Section A-Research paper

ADAPTIVE DEREVERBERATION OF SPEECH SIGNAL USING MULTI CHANNEL

[1] Praveen P.B., [2] Dr. Aravind H.S.

[1] Research Scholar, Research Centre, ECE Dept., JSS Academy of Technical Education, Bengaluru, Karnataka, India
[2] Professor, Department of Electronics & Communication Engg., JSS Academy of Technical Education, Bengaluru, Karnataka, India

praveen.patelb@gmail.com aravindhs@jssateb.ac.in

Abstract

The quality of the speech signal degrades mainly due to additive noise and/or reverberation. This in-turn degrades the performance of many speech-based applications like automatic speech recognition, telecommunications, speech recordings, etc. This paper deals with reducing the reverberation effect without prior knowledge of room impulse response (RIR) using multiple channels. In most cases, it is unknown, therefore, much research is done on blind dereverberation [2][3]. This paper proposes an adaptive method to perform de-reverberation of an unknown RIR using 2 or more channels. This method first estimates the impulse response of the echo path for each channel using the method described in [1], using it finds the transmission delay at 2nd mic, 3rd mic and so on with respect to 1st mic and finally cancels the echo signal. This method doesn't extract the clean signal, but the direct signal as explained in [1]. Performance and convergence of the estimated weights/coefficients of an adaptive filter purely depends on the type of adaptive filter used to estimate filter co-efficient, numbers of echo path and number of channels. This method is implemented and tested using the MATLAB tool.

Keywords

Dereverberation, speech enhancement, adaptive dereverberation, multi-channel adaptive dereverberation, blind dereverberation, adaptive filter.

INTRODUCTION

Enhancement of a speech signal is a very widely studied subject and much research has been done so far. Speech degrades mainly due to two reasons, additive uncorrelated noise and/or additive reverberated signals. This paper discusses mainly the reduction of the reverberated signal. The process of cancelling the reverberated signal is called de-reverberation. The de-reverberation method can be classified broadly as de-reverberation with prior knowledge of RIR and without. The de-reverberation method without prior knowledge of RIR is called blind dereverberation. This can be further classified into two classes. First-class, extract clean signal without estimating RIR [4][5]. This type of de-reverberation best works online, but performance will be poor for highly reverberated signals, i.e., low DRR (Direct to reverberant ratio).

The second class of methods uses an inverse filter of RIR. Estimation of RIR can be based on prior knowledge [2], this method works based on a statistical model of speech transmission and with the assumption of echo delay. The other method estimates RIR and applies inverse transformation [6], since it adopts a room transfer function without any prior knowledge this method is called adaptive de-reverberation. The paper [6] estimates the RIR or room regression coefficient (RRC) using the weighted RLS algorithm.

This paper proposes the method to cancel echo signal using multichannel by estimating RIR at each mic and by that it estimates the transmission delay of echo signal to 2nd mic, 3rd mic and so on with respect to mic 1. Here, mics are arranged in omni directional and desired signal shall be perpendicular to the mic, I.e., desired signal shall reach all mics without any delay. The performance of this method is much better compared to [1] because with this approach we need to extract approximation of RIR and estimate transmission delay. Where as in [1] better the estimation of RIR better the suppression.

Remaining of this paper is organized as follows; Section 2 discusses the basic principle to cancel single echo path using 2 channels. Section 3 discusses on estimation of delay and realization. Section 4 discusses effect of multi echo path on 2 channels method. Section 5 discusses de-reverberation using 3 or more channel for multi echo path. Section 6 and 7 discusses implementation and its result and finally section 8 concludes this paper.

DE-REVERBERATION OF SINGLE ECHO USING DUAL CHANNELS

Let's consider a two-mic being placed apart by few cm. A speaker stands in front of 2 mic such that transmission delay to 2 mics is same I.e., angle of speech signal incident is 0^0. An echo signal traverse to two mics at different interval as shown in below figure 1.

