An Effective Approach to Estimate Reverberated Channel of Speech Signal

Arif Ahmed Tanim[1], Samiul Hasan[2], Md. Sharifur Rahman[3]
Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology,
Dhaka-1000, Bangladesh
tanim0123@gmail.com[1], samiul001@yahoo.com [2], shourav.115@gmail.com[3]

Abstract—**Nowadays the use of voice recognition system is increasing day by day. Dereverberation of speech signal is necessary to make this voice recognition system effective and smooth. Beamforming technique and an adaptive process are useful for blind channel estimation which can be used for de-reverberation process. Reverberated channel estimation is a very important step for de-reverberation process. This paper emphasis on reverberated channel estimation and introduce a method that can accelerate the reverberated channel estimation process of existing blind channel estimation techniques.**

Keywords- Reverberation, Channel; Delay; LMS algorithm; Estimation; NPM, Phase, Beam formng, Iteration, MATLAB

I. Introduction

De-reverberation is a signal processing technique to enhance the quality of reverberated speech signal that appears in confined environment (room, chamber). De-reverberation is a blind problem i.e., the source signal is unknown to the receiver. No training pulses can be sent to estimate the long AIR (Acoustic Impulse Response), typically consisting of thousands of coefficients. The AIRs are inherently time-varying and a slight movement of the head, which is natural during conversation, causes the AIR to be changed. At some point, the reflections' arrival rate exceeds even the sampling rate. These issues must be taken into consideration for an effective, efficient and robust de-reverberation technique [1].

For de-reverberation process of reverberated speech signal there are two main objectives. These are *Reverberated channel estimation* and *De-reverberation using the channel characteristic.*

Hagai Attias et. al. in their work on Speech De-noising and De-reverberation Using Probabilistic Models mention some approach of de-reverberation process. It states that the difficulty of speech enhancement depends strongly on environmental conditions. If a speaker is close to a microphone, reverberation effects are minimal and traditional methods can handle typical moderate noise levels. However, if the speaker is far away from a microphone, there are more severe distortions, including large amounts of noise and noticeable reverberation.

David Halupka et.al. in their work on 'Low-Power Dual-Microphone Speech Enhancement Using Field Programmable Gate Arrays' proposed Phase-Based Time Frequency Masking for the de-reverberation of speech signal[2]. Phase-based time-frequency masking is based on the following simple observation. Under ideal conditions (no noise, no reverberations, and a single speaker), the microphone signals observed during each fixed time interval will be related by here represents the speaker's location, that is, the speaker's produced signal's inter-microphone Time delay of arrival, τ (TDOA)[2].

$$angle(M_1(\omega)) - angle(M_2(\omega)) = \tau\omega$$

Where, $M_1(\omega)$ and $M_2(\omega)$ are transformed microphone signal. However, in noisy and/or reverberant environments, the two microphones signals are no longer strictly related, and the resulting phase-error is given by,

$$\theta(\omega) = angle(M_1(\omega)) - angle(M_2(\omega)) - \tau\omega$$

P. Aarabi et. al. showed that is proportional to the amount of the noise and reverberation corrupting the desired signal [3]. Hence, the phase error can be used to construct a time-varying filter that attenuates the amplitude of signal frequencies that have a high phase error [4]. P. Aarabi et. al. [4] proposed a time-varying phase-error-based filter for de-reverberation by inserting a frequency-dependent weight vector.

E.A.P. Habets in his works on Single Channel Speech De-reverberation process derived model for Room Impulse Response as well as Reverberation Signal Model [5]. They are based upon channel identification to determine the Room Impulse Response (RIR) between the source and the receiver and use this information to equalize the channel. However, de-convolution methods have been shown to be little robust to small changes in the RIR. The model we use was developed by Polack [6]. This model describes a Room Impulse Response (RIR) as one realization of a non-stationary stochastic process.

