Simultaneous Suppression of Noise and Reverberation by Applying a Two Stage Process

[1]Marjan Joorabchi [2]Seyed Ghorshi [3]Ali Sarafnia

[1, 2]School of Science and Engineering, Sharif University of Technology, International Campus
Kish Island, Iran
[3]Department of Electrical and Computer Engineering, Concordia University, Montreal
Quebec, Canada
[1, 2]{joorabchi, ghorshi}@kish.sharif.edu, [3]sara_ali@encs.concordia.ca

Abstract: - In speech processing, reflections of sound wave in a bounded space have considered as speech reverberation. Although for musical instruments and their related recording devices these reflections are useful, however, some other applications face serious problems receiving them along with speech signal. Reverberation causes speech degradation and unintelligibility as well as quality reduction. Dereverberation algorithms are essential for Automatic Speech Recognition (ASR), telecommunication and hearing aid devices, which are some of the mostly used applications. Even though dereverberation itself is a challenging issue, dereverberating a speech signal recorded only by one microphone (or channel) requires more attention and resolution. Aside from that, in acoustic environments, interference signals such as White Gaussian Noise and any other type of noise decrease the quality of received signal even more. In this paper, we propose a two stage algorithm for enhancement of reverberant speech from one microphone recordings. In the first stage, Kalman filter has been employed to reduce the effects of noise on received speech signal. In this stage, different values of SNR of additive White Gaussian Noise have been practiced on a reverberated speech signal. Furthermore, the results of noise reduction stage have been compared to show the grade of effectiveness of each adaptive filter. In the second stage, an inverse filter has been applied for reducing the reverberation's effect on received speech signal.

Key-Words: - Speech Signal, Reverberation, Noisy Acoustic Environments, Single-Channel, Dereverberation, Adaptive Filter.

1 Introduction

Unlike echo that happens in at least one-sided open acoustic environment, reverberation is a phenomenon that occurs in enclosed rooms. The propagation of speech waves in these bounded spaces (rooms) and by hitting the surfaces of the room and many objects within the room will cause different time delays between different paths that waves travel to reach recording devices (receiver) or listeners. Although these time delays may seem unimportant individually, by the time they reach to the microphone of objective receiver the combination of these waves will effectively degrade the quality of received speech signal and in some cases the intelligibility of speech signal may highly degrade as well [1,2]. These delays and types of changes in speech quality may not be that severe if the receiver side of speech is a human being, however, for impaired person who uses hearing aids, speech dependable applications (i.e. ASR, Speech authentication, Speech-to-text, etc.) telecommunication, Video conferencing, and many other applications and devices the quality and intelligibility of received speech signal are extremely important [3]. If any of these two features does not meet the basic requirements of each specific application, it may cause some problems for the person and/or provide inaccurate results. Therefore, dereverberation algorithms are necessary to suppress and/or reduce the effects of reverberation from speech signals in any reverberant environment.

On contrary to the speech dependable devices, music takes unlimited advantage of reverberation effects. It is desirable for musicians to play and record tracks with more than average level of reverberation. The more reverberated sounds they have produced, the more audiences they could attract. As the audiences' points of view, it seems also that they get excited and enjoy more [4].

Due to the monologue nature of the music and use of amplifying devices and loud speakers, reverberation effects may not be too obvious. However, in cases that a sentence should be analysed in order to identify a person or to speed up process of writing employees using speech-to-text feature of an application even a smallest interference (i.e. reverberation, noise, etc.) in speech signal, which leads to the quality reduction that would make highly increase the error rate and/or completely provide inaccurate outcomes [5]. In recent decades that the importance of speech quality is known to researchers, there are many dereverberate techniques and approaches have been proposed so far to at least reduce the reverberation's destructive effects.

Noise is also one of indisputable types of interference signals that in any fields of science researchers try to eliminate its effects. In speech processing unlike other fields aside from good quality that is the first and most aspiration of noise reduction methods we insist on obtaining intelligible speech signals as well. The importance of intelligibility may make it even more crucial than quality itself. Therefore, in this work which is about a two-stage approach to dereverberate speech signal, we made use of additive noise that may exist in daily life conversations. To ensure how effective the applied method of noise reduction is also different Signal-to-Noise Ratios (SNRs) have been added to an artificially extremely reverberated speech signal. The results of adaptive filters used in the first stage (e.g. noise reduction procedure) of the proposed approach will be compared in order to provide better understanding of the decisions that have been made to employ such methods in each stage of the proposed approach. The proposed method offers a two-stage process for de-noising and dereverberating a drastically distorted noisy reverberated speech signal. The dereverberation procedure employs an inverse filter in the second stage in order to reduce reverberation effects on observed speech signal.

There are many different methods to either suppress or reduce the effects of reverberation. Few of them have been designed to reduce the existing noise on the observed speech signal as well. In general dereverberate speech algorithms are divided into Blind and Non-Blind approaches. Methods that extract and estimate every parameter that are needed for their calculation from received speech signal are Blind approaches. They basically have no prior knowledge of the reverberant room and any other known information of the recorded speech. Non-Blind approaches on the other hand, use prepared information such as Reverberation Time (RT), source and receiver's position in the room, and other parameters that are related to their desired acoustic environment [6]. The number of microphones in receiver side is also quite important. The more observed speech signals received that are related to a specific time, the estimated parameters would be more accurate. Therefore, compare to Multi-Microphone (two and more) or channels recorded speeches, dereverberation of a speech signal that has been received by just one microphone (single acoustic channel) is more challenging [7].