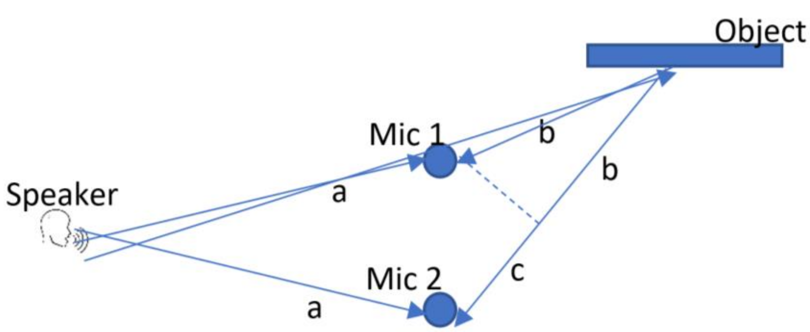

Figure 1: Pictorial representation of speaker, mic and object considered for echo cancellation.

Note that a, b and c are distances in cm.

Then signal at each mic in Z domain can be written as,

$$y1(z) = h1(z)s(z) + h2(z)s(z) \quad (1)$$

$$y2(z) = h1(z)s(z) + d1 * z^{-n1}h2(z)s(z) \quad (2)$$

whereas,

- y1(z) signal received at mic 1.
- y2(z) signal received at mic 2.
- h1(z) transfer function from speaker to mic 1 and mic2.
- s(z) clean signal
- h2(z) transfer function of echo signal at mic 1.
- $d1 * z^{-n1}$ is the transmission delay to second mic compared to first mic with attenuation factor of d1.
- Let d(z) is the direct signal such that,
- d(z) = h1(z) * s(z) (3)

Equation 1 and 2 can be written as

$$y1(z) = d(z) + h2(z)s(z) \quad (4)$$

$$y2(z) = d(z) + d1 * z^{-n1}h2(z)s(z) \quad (5)$$

The equations 4 and 5 can be written in matrix form as shown.

$$\begin{bmatrix} y1(z) \\ y2(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & d1 * z^{-n1} \end{bmatrix} * \begin{bmatrix} d(z) \\ h2(z)s(z) \end{bmatrix}$$

d(z) can be obtained by finding inverse matrix of transfer function as shown below,

$$\begin{bmatrix} d(z) \\ h2(z)s(z) \end{bmatrix} = \begin{bmatrix} \frac{d1 * z^{-n1}}{d1 * z^{-n1} - 1} & \frac{-1}{d1 * z^{-n1}} \\ \frac{-1}{d1 * z^{-n1}} & \frac{1}{d1 * z^{-n1}} \end{bmatrix} * \begin{bmatrix} y1(z) \\ y2(z) \end{bmatrix}$$

The direct signal d(z) can be extracted by multiplying the received signals with the inverse matrix as shown below,

$$d(z) = \frac{d1 * z^{-n1}}{d1 * z^{-n1} - 1} * y1(z) - \frac{1}{d1 * z^{-n1} - 1} * y2(z) \quad (6)$$

Since distance between mics are very small, we can neglect the attenuation factor d1 by setting it to 1 and hence equation 6 can be simplified as

$$d(z) = \frac{1}{z^{-n1} - 1} * \left(z^{-n1} * y1(z) - y2(z)\right) \qquad (7)$$

If we can find transmission delay of echo signal at 2nd mic w.r.t 1st mic, then it is possible to extract the desired signal.

ESTIMATION OF DELAY AND REALIZATION

Echo transfer function (h2) can be determined using the method described in [1]. Figure 2 depicts echo transfer function of echo signal received by mic1 and mic2 respectively. The transmission delay from mic1 to mic2 can be determined by determining the distance between peaks of two transfer function. This can be achieved by performing cross correlation. For example, length of transfer function of mic1 and mic2 is N taps then delay can be calculated by subtracting the peak index by N.