II. Problem Formulation

A. Beamforming using two microphone system:

Multichannel Least-Mean-Square (MCLMS) algorithm is an effective, simple and commonly used method for blind channel estimation like reverberation[1][7]. The algorithm is based on adaptive estimation and using beam-forming techniques MCLMS can be implemented. Beam-forming is a signal processing technique used in sensor arrays for directional signal transmission or reception. This is achieved by combining elements in the array. This is sending the same signal to all transducers but with varying information encoded by frequency in that signal, requiring a broadband signal - this is called using a "blazed array". Incoming acoustic waves will generally arrive at the different microphones with slightly different times. MCLMS suffers from slow convergence rate. In this paper we will emphasis on this shortcoming of MCLMS and use two microphones for beam-forming.

For channel estimation introducing two microphones is an effective way to estimate reverberated channel. A speech source is received by two identical microphones. These microphones are kept at two different distances from the source. That is equivalent de reverberated channel for two different microphone is not the same.
The overall condition is depicted on fig.1:

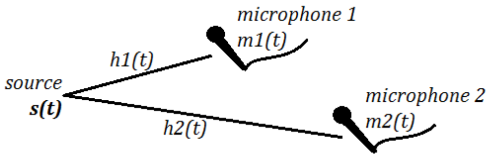

Figure 1: channel estimation using two microphones

Here,
$s(t) =$ Source signal
$h_1(t) =$ Reverberated channel for microphone 1
$h_2(t) =$ Reverberated channel for microphone 2
$m_1(t) =$ Received signal received by microphone 1
$m_2(t) =$ Received signal received by microphone 2

Let introduce a quantity such that,
$e = m_1(t) * d_1(t) - m_2(t) * d_2(t)$
Where, $d_1(t)$ and $d_2(t)$ are matrix of the dimension of $h_1(t)$ and $h_2(t)$ respectively, arbitrarily chosen. And '*' denotes convolution operator.

$$\Rightarrow e = s(t) * h_1(t) * d_1(t) - s(t) * h_2(t) * d_2(t)$$
$$\Rightarrow e = s(t) * [h_1(t) * d_1(t) - h_2(t) * d_2(t)]$$

To estimate channel response we need to choose $d_1(t)$ and $d_2(t)$ such that, $e = 0$

$$\Rightarrow s(t) * [h_1(t) * d_1(t) - h_2(t) * d_2(t)] = 0$$
$$\Rightarrow [h_1(t) * d_1(t) - h_2(t) * d_2(t)] = 0$$
$$\Rightarrow h_1(t) * d_1(t) = h_2(t) * d_2(t)$$

Solution of this relation occurs when,
$d_1(t) = h_2(t)$ and
$d_2(t) = h_1(t)$

By some initial assumption of $d_1(t)$ and $d_2(t)$ an adaptive process may be done for determining $d_1(t)$ and $d_2(t)$. That will be the estimation of reverberated channel $\hat{h}_1(t)$ and $\hat{h}_2(t)$. This paper of emphasis on the estimation of $\hat{h}_1(t)$ and $\hat{h}_2(t)$ and introduced a better method of the estimation of reverberated channel of existing blind channel estimation techniques.

B. LMS Algorithm for Blind Channel Identification:

Least mean algorithm (LMS) is a very common method of adaptive algorithm. The least mean square (LMS) algorithm is a linear adaptive filtering algorithm that consists of two basic algorithms-
A filtering process, which involves, (a) Computing the output of a transversal filter produced by a set of tap inputs and (b) Generating an estimation error by comparing this output to a desired response.

An adaptive process, which involve the automatic adjustment of the tap weights of the filter in accordance of the estimator (fig.2)

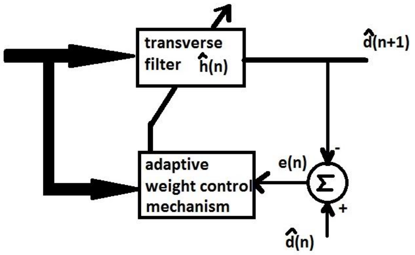
transverse filter $\hat{h}(n)$
$\hat{d}(n+1)$
adaptive weight control mechanism
e(n)
Σ
-
+
$\hat{d}(n)$

Figure 2: Block diagram for LMS algorithm

Generally a source signal passes through a reverberated environment. A reverberation is a sum of different delayed version of an actual signal (source). This source signal is made reverberant. For MATLAB implementation a speech signal (.wav) is considered as source. Then it is made reverberant by mounting some reverberating filter function with the source. We used a .wav file of 1 minutes and 40 seconds duration as our source. It was made reverberant by mounting reverberating functions $h_1(t)$ and $h_2(t)$ where $h_2(t)$ is some delayed version of $h_1(t)$. And finally got two reverberated signals, $m_1(t)$ and $m_2(t)$.