The following are brief reviews of the proposed dereverberation methods, which few of them are also designed to de-noise the received speech signals. Cepstrum based method in [8] uses the phase of signal and cepstral operation. Wavelet based filtering method [9] applies, both, weighted wavelet reverberant coefficients and parameters, by taking magnitude and general Spectral Subtraction in [10,11]. Maximum Kurtosis in [12] controls the filter by considering the use of LP residual kurtosis. Modulation Transfer Function (MTF) based dereverberation [13] is an adaptive time-frequency division technique. Maximum Likelihood [14] employs a linear filter to suppress the reverberation and then uses a nonlinear de-noising process. Transfer Function estimation with Overestimated Order blindly (without knowing exactly the channel's order) dereverberates the observed speech signal [15]. An online algorithm that applies Natural Gradient deconvolution [16] employs Multiples input/output Inverse Theory (MINT) in order to have equalizer impulse response. Weighted Recursive Least Squares (RLS) [17] is an adaptive algorithm that provides fast convergence rate and also affective update rule. Temporal and Spectral combination in [18] is employed to suppress and attenuate the early and late reverberant components respectively. Harmonic Structure and Fundamental Frequency [19,20] is also a blind algorithm that operates in frequency domain with vast database of utterances of the desired language. Linear Prediction in [21] with no whitening problem effects blindly dereverberates observed signals that are outputs of an Auto Regressive (AR) procedure. The proposed method in [22] uses selective-tap identification and has low computational overhead that is based on Multi Channel Least Mean Square (MCLMS). Adaptive Minimum Mean Square Error estimator (AMMSE) [23] in noisy reverberant environment with various SNRs reduces the effect of late reverberation component. Partial multichannel equalization by MINT [24], Non-Casual Minimum Variance Distortionless Response (MVDR) [25], Binaural Cues [26], Code Excited Linear Prediction (CELP) postfilter [27], Deep Neural Network (DNN) [28], Reverberation-Time-Aware approach [29], Complex Ratio Masks [30], and Coherence Matrix Estimation [31] are also few of many other dereverberation algorithms that have been proposed till now in order to precisely reduce reverberation effects on recorded speech signals.

In this work the proposed method is designed for de-noising and dereverberating an extremely noisy reverberated speech signal that is received by one microphone. In order to produce such signal the Image-Source Model [32] has been used. This model creates images of a defined source of sound [33]. The first step of the proposed method employs Kalman filter to reduce the noise then the de-noised reverberated speech enters into the second step to be dereverberated. In order to have a clean and high quality speech the source sound has been taken from TIMIT database [34]. Various additive noises have also been added with different SNRs to display the capability of each stage of the proposed method. The dereverberation step, which is in second stage, applies an inverse filter to reduce the effects of reverberation on speech signal.

The reminder of this paper is organized as follows. In section 2 the procedure of reverberation will be described. The Kalman filter model for noise reduction is presented in section 3. Section 4 described the dereverberation process. The experimental and evaluation results are presented in section 5. Finally section 6 concludes the paper.

2 Reverberation Procedure

As previously mentioned reverberation is a destructive effect on speech signals that drastically reduces the quality and intelligibility of speech. This effect happens in enclosed spaces where the sound waves hit the rigid objects and reflect. The reflections, which are the delayed version of the source sound, degrade the quality of speech signal. The reverberated speech signal at microphone can be written as [35],

$$x(m) = \sum_{k=0}^{L_R-1} s(m-k)h_r(k,m) \qquad (1)$$

where m is the discrete time index, $s(m-k)$ is source speech and its delay versions, L_R is the length of RIR and h_r is impulse response of acoustic environment. Reverberation consists of three components such as Direct Path that is the shortest path sound wave travel to reach the microphone only when the microphone is in sight of source. The two other components are reflections that arrive at microphone by few milliseconds delays from direct path and each other, respectively. The first of these reflections to reach the microphone are early reflections. Other reflections that arrive after early reflection are late reflections with larger delay ranges that specifically effect intelligibility and fidelity of speech signal [36]. By taking these components into consideration Equation (1) can be written as follows,

$$x(m) = \sum_{k=0}^{L_e-1} s(m-k)h_r(k,m) + \sum_{k=L_e}^{L_R-1} s(m-k)h_r(k,m) \qquad (2)$$

and briefly it can be expressed as,

$$x(m) = x_e(m) + x_l(m) \qquad (3)$$

where the first part of the equation is early $x_e(m)$ and the second part is the late reverberation $x_l(m)$. The room characteristics are also very important in a way that each object in the room and even the materials that make the room are able to absorb part of sound wave's energy and also the reflection of the rest of sound wave depends on the material structure. As the absorption of the materials increases there will be more attenuation on speech signal and if the room's materials absorb sound waves considerably that room will be called a "dead" room [37]. Therefore, the Reverberation Time or RT_{60} is defined as length of time that the energy of the signals degrades 60 dB below the initial or original value of the energy of the signal.

To produce the desired reverberated speech signal, we considered a 90 m^3 enclosed space as the reverberant environment where the RT of the produced speech is 0.84 s. There is only one acoustic channel and the source of the sound and the receiver are stationary and neither of them moves.