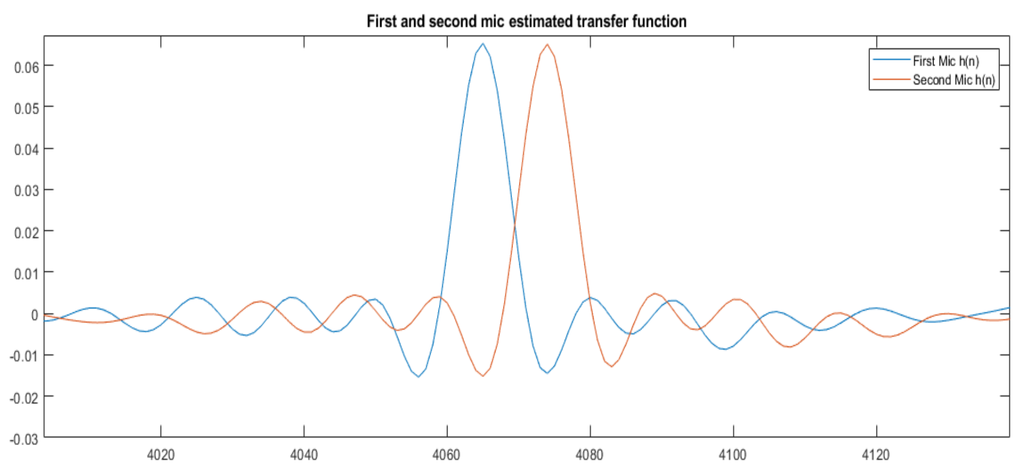

Figure 2: Estimated transfer function of echo signal received at mic1 and mic2.

Below diagram depicts the method for suppressing the single echo using 2 mics as per the equation 7.

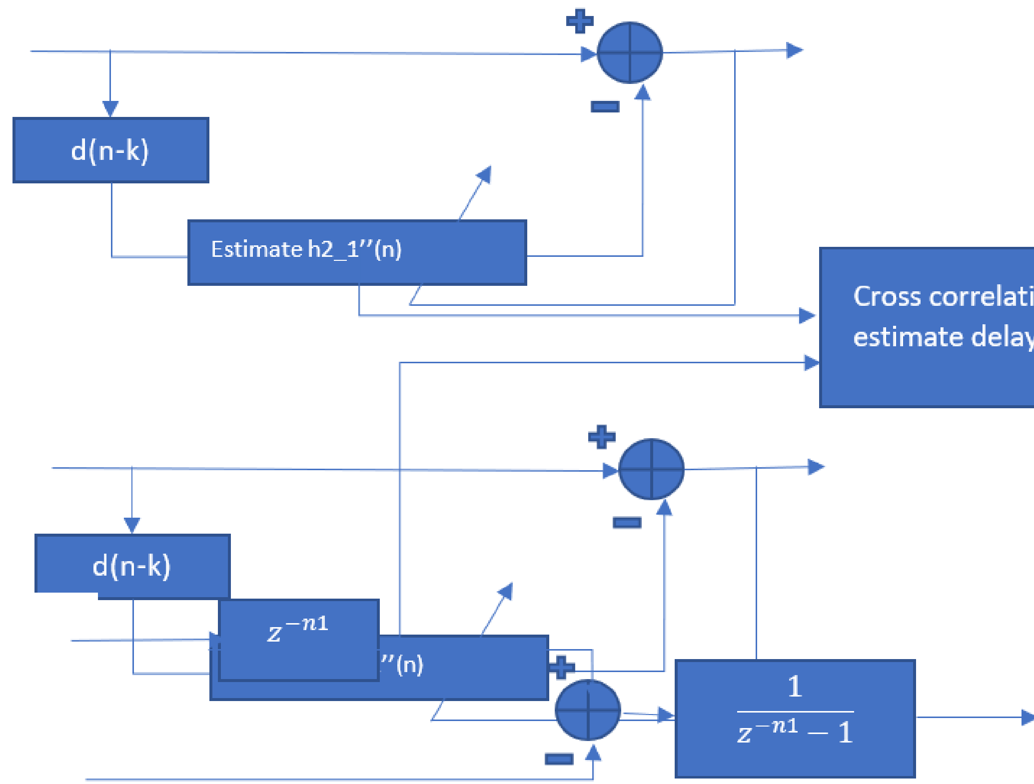

Figure 3: Realization of de-reverberation.