Thus, source signal, reverberated channel, output (reverberated signal) and delay between reverberated signals is initially known. Thus, both $m_1(t)$ and $m_2(t)$ can be used to estimate channel characteristics.

C. Introduce delay: A better estimation

We recall the equation-

$e = m_1(t) * d_1(t) - m_2(t) * d_2(t)$

For the estimation of $\hat{h}_1(t)$ and $\hat{h}_2(t)$ iteration is initiated by assuming initial $d_1(t)$ and $d_2(t)$ arbitrarily. This assumption is generally done by setting both $d_1(t)$ and $d_2(t)$ a matrix, $[1,0,0,0,0,.........]$. But this can be speeded up by making a very simple technique. Here, the microphones are not placed at equal distance and a delay exists between them. Eventually reverberated channel will also experience this and there must be some delay between two channels, $h_1(t)$ and $h_2(t)$. Eventually both the estimations, $\hat{h}_1(t)$ and $\hat{h}_2(t)$ will also have this delay between them and we are inserting the 'delay' at $d_1(t)$ and $d_2(t)$ where initialization of iteration begins. So Instead of setting both the same matrix we will set as below-

$d_1(t) = [1,0,0,0,0,.........]$ and,

$d_2(t) = [0,0,0,0,1,0,0,0,0,.........]$

Here 4 zeros $\underline{0,0,0,0}$ are placed as delay value.

D. Delay estimation:

A simple model for the signals observed by two microphones in a non-reverberant environment is given as-

$m_1(t) = s(t) + n_1(t)$

And,

$m_1(t) = s(t+\tau) + n_1(t)$

Where $s(t)$ is the signal source of interest, and $n_1(t)$ and $n_2(t)$ are used to model microphone noise, environmental noise, and possibly signals from other speakers. Attenuation of $s(t)$ has been neglected, which is a reasonable assumption when the inter microphone distance is much smaller than the source-to-microphone distance. The goal of sound localization is to use the observed signals $m_1(t)$ and $m_2(t)$ to deduce τ, the TDOA of the source between the two microphones.

Typical acoustic environments are reverberant and therefore cause signal-correlated noise. This complex environment can be modeled as-

$$m_1(t) = h_1(t) * s(t) + n_1(t)$$

And,

$$m_2(t) = h_2(t) * s(t) + n_2(t)$$

Where, $h_1(t)$ and $h_2(t)$ are the impulse responses of the environment with respect to each microphone's position. Noise that is not due to the signal reverberation is modeled by $n_1(t)$ and $n_2(t)$.

Phase delay between microphones at clean environment (no noise/no reverberation)-