Fig. 1 shows the clean speech signal as the input to the reverberation production system and the noise-

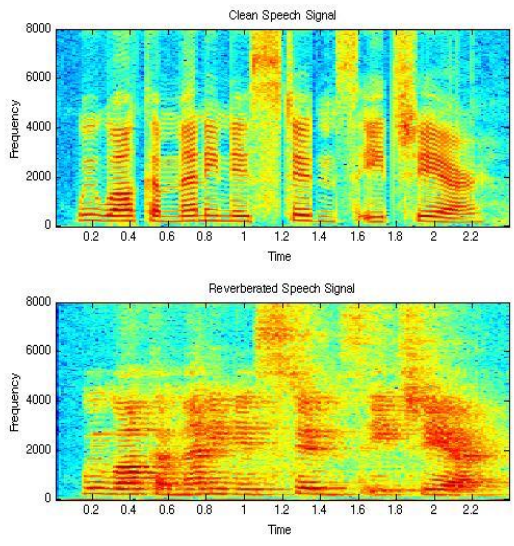

Fig. 1. Clean speech signal (Top Panel), reverberated speech signal (Bottom Panel).

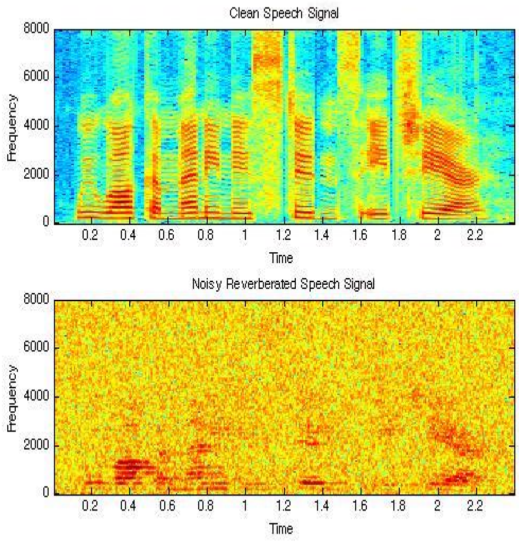

Fig. 2. Clean speech signal (Top Panel), noisy reverberated speech signal with 0 dB SNR (Bottom Panel).

free reverberated speech as the output. As it can be seen in Fig. 1 the reverberation in the speech signal (bottom panel) caused extreme distortion and dropped the quality of speech signal significantly. However, in Fig. 2 it is clearly observed that by adding noise to the reverberation speech signal the distortion will be more drastic than in Fig. 1. Also, in Fig. 3 the spectrums of clean speech, noise-free reverberated and noisy reverberant signals have been provided and by comparison it is clear that the noisy reverberant speech signal is exceedingly distorted.

3 Noise Reduction

The noise in acoustic environment also causes severe quality, intelligibility and fidelity damages.

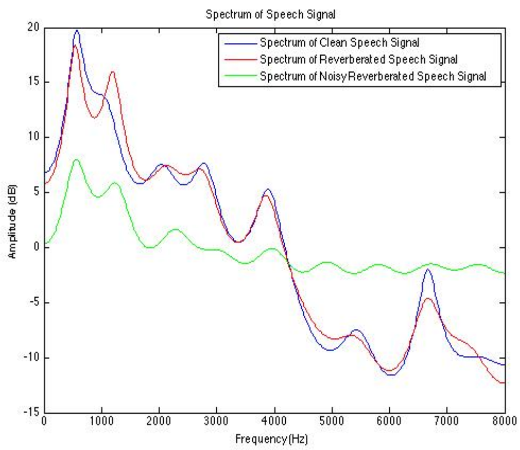

Fig. 3. Spectrums of clean speech signal, reverberated and noisy reverberated speech signal with 0 dB SNR.

Sometime this noise has unwillingly been produced by electronic devices' fans (i.e. air conditioners and computer fans) or whispers of people talking to each other in enclosed spaces. In the case of noisy reverberated speech signal, the received speech signal that is observed at microphone contains both reverberation and noise at the same time [7]. The noisy reverberated speech signal can be produced by adding noise in reverberant environment to the speech signal and it can be expressed as follows,

$$y(m) = x(m) + N(m) \tag{4}$$

where $y(m)$ is the noisy reverberated speech signal, $x(m)$ is the reverberant speech and $N(m)$ is the additive noise in the reverberant room. Removing noise from the observed signal requires an effective noise suppression technique. Kalman filter algorithm is one of the methods that can be used to reduce the effect of noise. Kalman filter is a Bayesian recursive model that has been used to estimate the desired signal from a noisy signal. It is widely used in noise reduction, prediction and system identification applications. It is a state-space approach that uses the state equation for modelling the noisy distorted observed signal. In the first stage of our proposed method which is the noise reduction stage we have considered the Auto Regressive (AR) model of speech signal that can be defined as,

$$x(m) = A_x x(m-1) + e(m) \tag{5}$$

$$x(m) = \sum_{I=1}^{P} a_i\, x(m-i) + e(m) \tag{6}$$

where a_i is AR's coefficient vector of P^{th} order [38], $x(m)$ is the speech signal and it is a $P \times 1$ dimensional matrix, A_x is also another matrix with $P \times 1$ dimensions and displays the state transition at times $m-1$ and m. The P dimensional matrix $e(m)$ is uncorrelated input excitation vector of the state equation. In this work after the reverberated speech signal has been produced the White Gaussian Noise (WGN) has been added with various SNR values ranging from -5 to 15 dB in order to produce a noisy reverberated speech signal. Then the noisy reverberated speech signal fed into Kalman filter for reducing the effects of noise. The aim of adding various SNRs is to show that this approach of noise reduction works perfectly with different ratios of daily life interference signals. In this work the noisy reverberated speech signal can be generated by Equation (7)

$$Noisy\ Reverberated\ Speech = Reverberated\ Speech + (\beta * Noise) \tag{7}$$

where the parameter β can be expressed as,

$$\beta = c * P_{Noise} \tag{8}$$

where c is the coefficient that picks a value from 0 to 10000 and P_{Noise} is the power of noise which can be expressed as