Substituting y1(n) and y2(n) into equation 7 and simplifying yields to

$$\mathrm{d'(z)} = \frac{1}{\mathrm{z^{-n1}} - 1}\left((\mathrm{z^{-n1}} - 1) * \mathrm{d(z)}\right) \tag{8}$$

In the above equation numerator $z^{-n1} - 1$ is a comb filter which suppress some frequency component completely and can't be recovered with inverse comb filter which is $\frac{1}{z^{-n1}-1}$ and this inverse comb filter is unstable. Hence this can't be realized in practise. To overcome from this issue, pass the y1(n) through delay filter with attenuation which is very near to 1 and use the same attenuation in inverse filter as shown below

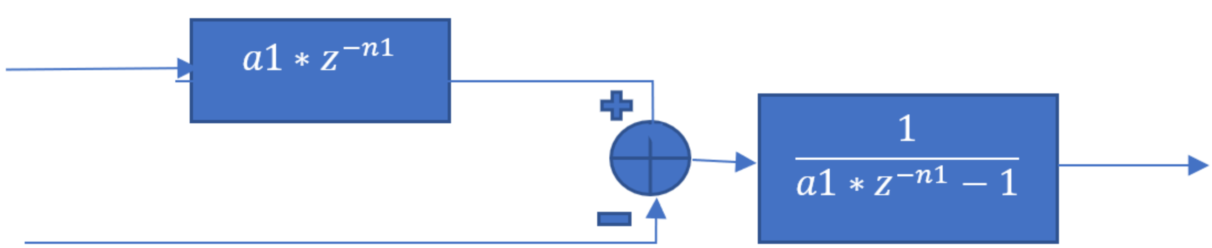

Figure 4: Practical implementation of de-reverberation

Hence replacing z^{-n1} by $a1 * z^{-n1}$ in equation 7 and substituting y1(n) and y2(n) yields to

$$d'(z) = d(z) + \frac{(a1-1) * z^{-n1}}{a1 * z^{-n1} - 1} * h2(z)s(z) \quad (9)$$

Since a1 chosen is near to 1 the effect second term is very minimal, and inverse comb filter will be stable.

EFFECT OF MULTI ECHO USING 2 CHANNELS

What happens when multiple echoes exist? Does this method still work? Let's take signal at mic 1 and mic 2 where 2 echo at mic1 and mic 2 hits at different interval of time.

$$y1(z) = x(z) + h2\_1(z) * s(z) + h2\_2(z) * s(z) \quad (10)$$

$$y2(z) = x(z) + z^{-n1}h2\_1(z) * s(z) + z^{-n2}h2\_2(z) * s(z) \quad (11)$$

Let's assume gain of h2_1 is larger than h2_2. Estimation of transfer function at mic gives 2 peaks as shown below.

