Multi-Microphone Speech Dereverberation and Denoising using Inverse Sparse approximation based GSC

Seema V. Arote [a], Vijay M. Mane [b], Rajesh M. Jalnekar [c]
Vishwakarma Institute of Technology, Pune, India

Abstract:

Speech signals collected by receivers in speech communication applications contain reverberation and additive sounds in addition to the intended speech signals. This audio signal collected by a nearby microphone is just a collection of delayed and fading copies of anechoic speech, and differentiating direct-path speech from reverberated speech may be difficult, particularly when room reverberation and background noise are present. Motivated by the historical rise of sparse estimation in a number of signal processing applications, we explore its applicability to improving the speech signal in a noisy environment. This research focuses on optimizing the front-end of the Sparse-based Generalized Sidelobe Canceller (GSC) in order to achieve an improved speech signal for the corresponding reverb and noisy signal. The enhancement includes multi-channel speech signal processing using a generalized side-lobe canceller with inverse sparse estimation followed by a single-channel Wiener post filter. Speech samples from the IEEE corpus dataset are used to validate the proposed procedure. The obtained results have been compared to those well-known state of the art approaches such as MCWF-WPE, WPE-MVDR, ISC-LP and BS-WPE under the same room conditions. The PESQ achieved with the proposed algorithm is up to 3.4, whereas other methods have a PESQ value of between 2.4 and 3.1. The LSD for the proposed algorithm is decreased to 1.11 and others have 1.4 to 1.9, which is better than the reviewed methods. According to the results, the proposed methodology significantly outperformed the competing approaches in all instances.

Keywords : Speech dereverberation, denoising, two-stage model, room impulse response, sparse estimation, wiener filter, GSC

DOI Number: 10.14704/nq.2022.20.9.NQ44725 **Neuro Quantology 2022; 20(9):6203-6215**

I. INTRODUCTION:

In most real-world speech communication applications, speech signals acquired by receivers contain reverberation and additive sounds in addition to the intended speech signals. Audiences' sound constancy and speech intelligibility suffer as a result. Speech signals move through the air inside a reverberant enclosure, eventually being echoed by the hall's walls, ceiling, floor, or any artefacts. As a result, the speech sound received by a distant receiver would be the sum of many delayed and fading copies of an anechoic speech signal. Historically, the concomitant subject of dereverberation with denoising has received little attention. Speech dereverberation is a tough operation since reverberation is a convolutive disruption when contrasted to background noise, making it difficult to separate the original sent signal from its echoed counterparts, particularly when ambient noise is present. Sparse approximation is concerned with finding sparse solutions to systems of linear equations. This approach is frequently used in signal processing, array processing, machine learning, and image processing [1]. The purpose of this study is to create a dereverberation and denoising technology that enhances speech in the same way as the original source does. This work proposes merging dereverberation and denoising approaches employing GSC with inverse sparse estimation to overcome this difficulty. The remainder of the paper is organised as follows. Section II offers a literature review on speech dereverberation and denoising approaches. Sections III and IV cover the background as well as the recommended methods for dereverberation and denoising.

Section V presents the experimental results and remarks. Section VI is conclusion of paper.