$$P_{Noise} = E[(N(m))^2] \tag{9}$$

where P_{Noise} is power of noise and $N(m)$ is the additive noise. If we assumed that the additive noise is a zero mean Gaussian random process. Then, the noise power can be represented in form of noise variance

$$P_{Noise} = variance\ (N(m)) \tag{10}$$

For the noise reduction process some prior knowledge about clean speech such as the covariance matrix $\boldsymbol{Q}$ of the input $\boldsymbol{e}(m)$ of the state equation and the covariance matrix $\boldsymbol{R}$ of the additive noise $\boldsymbol{n}(m)$ have to be calculated. Also, to provide State Transition Matrix A_x an estimation of Linear Predictor Coefficients (LPC) have been applied in "Time Table" (predict phase) of Kalman filter. The state matrix transition of every frame of observed speech signal then would be calculated by Minimum Mean Squared Error (MMSE) method. The aim of minimizing error between observed speech signal and its linear estimation is to employ the result in MMSE approach. Therefore Equation (11) needs to be minimized as

$$E[e_y^2(m)] = E\big[(y(m) - \textstyle\sum_{j=1}^{P} \hat{a}_m y(m-j))^2\big] \tag{11}$$

In Equation (11), $e(m)$ is as an error function and $y(m)$ and P are observed speech signal and linear prediction's order, respectively. The order can be calculated by [39] where f_s is defined as sampling frequency,

$$P = \left(\frac{f_s}{1000}\right) + 2 \tag{12}$$

Also, Equation (13) is the succeeding version of Equation (11),

$$E\big[e_y^2(m)\big] = r_{yy}(0) - 2r_{yy}^T\hat{a} + \hat{a}^T R_{yy}\hat{a} \tag{13}$$

where $R_{yy} = E[yy^T]$ in this equation is defined as autocorrelation matrix of observed input vector y^T, where $y^T = [y(m-1), y(m-2), \dots, y(m-P)]$. The autocorrelation vector of observed speech is r_{yy} and an estimation of predictor coefficient vector is $\hat{a}^T = [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_P]$. If the gradient of Equation (13) is set to zero, the prediction coefficient vector $\hat{a}$ will be obtained

$$R_{yy}\hat{a} = r_{yy} \tag{14}$$
$$\hat{a} = R_{yy}^{-1}\, r_{yy} \tag{15}$$

Equation (15) could also be represented in expanded form as

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_P \end{bmatrix} = \begin{bmatrix} r_{yy}(0) & r_{yy}(1) & r_{yy}(2) & \cdots & r_{yy}(P-1) \\ r_{yy}(1) & r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(P-2) \\ r_{yy}(2) & r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(P-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{yy}(P-1) & r_{yy}(P-2) & r_{yy}(P-3) & \cdots & r_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{yy}(1) \\ r_{yy}(2) \\ r_{yy}(3) \\ \vdots \\ r_{yy}(P) \end{bmatrix} \tag{16}$$

To solve Equation (16) we can use regular Toeplitz structure of autocorrelation matrix R_{yy}. In a Toeplitz matrix all elements on a left-right diagonal are equal. Also, the autocorrelation matrix is cross-diagonal symmetric and there are only $P+1$ unique element $[r_{yy}(0), r_{yy}(1), \dots, r_{yy}(P)]$ in autocorrelation matrix and cross-correlation vector. An algorithm that can effectively solve Equation (16) is Levinson-Durbin algorithm. In our noise reduction process the outputs of Kalman filter that are now de-noised frames will be concatenated to finally construct an estimation of the de-noised speech signal.

Fig. 4 illustrates spectrograms of clean speech signal and applied additive noise (WGN) with different SNR values. It can be seen in Fig. 4 that at 0 dB SNR the quality, fidelity and intelligibility of speech signal will significantly drop. However, in Fig. 5 it can be clearly seen that by employing Kalman filter the majority of the noise has been reduced if not completely suppressed. It is also observed in Fig. 5 that as SNR value increases we will have better results comparatively. Fig. 6 shows the result of applying Wiener filter in noise reducing procedure instead of Kalman filter. Having compared the results of Figs. 5 and 6, this statement can be claimed that the noise has tremendously reduced by employing Kalman filter. Therefore, we can conclude that compared with Wiener filter which is a common method of noise reduction, Kalman filter can work

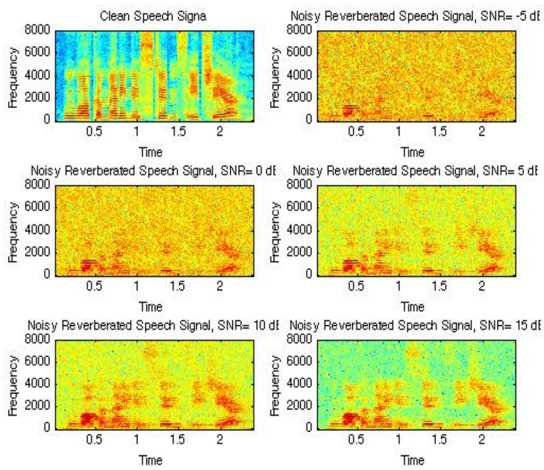

Fig. 4. Spectrograms of clean and noisy reverberated speech signals with different SNR values.