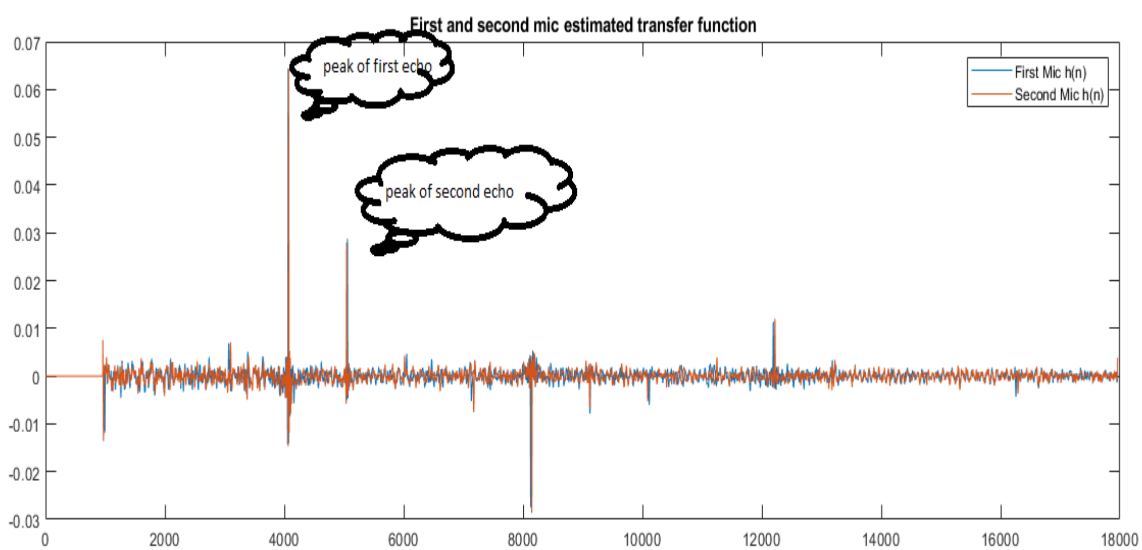

Figure 5: Estimated transformation at mic1 and mic2 for echo signal

Delay estimation algo will provide delay of highest amplitude peak. In the above figure h2_1 has highest peak. Hence it estimates delay for first echo. Passing the y1(n) through $a1 * z^{-n1}$ filter and subtracting with y2(n) gives,

$$d'(z) = d(z) + (a1 - 1) * \frac{z^{-n1}}{a1 * z^{-n1} - 1} * h2\_1(z) * s(z) + \frac{a1 * z^{-n1} - z^{-n2}}{a1 * z^{-n1} - 1} * h2\_2(z)s(z) \qquad (12)$$

where n2 is the delay at mic2 due to second echo. Since a1 is set to very near to 1 second term in equation 12 can be neglected but second echo still exist with transfer function $\frac{a1*z^{-n1}-z^{-n2}}{a1*z^{-n1}-1}$. Since gain of h2_2 is lesser than h2_1 the DRR will be improved considerably. This can be further improved by passing the d(n) to single channel auto dereverberation method [1] as shown below.

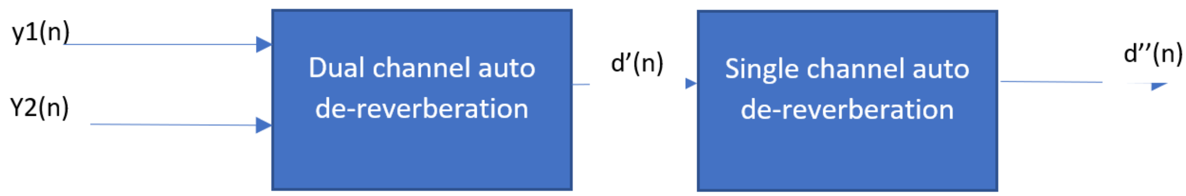

Figure 6: Multistage de-reverberation realization.

DE-REVERBERATION OF MULTI ECHO USING THREE CHANNEL AND ABOVE

Similar way to dual mic, we can use 3 mics to minimize 2 echo path, 4 mic for 3 echo path and so on. Let's consider for 3 mics. Which are separated by few cm apart in omni direction. Assuming there are two echo signals incidents 2nd mic and 3rd mic with respect to 1st mic. The signal captured at all 3 mics can be written in matrix form as,

$$\begin{bmatrix} y1(z) \\ y2(z) \\ y3(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & z^{-n1} & z^{-n2} \\ 1 & z^{-n3} & z^{-n4} \end{bmatrix} * \begin{bmatrix} x(z) \\ h2_{1(z)} * s(z) \\ h2_{2(z)} * s(z) \end{bmatrix} \qquad (13)$$

where n1 and n2 are transmission delay of 2 echoes at mic 2 and similarly n3 and n4 are transmission delay of 2 echoes at mic 3 w.r.t mic 1. Transmission delay n1 and n2 can be found by correlating estimated echo of mic 1 and mic 2 by finding two peaks. Similarly, n3 and n4 can be found out by correlating estimated echo of mic1 and mic3 by finding two peaks. Using the inverse matrix of equation (13) d'(z) can be extracted from y1(z), y2(z) and y3(z).