$angle(M_1(\omega)) - angle(M_2(\omega)) = \tau\omega$

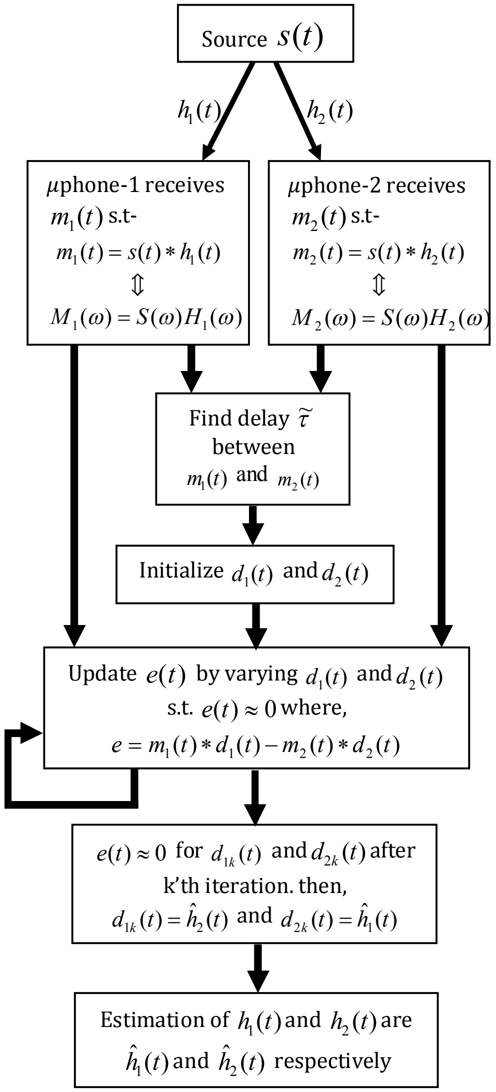
Source $s(t)$
$h_1(t)$
$h_2(t)$
μphone-1 receives $m_1(t)$ s.t-
$m_1(t) = s(t) * h_1(t)$
$\Updownarrow$
$M_1(\omega) = S(\omega)H_1(\omega)$
μphone-2 receives $m_2(t)$ s.t-
$m_2(t) = s(t) * h_2(t)$
$\Updownarrow$
$M_2(\omega) = S(\omega)H_2(\omega)$
Find delay $\tilde{\tau}$ between $m_1(t)$ and $m_2(t)$
Initialize $d_1(t)$ and $d_2(t)$
Update $e(t)$ by varying $d_1(t)$ and $d_2(t)$ s.t. $e(t) \approx 0$ where,
$e = m_1(t) * d_1(t) - m_2(t) * d_2(t)$
$e(t) \approx 0$ for $d_{1k}(t)$ and $d_{2k}(t)$ after k'th iteration. then,
$d_{1k}(t) = \hat{h}_2(t)$ and $d_{2k}(t) = \hat{h}_1(t)$
Estimation of $h_1(t)$ and $h_2(t)$ are $\hat{h}_1(t)$ and $\hat{h}_2(t)$ respectively

Figure 3: Flow chart for Channel estimation by inserting delay

But when the environment is not perfectly clean (like reverberant situation) then, Phase delay between microphones is-

$$\theta(\omega) = angle(M_1(\omega)) - angle(M_2(\omega)) - \tau\omega$$

Phase error $\theta(\omega)$ arises due to reverberation (same for noisy situation too) Here τ is a variable that indicates time. Let denote this variable τ as β then-

$$\theta(\omega) = \angle M_1(\omega) - \angle M_2(\omega) - \omega\beta$$

Now introduce a variable τ such that -

$$\tilde{\tau} = \arg\max_{\beta} \int_{-\infty}^{\infty} \cos(\angle M_1(\omega) - \angle M_2(\omega) - \omega\beta) d\omega$$

Here $\tilde{\tau}$ is nothing but the value of the integration value.

If we vary β then $\tilde{\tau}$ will also be varied. That means for some value of β once we will get the maximum value of $\tilde{\tau}$. That means-

$\cos(\angle M_1(\omega) - \angle M_2(\omega) - \omega\beta)$ maximizes. Eventually,

$(\angle M_1(\omega) - \angle M_2(\omega) - \omega\beta)$ minimizes. Eventually,

$(\angle M_1(\omega) - \angle M_2(\omega) - \omega\beta) \rightarrow 0$

$\omega\beta = \angle M_1(\omega) - \angle M_2(\omega)$

This β indicates delay [4].

The total process is implemented on MATLAB based on below flow chart (fig.3).

E. De-reverberation of Speech Signal from estimated channel

If a signal $s(t)$ is allowed to go through a reverberated environment and received by a microphone as $m(t)$ then, mathematically-

$$s(t) * h(t) = m(t)$$

$$\Rightarrow S(\omega)H(\omega) = M(\omega) \qquad ;[m(t) \Leftrightarrow M(\omega)]$$

As $\hat{H}_1(\omega)$ is estimated reverberant environment so we can expect reconstructed signal as-

$$s(\omega) = \frac{M_1(\omega)}{\hat{H}_1(\omega)}$$

And convolution between inverse filter and observed reverberated signal may give actual signal.