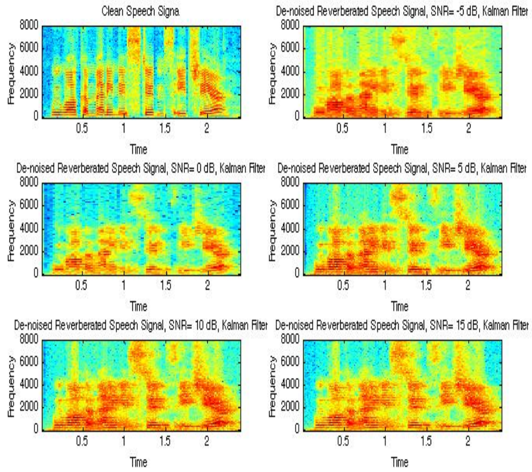

Fig. 5. Spectrograms of clean and de-noised reverberated speech signals with different SNR values by employing Kalman Filter.

effectively and better in case of existence of reverberation in any speech signal.

4 Dereverberation Process

In this paper the dereverberation algorithms, as previously mentioned, have been proposed as the second stage to reduce or suppress reverberation effects after receiving speech signals and enhance their quality, fidelity and intelligibility. There are many multi-channel dereverberation methods, which will be more effective and commercially successful but they are only applicable for more than one microphone cases. This will be an advantage for their respective methods of dereverberation, which would consider various aspects of speech signal. Therefore,

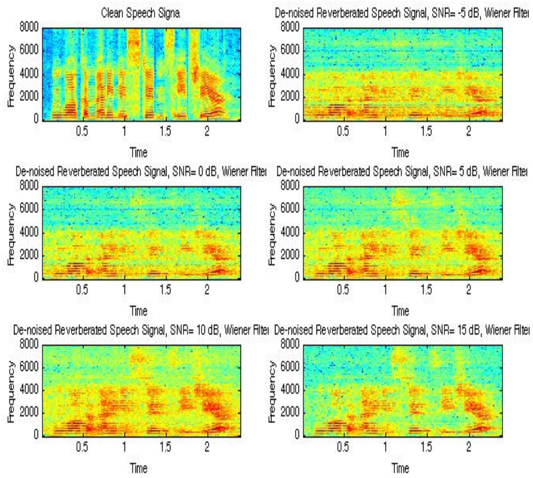

Fig. 6. Spectrograms of clean and de-noised reverberated speech signals with different SNR values by employing Wiener Filter.

in these methods the accuracy of the estimated speech signal will increase, as the error rate would decrease. However, in the case of single-channel dereverberation the observed speech signal will be the sole signal that is received through the acoustic channel. Thus it is extremely important to know that if the received speech signal accompanied with noise and/or echo because these effects might cause damages to the speech signal.

In this paper we have used a dereverberation algorithm as the second stage of our proposed method in order to reduce the effect of reverberation. Inverse filtering procedure has been used in many dereverberation methods in order to reduce the reverberation's effects on received speech signals. The use of inverse filter in this work will bring the closest results to the source of speech signal. The dereverberation procedure in this paper uses a de-noised reverberated speech signal that can be explained as follows [40]. Consider the following equation

$$s(m) = z(m) * \omega(m) \qquad (17)$$

where $s(m)$ is the source speech signal or an estimation of the source and $z(m)$ has been considered as de-noised reverberated speech signal which is an output of the first stage of the proposed method. In Equation (17) $\omega(m)$ is the inverse filter of $h_r(m)$ in Equation (1). Therefore, Equation (18) can be written as

$$h_r(m) * \omega(m) = \delta(m - \tau) \qquad (18)$$

where $\delta(m)$ displays the unit sample sequence with τ delay. By estimating the room impulse response (RIR) of the room [32] and convolving the obtained result through the inverse filtering procedure, an approximate estimation of the source speech signal can be achieved.

5 Experimental and Evaluation Results

In this section the experimental and evaluation results of the proposed method will be presented. Fig. 7 illustrates the spectrums of clean and de-noised reverberated speech signals with different SNR values ranging from -5 dB to 15 dB. It should be noted that Kalman filter has been used for the de-noising part. It is clear from Fig. 7 that these de-noised reverberated speech signals more or less follow the same spectrum as of the clean/source speech signal. In Fig. 8, the spectrograms of clean speech and dereverberated speech signals of various SNRs have been displayed. It is observed from Fig. 8 that as the SNR value increases the quality of dereverberated speech signal increases too.

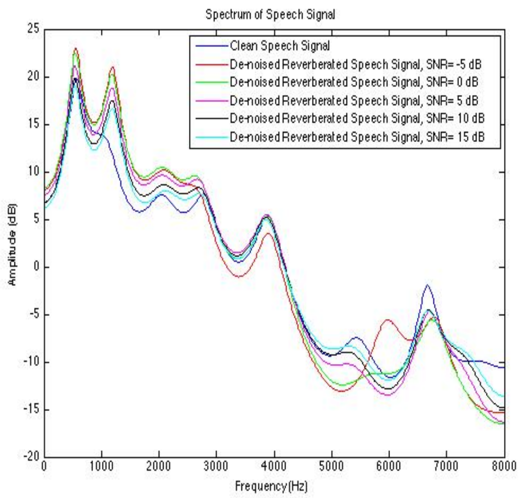

Fig. 7. Spectrums of clean speech signal and de-noised reverberated speech signal with different SNR values by employing Kalman Filter.