In a similar way, for higher echo path say N can be minimized using N+1 mics.

IMPLEMENTATION

This method is implemented in MATLAB tool. For this, a clean signal is collected from the sources. The room impulse response is simulated and same is used to generate reverberated signal from clean signal. Signal is sampled at 48 kHz and cancellation is done using 2 mics. Mics are spatially separated by 6 cm and kept in omni direction. Desired source is placed at 1 meter away from the centre of 2 mics. One or more objects are placed at random places and RIR is estimated for this setup to simulate echo signal. Test are conducted with one object, two objects and three objects keeping them at random places. LSM method is used to estimate RIR. The Figure 7 shows the flow chart of implementation for the discussed method and Figure 8 shows the plot of direct signal, reverberated signal at mic1, and enhanced signal for single echo path. The echo cancellation can be visually observed from the plot.

TESTS & RESULT

Tests are conducted with single echo path, two echo path and three echo paths. Below table shows the DRR using dual channel respectively. In this test, for the 2 and 3 echo paths, the output of dual channel is not passed through single channel auto de-reverberation method as explained in section 4.

Input/Output	Single Echo Path	Two Echo Path	Three Echo path
Input DRR	-6.173120	-6.173120	-6.173120
Output DRR	-0.000503	-0.977886	-1.000569