The time-varying phase-error-based filter proposed by G. Shi and P. Aarabi [4] is,

$$Y(\omega) = \psi(\omega) M_1(\omega)$$

With the frequency-dependent weight given by

$$\psi(\omega) = \frac{1}{1 + \gamma\theta^2(\omega)}$$

It is assumed that $\theta(\omega)$ is wrapped to be in the range *[-π, π]*. The term γ is an adjustable parameter that controls the aggressiveness of the filter. In low-SNR conditions, a high value of γ is favorable, whereas in high-SNR conditions, a low value of γ is favorable as a high value of γ will actually corrupt the signal of interest[3][4]. It was shown that $\gamma = 5$ results in good speech enhancement; the results presented in this paper utilize this fixed value of γ. However, this parameter can be adjusted via configuration pins or a programmable register.

III. SIMULATIONS AND RESULTS

We have generated two sample delay between the reverberated channels $h_1(t)$ and $h_2(t)$. We applied the above techniques to measure delay. From fig.4 maximum integral value occurs when β=2. So, time delay of arrival, $\tilde{\tau} = 2$

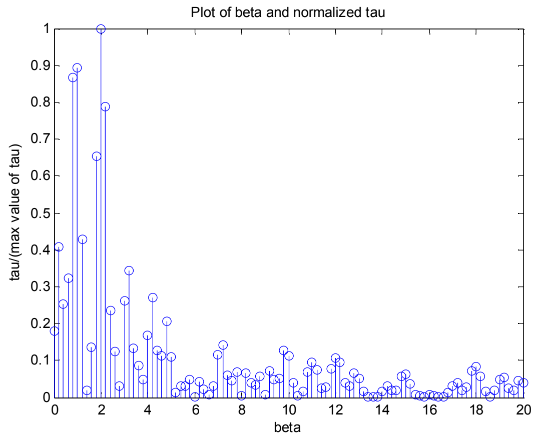

Figure 4: Integration (abs) Vs beta graph

Two different methods were used for channel estimation. These are- (a) Estimation without inserting delay and (b) Estimation by inserting delay

A comparative analysis between these methods is depicted at fig.5. Normalized Projection Misalignment (NPM) of channel estimation vs. iteration number of the two different methods is shown in the figure in same scale by MATLAB simulation. Here the dotted curve (line 1) indicates the situation where delay is not inserted. Plane curve (line 2) indicates the situation where delay is inserted. It is obvious that, NPM decades first by inserting delay during initialization. Eventually channel estimation gets faster.

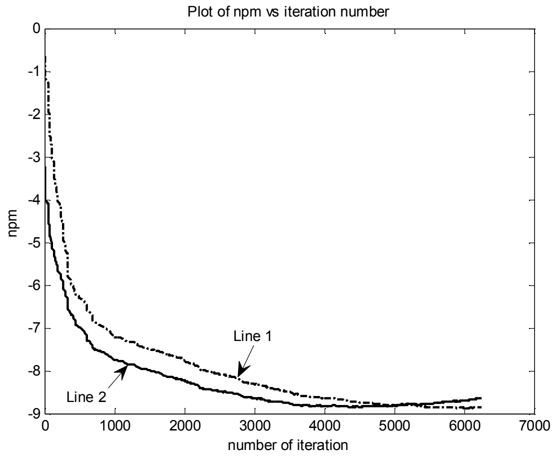

Figure 5: Estimation with and without using delay

From fig.6 we get an npm of -8.71 after 4342 iteration (line 1) while the same npm arrives after only 3204 iteration (line 2) by inserting delay at initialization of estimation. Inserting delay gives a better estimation as it requires less iteration and is computationally efficient.