Also, the spectrum of clean and dereverberated speech signals can be seen in Fig. 9. Fig. 9 shows that the spectrum of dereverberated speech signals in different SNR value almost follow the spectrum of clean speech signal. Figs. 8 and 9 perfectly described the effectiveness of inverse filtering method for dereverberating a reverberated speech signal. In addition, the estimated fundamental frequency traces and the estimated fundamental frequency contours of clean speech signal and dereverberated speech signals at different SNR values have been depicted in Figs. 10 and 11, respectively. To obtain the estimated fundamental frequency traces, the Cepstrum method had been employed. However, the Autocorrelation method has been used in order to obtain the estimated fundamental frequency contours of clean speech and dereverberated speech signals. Figs. 10 and 11 clearly

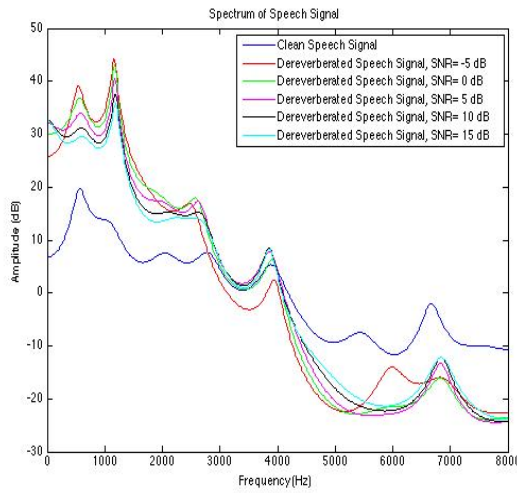

Fig. 9. Spectrums of clean speech signal and dereverberated speech signal of different SNR values.

show the voiced, unvoiced and silence regions of clean speech signal and dereverberated speech signals at different SNR values. Therefore, we can conclude that the vector of RIR's coefficients that had been applied to the inverse filtering procedure improved intelligibility and quality of the speech signals. Furthermore, Table 1 shows the comparison of formants of clean, noisy reverberated and de-noised reverberated speech signals. It is clear from Table 1 that the formants of de-noised reverberated speech signal are quite closed to the formants of clean speech signal in all SNR values.

Also, Table 2 provides formants listing of dereverberated speech signals along with clean and reverberated speech signals where no additive noise was existed. It is shown in Table 2 that the formants of de-noised reverberated speech signal at SNR 5 dB is closed to the formants of clean speech signal.

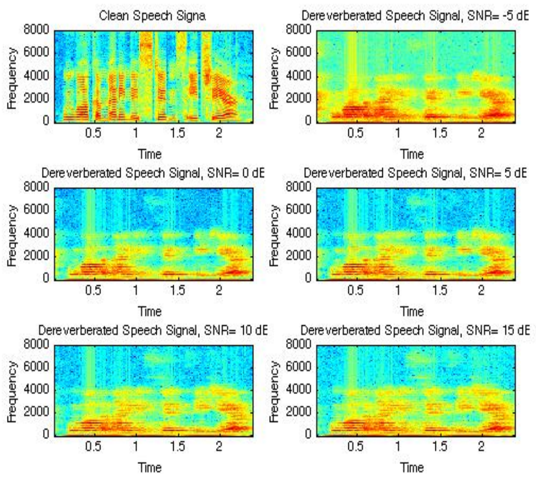

Fig. 8. Spectrograms of clean speech signal and Dereverberated speech signal using different SNR values by applying Kalman Filter.

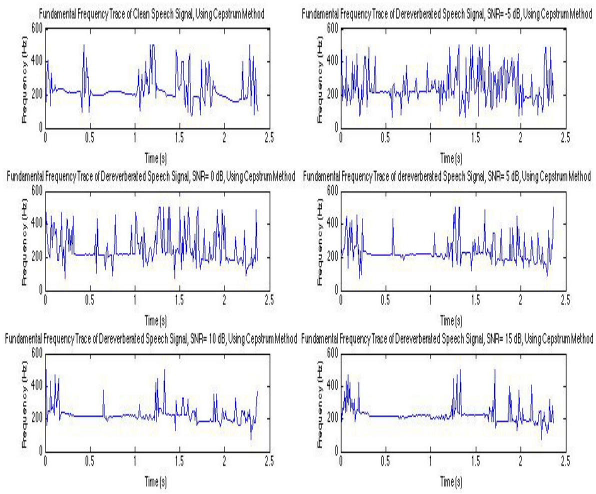

Fig. 10. Fundamental frequency traces of clean speech signal and dereverberated speech signal of different SNR values by employing Cepstrum-based method.

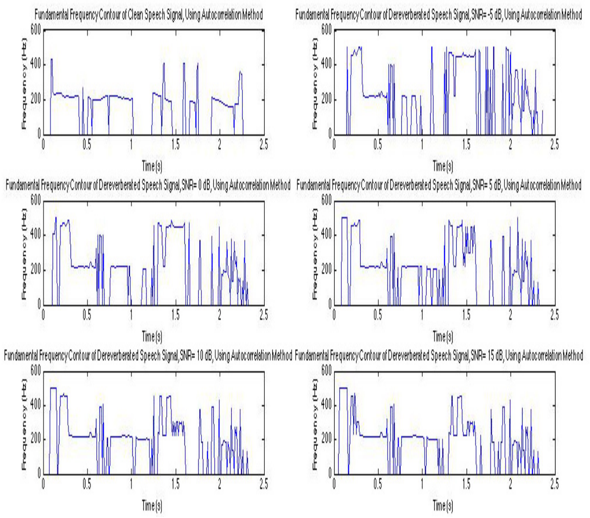

Fig. 11. Fundamental frequency contours of clean speech signal and dereverberated speech signal of different SNR values by employing Autocorrelation method.