II. LITERATURE SURVEY:

Traditional multimicrophone speech enhancement methods are categorised based on whether they include beamforming [2–6], post-filtering [7–9], a Generalized Sidelobe Canceller [10, 11], or a multichannel based Wiener Filter [12–15]. Concentrating on the relative transfer function (RTF) and the steering vector, which determine the spatial correlation, delay, and sum Beamformers, such as fixed type [16] and minimum variance distortion response beamformers, such as adaptive beamformers [17] and linearly constrained minimum variance [18], can be thought of as spatial filters that efficiently extract information from a specific direction while suppressing noise from other directions. After the beamformer eliminates some disturbances, single or multi adaptive noise reduction is used to further reduce residual noise in the beamforming output. Hashemgeloogerdi et al. [19] suggested a hybrid denoising and dereverberation technique based on a multichannel linear prediction mechanism relayed on a limited Kalman filter. Throughout this method, they apply a steering vector based on a relative transfer function, with the purpose of reducing direct sound while keeping early reflections. L. Yi et al. [20] suggested a strategy for learning signal approximation artificial neural networks employing the lower range of frequencies of the speech mixture to enhance efficiency. The lower band approach for single-channel speech augmentation requires a long short-term memory block that uses the short-time Fourier transform of the relevant frequency spectrum to be implemented. The multi-channel Wiener filters, as well as the weighted estimation error, are critical components of this method. S. Song et al. [21] suggested a multi-channel speech amplification method that is tuned for simultaneous noise reduction and dereverberation. A multi-channel Wiener filter is applied using a minimum variance distortionless response beamformer, which is subsequently complemented by a single-channel Wiener filter. Nakatani et al. [22] proposed techniques for optimising a convolutional beamformer for computationally efficient denoising, dereverberation, and source separation. A hybrid design consisting of a weighted prediction error reduction dereverberation filter followed by a minimum variance distortionless response (MVDR) beamformer has traditionally been employed. T. Dietzen et al. [23] offer a spectral wiener gain post filter that is linked to the posterior state approximation of the Kalman filter. They combine MCLP and GSC in a framework known as integrated sidelobe cancellation and linear prediction to achieve both deconvolution and spatial filtering, with the sidelobe-cancellation filter and the linear prediction filter coupled in parallel on separate microphone signal frames. Wang et al. [24] investigated single and multi-channel speech dereverberation ensemble learning with magnitude domain masking, as well as mapping-based dereverberation with sophisticated modelling, in which deep learning models are trained to distinguish between the real and imaginary elements of a direct-path signal from reverberant ones. To suppress the transmitted signal in multi-channel synthesis, a MVDR estimator is calculated beforehand, and then the real imaginary elements of the terminated signal, which is considered as a filtered version of unwanted signals, are fed as dereverberation features. Xue et al. [25] describe a modulation field multi-channel Kalman filtering (MKF) approach for speech enhancement that combines the sequence propagation system with the temporal variance of speech and the cross correlation coefficient to estimate the clean speech signal. J. M. Martn-Doas et al. [26] established a framework for interactive multichannel speech amplification using a recursive expectation maximisation method. The speech existence probability is determined using a deep learning model mask estimation approach, which is then improved by employing mathematical spatial representations of both noise and noisy speech signals. R. Ikeshita et al. [27] suggested a dereverberation approach based on weighted prediction error (WPE) in which temporal scales are separated into clusters for each frequency bin and the weighted prediction error's causal FIR filter is adjusted for each cluster. J. Wung et al. [28] investigate statistical responsiveness; including computing efficiency, by developing numerically stable interactive weighted least squares algorithms through householder RLS and householder least squares lattice that preserve the RLS algorithm's rapid convergence capabilities. They also proposed and verified the robustness of an angle normalised variant of the householder least squares lattice approach. S.Gannot et al. [29] developed a novel approach for simultaneously reducing early reflections, late reverberation, and ambient noise. To create the filtered variation of an early speech component, the estimate approach is designed as a MVDR beamformer with post filter (PF). A nonorthogonal generalised sidelobe canceller structure is used to achieve MVDR beamformer. Compressed sensing (CS) catches the speech signal with fewer measurements than would otherwise be required. Sparse estimate is a reconstruction approach that takes use of sparsity prior to reconstruction. To put it another way, the signal is scarce in certain areas. Sparse coding is utilised to tackle inverse issues in most applications, including denoising [30], [31], deblurring, [32], super-resolution [33], and CS reconstruction [34]. One example of one-of-a-kind software is CS-reconstruction. Sparse coding grew in favour as computer science progressed. The methodologies proposed aim to provide a redundant sparse decomposition of the speech signal. While the work given here is based on the same sparsity concepts, a novel technique is utilised to determine the most sparse viable option in noisy situations.

III. BACKGROUND:

Anechoic and denoised speech signal is estimated from the noisy data using two-phase framework that consists of the GSC and the post filter. A beamformer, a blocking matrix, and a noise canceller (NC) comprise the GSC. Filtering microphone signals and combining the results to extract the important signal while rejecting interfering signals depending on their spatial location is how beamforming is performed. The present study's primary aim is to develop NC. Inspired by robust estimation hypothesis [35], [36], and sparse coding-based restoration strategies [37], the authors propose a unique NC to remove noise that does not include a particular impulse step and hence can manage channel noise successfully in this work. Following GSC, the Wiener filter is used as a post-processing filter.

