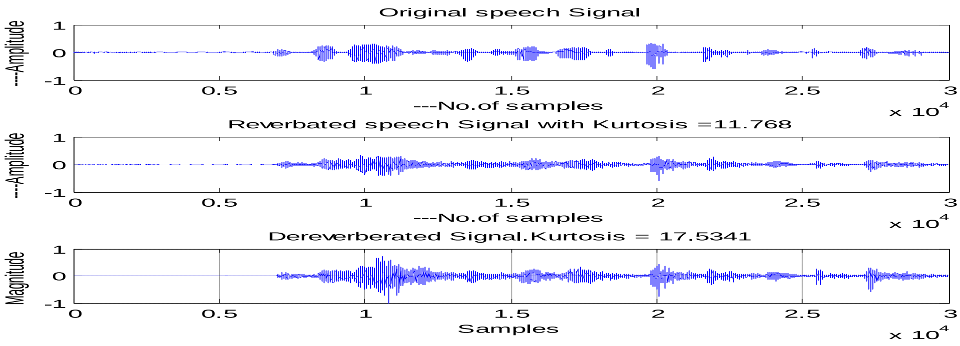

Fig 4.3 Waveforms of clean, reverberant and processed speech - Exp 1

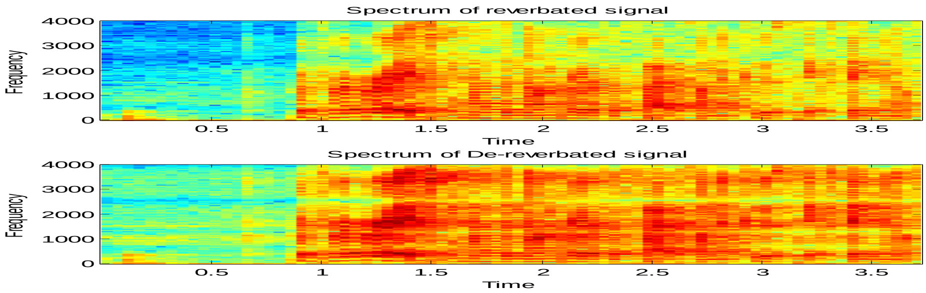

Fig 4.4 spectrograms of reverberant, dereverberant signals-Exp1

Bychanging reflection coefficient r=0.9, and perform dereverberation operation and the results are listed in the tables and the reverberation is improved in this condition.

These are the schematic and spectrograms of kurtosis of Lp residual maximising algorithm. The experimental results are shown below.

1. In this speech source position is src=[5 2 1],mic=[9 8 1.2],room size rm=[20 19 21],N=10,

And R=0.7

	SD	RR_{IMP}
KURTOSIS WITH VSS	-28.3525	282.3258
KURTOSIS WITH OUT VSS	-22.3055	282.3258

Table 4.2 shows SD and RR imp values for exp2, when R=0.9

	SD	RR_{IMP}
KURTOSIS WITH VSS	-35.3322	285.6443
KURTOSISWITH OUT VSS	-24.7443	277.2948

Table 4.1 shows SD and RR $_{imp}$ values for exp1, when R=0.7

5. REFERENCES

[1]. P. A. Naylor and N. D. Gaubitch, Eds., Speech Dereverberation. Springer 2012- 05-09.

[2]. B.Yegnanarayana and P. S. Murthy, Enhancement of reverberant speech using LPResidual Signal," IEEE Transactions on Speech and Audio Processing, vol. 8, no. 3,pp 267{281}, May 2000.

[3]. B. W. Gillespie, D. A. F. Florencio, and H. S. Malvar, Speech dereverberation via Maximum-kurtosis subband adaptive filtering," in ICASSP '01: IEEE International Conference on Acoustics, Speech, and Signal Processing 2001, vol. 6, pp. 3701{3704}, IEEE.

[4]. J. Benesty,T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay,Advances in Network and Acoustic Echo Cancellation. Berlin, Germany:Springer-Verlag, 2001.

[5]. H.-C. Shin,A.H. Sayed, and W.-J. Song,“Variable step-size NLMS and affine projection algorithms,” IEEE Signal Process. Lett.vol. 11, no. 2, pp. 132–135, Feb. 2004.

[6]. J. Benesty, H. Rey, L. Rey Vega, and S.Tressens.“A nonparametricVSSNLMS algorithm,” IEEE Signal Process. Lett, vol. 13, no. 10, pp. 581–584, Oct. 2006.