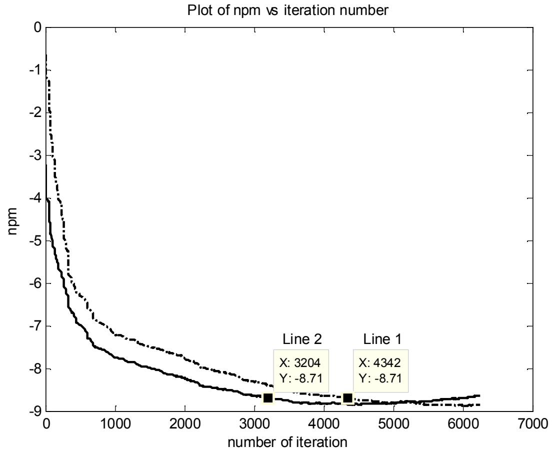

Figure 6: Estimation with and without using delay

Table: 1 Simulation output: Iteration Vs NPM

No. of iteration	NPM (without using delay) (a)	NPM (using delay) (b)
500	-6.306	-6.994
1000	-7.188	-7.741
1500	-7.486	-7.996
2000	-7.754	-8.234
2500	-8.088	-8.485
3000	-8.313	-8.649
3500	-8.485	-8.756
4000	-8.644	-8.822

Table 1 is the comparison of NPM value between two approaches at different iteration level. Inserting delay ensures the NPM to decade earlier that accelerates the estimation of reverberant environment.

IV. Conclusion

In this paper we presented an approach to estimate the de-reverberated channel more efficiently. We used two microphones for beam-forming and inserted microphones relative position (delay). That can improve iteration process during estimation of beam-forming methods.

De-reverberation was conducted by directly inversing channel with the reverberated signal and a slight improvement was observed by physical observations. But, this reconstructed signal got some new problems. This signal was not smooth. De-reverberation was not conducted over all the data point simultaneously. A group or data sequences are taken into consideration for processing. This is also needed for continuous processing of a signal especially in real time environment like broadcasting commentary or in teleconference. For above mentioned reasons filter length has a limitation. After each consecutive processing processed data are not matched at the beginning and end. And nonlinearity is observed. We propose an IIR based filter in this regard.

During channel estimation, a noisy reverberated channel is not considered. Noise component will consist of microphone noise as well as Gaussian noise of channel. This noise component can be considered at further analysis. Also de-reverberated signal needs to be analyzed using Signal-to-Reverberation Ratio (SRR).

V. Acknowledgement

The authors wish to express their heartiest gratitude to Dr. Mohammad Ariful Haque, Associate Professor, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka for his guidance, inspiration, devotion, close supervision throughout the entire period of the work.

VI. References

[1] Mohammad Ariful Haque, Toufiqul Islam and Md. Kamrul Hasan, “Robust Speech Dereverberation Based on Blind Adaptive Estimation of Acoustic Channels”, IEEE transactions on audio, speech, and language processing, ISSN 1558-7916, 2011, vol. 19, no4, pp. 775-787 [13 page(s) (article)] (24 ref.)

[2] David Halupka, Alireza Seyed Rabi, Parham Aarabi, Ali Sheikholeslami, “ Low-Power Dual-Microphone Speech Enhancement Using Field Programmable Gate Arrays”, IEEE transactions on signal processing, vol. 55, no. 7, July 2007

[3] P. Aarabi and G. Shi, “Phase-based dual-microphone robust speech enhancement,” IEEE Trans. Syst., Man, Cybern., vol. 34, no. 4, pp. 1763–1773, Aug. 2004

[4]G. Shi and P. Aarabi, “Robust digit recognition using phase-dependent time-frequency masking,” in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Hong Kong, Apr. 2003, pp. 684–687

[5] Single Channel Speech De-reverberation based on Spectral Subtraction by E.A.P. Habets Technische Universiteit Eindhoven, Department of Electrical Engineering, Signal Processing Systems Group, EH 3.27, P.O. Box 513Eindhoven, The Netherlands

[6] A. Dembo and O. Zeitouni. “Maximum a posteriori estimation of time-varying ARMA processes from noisy observations”. IEEE Trans. Acoustics, Speech, and Signal Processing, 36(4):471–476, 1988.

[7] Y. Huang and J. Benesty, Adaptive multi-channel least mean square and Newton algorithms for blind channel identication," Signal Process, Aug. 2002, vol. 82, no. 8, pp. 11271138.