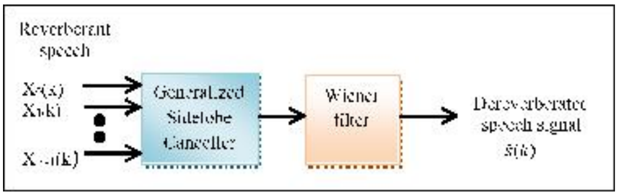

Figure 1. GSC Beamformer with Post Filter

Problem Formulation:

Consider a speech signal that was picked up by i^th microphone. A reverberant signal is created by convolving anechoic signals with simulated room impulse response and adding white Gaussian noise. Eq. (1) provides the reverberant signal received by i^{th} microphone.

$$y_i[n] = h_i[n] * s[n] + v_i[n] = x_i[n] + v_i[n] \quad (1)$$
$$i = 1,2,3,4 \ldots$$

Where

$h_i[n]$ Impulse response (RIR)of acoustic channel from source to microphone i

$s[n]$ is the anechoic signal in the discrete time domain .

$v_i[n]$ is additive noise component in i^{th} microphone signal

$x_i[n]$ is reverberant speech component

The objective of this paper is to estimate desired signal $\hat{s}(k)$for given the corrupted observations, $\{y_i[n]\}_{i=1}^{N}$

IV. METHODOLOGY:

Dereverberation is an acoustic signal processing approach that attempts to retrieve the original speech signal from one or more reverberant signal observations. Beamforming is used in this study to produce dereverberation.

Generalized Sidelobe Canceller (GSC):

The Frost Beamforming method described in [38] is simplified in the generalized sidelobe canceller. This is an example of an LCMV beamforming technology. The GSC is separated into two parts: upper and lower. The top section comprises delay and sum beam former, while the lower part comprise blocking matrix with just a noise attenuator. Frost beamforming provides a speech that is mainly in the desired direction after eliminating practically all unnecessary signals. The stated beamformer is nothing more than a delay and sum beamformer

(DSB), with the delay injected after each microphone to compensate for changes in the time of arrival of a speech at specific microphones, and delay and sum beamformer output is provided by Eq. (2).

$$d(k) = \sum_{m=1}^{M} g_m . x_k(k - \tau_m) \qquad (2)$$

Where g_m are scalars

τ_m are delays

d(k) the beamformer output.

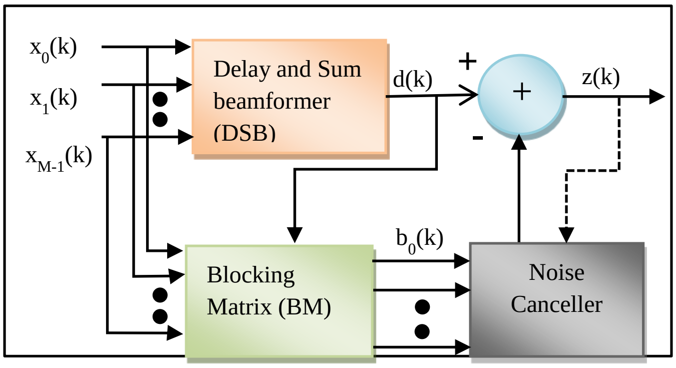

Fig.2. Generalized Sidelobe Canceller

The blocking matrix seems to be in favour of suppressing the desired speech and forwarding all other speech with the Frost beamformer output as a reference speech to inhibit the desired speech and transfer all other speech. The performance of the blocking matrix is given by Eq.(3).

$$b_m(K) = x_m(k) - H_m^T D(k) \qquad (3)$$

Where,

$b_m(\kappa)$is m^{th} the outcome of the blocking matrix

$x_m$is the m^{th} microphone speech

H_m is response of m^{th} adaptive filter

$D(\kappa)$is speech array resulting from Beamformer

The noise canceller takes input from the blocking matrix's output, which would be an interfering speech signal or surrounding noise. Both speech components that have a similarity to the blocking matrix output signals are excluded by the noise canceller. The noise canceller response is deducted first from the beamformer output speech. As a result, the generalized sidelobe canceller's results consist of an improved desired signal and a few partial reverberations along with background noise. The GSC output is given by Eq. (4).

$$z(\kappa) = d(\kappa) - \sum_{m=0}^{M-1} W_m^T(\kappa) Y_m(\kappa) \qquad (4)$$

Where,

$z(\kappa)$ is output of generalized sidelobe canceller.

$d(\kappa)$ is output of delay and sum Beamformer

$W_m(k)$ is m^{th} adaptive filter coefficient vector

$Y_m(\kappa)$ is m^{th} adaptive filter signal vector

Noise Estimation using Inverse Sparse approximation:

To address the l1-norm regularized least square problem of sparse speech signal retrieval, this paper presents an Inverse Sparse approximation with a unique pre-conditioner. This methodology would retrieve a speech signal $x \in \mathbb{R}^\eta$ from its noisy measurements. General sparse approximation can be estimated as [39]

$$\mathfrak{b} = \mathcal{A}x + n \in \mathbb{R}^M \qquad (5)$$

Where $\mathcal{A} \in \mathbb{R}^{MxN}$and $n \in \mathbb{R}^M$ is the environmental noise.