Moreover, the objective quality score from the Perceptual Evaluation of Speech Quality (PESQ) defined in ITU-T recommendation P.862 and subjective Listening Quality (LQ) of 10 listeners have been provided in Table 3. It is seen from Table 3 that the PESQ and LQ of dereverberated speech signal at SNR 15 dB provide better results compared to other SNR values. Also, the subjective Listening Quality (LQ) provides better results compared to PESQ in all SNR values. Therefore, we can conclude that the intelligibility of these speech signals is above average. However, in presence of a live receiver the reverberation effect on noisy observed speech signals will be reduced.

6 Conclusion

In this work we presented a two stage method of de-noising and dereverberation where an adaptive filter has been used to reduce the destructive damages that noise had brought to observed speech signal. The enclosed acoustic environment that was stimulated for this purpose also added yet another harmful effect to the received speech signal. The complicated issue in case of having one acoustic channel is that only one microphone is presented, which records an exclusive version of source speech. As the distance between source and receiver that is the microphone and the noise's source increase, the effects of reverberation and noise will increases too. This two stage dereverberation method proposes noise reduction at first stage and then dereverberation procedure as second part of algorithm. The recorded speech signals will be de-noised by an adaptive Kalman filter as it proved to be a better choice of noise reduction and the respective output of this process will be the entry of the next part. The inverse filtering is processed by convolving the inversed of the estimated vector of RIR's coefficients which provides an approximate estimate of source speech signal. The results depict excellent improvement of the dereverberated speech signal compared to the noisy reverberated speech signals that had been recorded by single microphone. Furthermore, PESQ scores show maintaining good quality of dereverberated speech signals whereas the subjective listener could clearly understand the meaning of the dereverberated speech signals.

References:

[1] Naylor P. A., Gaubitch N.D., *Speech Dereverberation*, London: Springer-Verlag, 2010.

Table 1. Formants of clean, noisy reverberated and de-noised reverberated speech signals.

Formant Number	Clean Speech	Noisy Reverberated Speech					De-noised Reverberated Speech				
		SNR (dB)									
		-5	0	5	10	15	-5	0	5	10	15
1	558.4	546.1	546.1	543.1	539.7	538.0	549.7	539.2	533.3	533.0	533.9
2	1087.6	1311.3	1267.2	1236.0	1218.2	1220.3	1197.3	1193.0	1190.9	1195.0	1198.6
3	2048.9	2286.2	2271.0	2231.5	2183.1	2154.4	2085.8	2061.4	2057.2	2082.2	2101.9
4	2807.7	3143.2	3066.6	2940.7	2855.2	2819.2	2698.0	2716.6	2728.1	2752.6	2769.6
5	3904.5	4010.2	3970.7	3917.0	3883.7	3873.1	3917.9	3908.6	3906.9	3882.2	3872.9

Table 2. Formants of clean, noise-free reverberated and dereverberated speech signals.

Formant Number	Clean Speech	Noise-Free Reverberated	Dereverberated Speech				
			SNR				
			-5 dB	0 dB	5 dB	10 dB	15 dB
1	558.4	535.1	535.7	566.4	591.3	611.3	625.1
2	1087.6	1201.5	1163.9	1172.2	1177.9	1186.3	1191.7
3	2048.9	2115.9	1688.3	1923.7	2024.8	2137.5	2216.5
4	2807.7	2779.6	2519.0	2607.8	2657.0	2694.3	2716.2
5	3904.5	3868.8	3360.5	3783.0	3879.1	3862.2	3864.4

[2] Doclo S., *Multi-Microphone Noise Reduction and Dereverberation Techniques for Speech Applications*, Ph.D. dissertation, Catholic University of Leuven, Belgium, 2003.

[3] Takata Y., Nábelek A.K., English Consonant Recognition in Noise and in Reverberation by Japanese and American Listeners, *Journal of the Acoustical Society of America*, vol. 88, pp. 663-666, 1990.

[4] Beranek L., *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, New York: Springer-Verlag, 2004.

[5] Thomas S., Ganapathy S., Hermansky H., Recognition of Reverberant Speech Using Frequency Domain Linear Prediction, *IEEE Signal Processing Letters*, vol. 15, pp. 681-684, 2008.

[6] Passero C.R.M., Zannin P.H.T., Statistical Comparison of Reverberation Times Measured by the Integrated Impulse Response and Interrupted Noise Methods, Computationally Simulated with ODEON Software, and Calculated by Sabine, Eyring and Arau-Puchades' Formulas, *Applied Acoustics*, vol. 71, no. 12, pp. 1204-1210, 2010.

[7] Habets E. A. P., *Single- and Multi-microphone Speech Dereverberation Using Spectral Enhancement*, Ph.D dissertation, Eindhoven University of Technology, Netherlands, 2000.

[8] Subramaniam S., Petropulu A.P., Wendt C., Cepstrum-Based Deconvolution for Speech Dereverberation, *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 5, pp. 392-396, 1996.

[9] Gomez R., Kawahara T., An Improved Wavelet-Based Dereverberation for Robust Automatic Speech Recognition, *11th Annual Conference of the International Speech Communication Association (ISCA)*, INTERSPEECH2010, pp. 578-581, (2010).

[10] Habets E.A.P, Single-channel Speech Dereverberation Based on Spectral Subtraction, *15th Annual Workshop on Circuits*, pp. 250-254, 2004.

[11] Wang L., Odani K., Kai A., Dereverberation and Denoising Based on Generalized Spectral Subtraction by Multi-Channel LMS Algorithm Using a Small-Scale Microphone Array, *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, 2012.

Table 3. PESQ and LQ scores of dereverberated speech signals.