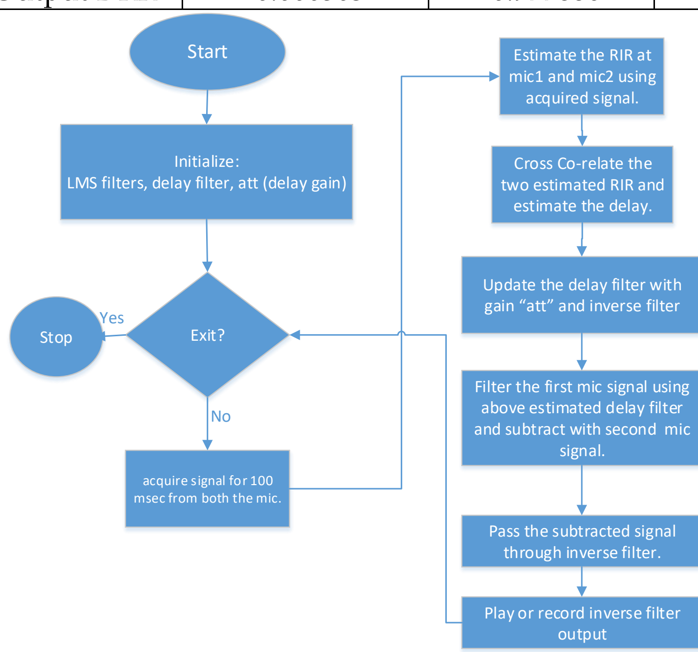

Figure 7: Flow chat for de-re implementation

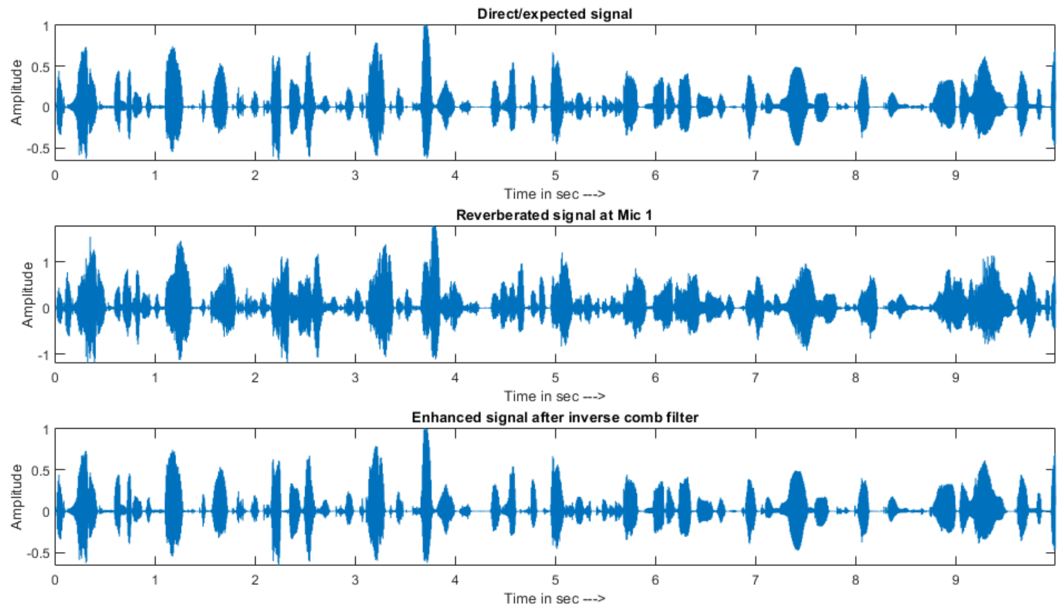

Figure 8: Audio signals at different places

CONCLUSION

The performance of this method is purely depending on algorithm used in the adaptive filter and transmission delay of echo path is much larger than direct path and number of mics greater than or equal to the number of echo signal. Practically it is not possible to estimate exact transmission delay due to digitalization of the system. The transmission delay estimation error TErr is given by 0 < TErr <1/(2*Fs) where Fs is sampling frequency. Also, lesser the distance between mic it is difficult to estimate transmission delay for smaller path. Hence, lesser the sampling frequency, distance between mics shall be more and vice versa. This method becomes more complicate for the higher number of mics I.e., greater than 2. Also, for better performance number of mics should be at least number of echo path. Choosing mics dynamically based on echo path also a challenge. Computational complexity and memory requirement will be more as this method demands processing at higher sampling rate.

REFERENCES

[1] Praveen P B, Dr Aravind H S, "Adaptive de-reverberation of speech signal" Design Engineering ISSN: 0011-9342, Issue 01, Page 3515-3527, 2022

[2] E.A.P. Habetts, " Multi-channel speech dereverberation based on a statistical model of late reverberation in 2005 Acoustic, speech, and signal processing (ICASSP) IEEE International conference,

[3] M. Delcroix, T.Hikichi, & M Miyoshi, “Blind dereverberation algorithm for speech signals based on multi-channel linear prediction,” Acoustic. Sci. Technol., vol. 26, no. 5, pp. 432–439, 2005.

[4] Harish Padaki, Karan Nathwani, and Rajesh M Hegde, “Multi-channel speech dereverberation using the LP residual cepstrum,” in Communications (NCC), 2013 IEEE National Conference. pp. 1–5.

[5] Harish Padaki, Karan Nathwani, and Rajesh M Hegde, “Single-channel speech dereverberation using the LP residual cepstrum,” in Communications (NCC), 2013 National Conference on. IEEE, 2013, pp. 1–I

[6] Takuya Yoshioka, Hideyuki Tachibana, Tomohiro Nakatani, and Masato Miyoshi, “Adaptive dereverberation of speech signal with speaker position change detection", in 2009 Acoustic, speech and signal processing (ICASSP) IEEE International conference,

[7] B. W. Gillespie, H. S. Malvar, and D. A. F. Florˆencio, “Speech dereverberation via maximum-kurtosis sub band adaptive filtering, “in Proc. Int’l Conf. Acoust., Speech, Signal Process., 2001, vol. VI, pp. 3701–3704.

[8] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, “Incremental estimation of reverberation with uncertainty using prior knowledge of room acoustics for speech dereverberation,” in Int’l Worksh. Acoust. Echo, Noise Contr., 2008, CD-ROM Proceedings.

[9] S. Haykin, Adaptive Filter Theory, Prentice Hall, fourth edition, 2001.