The traditional Least Squares (LS) approach necessitates a large number of observations.

$\mathcal{M} \geq \mathcal{N}$ and $\mathcal{A}$ has full rank $\mathcal{N}$ to recover $\dot{x} = (\mathcal{A}^T\mathcal{A})^{-1}\mathcal{A}^T\mathfrak{b}$. Current Compressed Sensing (CS) techniques could recreate x from a much smaller number of observations$\mathcal{M} \leq \mathcal{N}$. As long as the speech signal appears sparse, the aforementioned basis pursuit's speech denoising dilemma can be solved:

$$\min_{x \epsilon \mathbb{R}^N} \frac{1}{2}\|Ax - b\|^2 + \tau\|x\|_1 \qquad (6)$$

Where $\tau > 0$ is a specified normalization coefficient, $\|x\| = \sqrt{\sum_{i=1}^{N} x_i^2}$ and $\|x\|_1 = \sqrt{\sum_{i=1}^{N} |x_i|}$ denote the l_2 and the l_1 norms of x, respectively,

Including the constraint clearly defines the solution key space in (6) the pseudo recovery, proposed methodology would use a simple optimization strategy to get them out.

$$\hat{x} = [x^+; x^-] \in \mathbb{R}^{2N} \geq 0 \text{ and } \hat{A} = [A, -A] \in \mathbb{R}^{Mx2N} \quad (7)$$

where,

$x_i^+ = \max(x_i, 0)$ and $x_i^- = \max(-x_i, 0)$, the $Ax = \hat{A}\hat{x}$ and, $\|x\|_{1=}\|\hat{x}\|_1$ and hence Eq. 6) can be solved with respect to $\hat{A}$ and. $\hat{x} \geq 0$.

As a result, proposed method just need to acknowledge the version of Eq.(6) shown below for $x \geq 0$

$$\min_{xe\mathbb{R}^N} \frac{1}{2}\|Ax - b\|^2 + \tau e^\tau x \quad (8)$$

$$st x \geq 0$$

Since in Eq.(8) an optimization problem is convex with nothing but linear constraints that fulfills Slater's condition, then this can discover optimized solution by addressing its Karush-Kuhn-Tucker (KKT) system:

$$A^\tau Ax - s - A^\tau b + \tau e = 0 \quad (9a)$$

$$X.Se = 0 \quad (9b)$$

$$(x, s) \geq 0 \quad (9c)$$

Where

$$X = Diag(x) and S = Diag(s)$$

The above equation indicates diagonal matrices consist of primal coefficient x and dual coefficient value x and dual coefficient s, respectively, and 0 and e indicate an entirely zero or all one array whose size should be apparent from reference, respectively. The Inverse Sparse Approximation (ISA) addresses a transformed Karush-Kuhn-Tucker method by merely substituting Eq. (9b) for Eq. (9b) in the basic Karush-Kuhn-Tucker framework.

$$Xse = \sigma\mu e \quad (10)$$

where

$\mu = x^\tau s/N$ goes to 0 , Whenever it converges, it returns to zero and $\sigma \in [0,1]$ is a centeredness element. A σ closer to 1 will prompt search results further towards the interior $(x, s) > 0$. Moving from a specific point (x, s), The novel Karush-Kuhn-Tucker system's orientation could be calculated as

$$A^\tau A\Delta x - \Delta s = r_d \quad (11a)$$

$$S\Delta x + X\Delta s = r_e \quad (11b)$$

Where r_d indicated the stationary residual and complementary slackness residual r_e can be expressed by

$$r_d = s - \nabla h(x) \quad (12a)$$

$$r_e = \sigma\mu e - XSe \quad (12b)$$

Here,

$\nabla h(x) = A^\tau Ax - A^\tau b + \tau e$ is the gradient of the objective function