	Dereverberated Speech				
	SNR= -5 dB	SNR= 0 dB	SNR= 5 dB	SNR= 10 dB	SNR= 15 dB
PESQ	2.0	2.1	2.2	2.6	2.9
LQ	2.3	2.7	2.9	3.1	3.4

[12] Gillespie B.W., Malvar H.S., Florencio D.A.F, Speech Dereverberation via Maximum-Kurtosis Subband Adaptive Filtering", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, pp. 3701-3704, 2001.

[13] Unoki M., Toi M., Akagi M., Development of the MTF-Based Speech Dereverberation Method Using Adaptive Time-Frequency Division, *Forum Acusticum*, pp. 51-56, 2005.

[14] Yoshioka T., Nakatani T., Miyoshi M., Enhancement of the Noisy Reverberant Speech by Linear Filtering Followed by Nonlinear Noise Suppression, *International Workshop on Acoustics Echo and Noise Control (IWAENC)*, 2008.

[15] Hikichi T., Delcroix M., Miyoshi M., Speech Dereverberation Algorithm Using Transfer Function Estimates with Overestimated Order *Acoustic Science and Technology*, vol. 27, no. 1, pp. 28-35, 2006.

[16] Tonelli M., Davies M., A blind multichannel dereverberation algorithm based on the natural gradient, *International Workshop on Acoustics Echo and Noise Control (IWAENCE)*, 2010.

[17] Yoshioka T., Tachibana H., Nakatani T., Miyoshi M., Adaptive Dereverberation of Speech Signals with Speaker-Position Change Detection, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3733-3736, 2009.

[18] Krishnamoorthy P., Prasanna S., Reverberant Speech Enhancement by Temporal and Spectral Processing, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 253-266, 2009.

[19] Nakatani T., Miyoshi M., Blind Dereverberation of Single Channel Speech Signal Based on Harmonic Structure, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 92-95, 2003.

[20] Nakatani T., Kinoshita K., Miyoshi M., Harmonicity-Based Blind Dereverberation for Single-Channel Speech Signals, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 80-95, 2007.

[21] Delcroix M., Hikichi T., Miyoshi M., Blind Dereverberation Algorithm for Speech Signals Based on Multi-Channel Linear Prediction, *Acoustical Science and Technology*, vol. 26, no. 5, pp. 432-439, 2005.

[22] Kokkinakis K., Loizou P.C., Selective-Tap Blind Dereverberation for Two-Microphone Enhancement of Reverberant Speech, *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 961-964, 2009.

[23] Abutalebi H.R., Dashtbozorg B., Speech Dereverberation in Noisy Environments Using An Adaptive Minimum Mean Square Error Estimator, *IET Signal Processing*, vol. 5, no. 2, pp. 130-137, 2011.

[24] Kodrasi I., Doclo S., Robust Partial Multichannel Equalization Techniques for Speech Dereverberation, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 537-540, 2012.

[25] Song M.S., Kang H.G., Single-Channel Dereverberation Using a Non-Casual Minimum Variance Distortionless Response Filter, *Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 29-35, 2012.

[26] Jeub M., Schäfer M., Esch T., Vary P., Model-Based Dereverberation Preserving Binaural Cues, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1732-1745, 2010.

[27] Jeub M., Vary P., Enhancement of Reverberant Speech Using the Celp Postfilter, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3993-3996, 2009.

[28] Han K., Wang Y., Wang D., Learning Spectral Mapping for Speech Dereverberation, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[29] Wu B., Li K., Yang M., Lee C.H. A Reverberation-Time-Aware Approach to Speech Dereverberation Based on Deep Neural Networks, *IEEE/ACM Transaction on Audio, Speech, and Language Processing*, vol. 25, issue. 1, pp.102-111, 2017.

[30] Williamson D.S., Wang D., Speech Dereverberation and Denoising Using Complex Ratio Masks, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 5590-5594, 2017.

[31] Schwartz O., Gannot S., Habets E.A.P., An Expectation-Maximization Algorithm for Multimicrophone Speech Dereverberation and Noise Reduction with Coherence Matrix Estimation, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, issue. 9, pp.1495-1510, 2016.

[32] Allen J.B., Berkley D.A., Image Method for Efficiently Simulating Small-room Acoustics, *Journal of The Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, 1979.

[33] Lehmann E.A., Johansson A.M., Nordholm S., Reverberation-Time Prediction Method for

Room Impulse Simulated with the Image-Source Model, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.159-162, 2007.

[34] TMID Dictionary available at: www.ldc.upenn.edu/

[35] Löllmann H.W., Vary P., Low Delay Noise Reduction and Dereverberation for Hearing Aids, *EURASIP Journal on Advances in Signal Processing*, 2009.

[36] Florensa A.T., *Single Channel Speech Dereverberation*, M.S. thesis, Eindhoven University of Technology, Netherlands, 2005.

[37] Eyring C.F., Reverberation Time in Dead Rooms, *Journal of the Acoustical Society of America*, pp. 217-241, 1930.

[38] Vaseghi S., Rayner P.J., Detection and Suppression of Impulsive Noise in Speech Communication Systems, *Communications, Speech and Vision, IEE Proceeding I*, vol. 137, no. 1, pp. 38-46, 1990.

[39] O'Cinneide A., Dorran D., Gainza M., Linear Prediction: The Technique, Its Solution and Application to Speech, *Dublin Institute of Technology, Internal technical report*, 2008.

[40] Radlović B.D., Kennedy R.A., Iterative Cepstrum-Based Approach for Speech Dereverberation, *The 5[the] International Symposium on Signal Processing and its Application (ISSPA)*, pp. 55-58, 1999.