$$h(x): \frac{1}{2}\|Ax - b^2\| + \tau e^\tau x \quad (13)$$

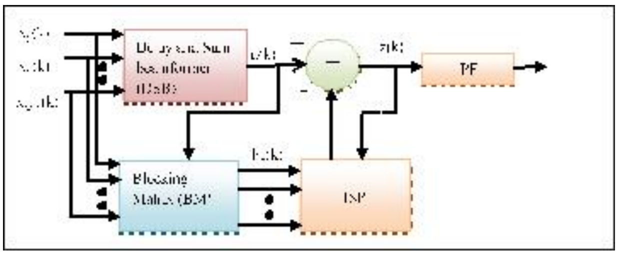

Fig.3. Proposed ISP in Generalized Sidelobe Canceller

In Algorithm 1, the proposed method represents Inverse Sparse Approximation (ISA) with predictor-corrector steps, which employs the Inverse Sparse Approximation (ISA) framework. It can be widely regarded as one of the most effective of the different SAs. To ensure quicker convergence, the proposed ISA uses different initializations, which have simplified but appropriate coefficients, a new pre-conditioner and adaptive tolerance. Although Eq. (9a) must be fulfilled at all times, the proposed ISA allows quite versatile x, s that violate Eq. (9a) during initial setup and subsequent iterations, that only need Eq. (9a) to be fulfilled at convergence.

Algorithm: 1 Inverse Sparse approximation (ISP) Framework

Inputs: ϵ : $Choose\ (x^0, s^0) > 0\ stop\ accuracy$ ϵ (e.g. 1e − 6),

Total Epochs is k_{max}.

$for\ k = 1,2, \ldots\ldots, k_{max}\ \ do$

Perform Prediction Step : set $\sigma \leftarrow 0.001$

$(x^k, s^k, \alpha_p, \alpha_d) = UPDATE(x^{k-1}, s^{k-1}, \sigma)$

if $\min(\alpha_p, \alpha_d) \leq 0.1$ then

Perform Correction Step: set $\sigma \leftarrow 0.99$.

$if\ \mu_k \leq \epsilon h(x^k)\ and\ \|r_d^k\| \leq \epsilon$ then

Break

Output: x^k

Function: UPDATE$(x^{k-1}, s^{k-1}, \sigma)$

Compute $\Delta x, \Delta s$ with σ , x^{k-1}, s^{k-1}

Compute α_p, α_d with $x^{k-1}, s^{k-1}, \Delta x, \Delta s$

Update $(x^k, s^k) \leftarrow (x^{k-1} + \alpha_p \Delta x, s^{k-1} + \alpha_d\ \Delta s)$

return $(x^k, s^k, \alpha_p, \alpha_d\ )$

Post Filtering (PF):

The majority of the Speech de-reverberation algorithms include beamforming after which the Wiener filter is commonly employed as a post-processing filter. The Wiener filter is a linear filter that is known to be optimum. GSC output is z(k), which will provide an artificially loud speech. It will be utilized as one input to the Wiener filter, which is a circuit that applies the signal with the necessary signal s(k) and noise n(k).Wiener also receives a noise signal as an input n(k). The Wiener filter estimates the level of noise in a speech $\hat{n}$(k). To obtain an approximation of the desired signal $\hat{s}$(k), the measured noise is deducted from the noisy input signal z(k).

$$\hat{s}(k) = z(k) - \hat{n}(k) \tag{14}$$

$$\hat{s}(k) = z(k) + n(k) - \hat{n}(k) \tag{15}$$

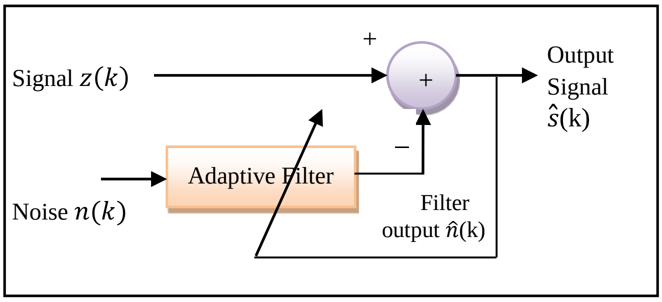

Fig.4. Adaptive Wiener Filter

V. EXPERIMENTS AND RESULT:

To decrease computational and memory expenses, all produced data is sampled at 8 kHz before processing. The IEEE [40] corpora collection was used to acquire the clean speech samples. This dataset comprises 720 examples of clear speech. 300 clean talks are chosen at random from the IEEE corpus. White noise, chatter, and restaurant sounds are blended into clean speech to evaluate the system responsiveness in real time. SNR values for this noise signal are 0 dB, 10 dB, 20 dB and 30 dB. The picture approach was used to construct the Room Impulse Response (RIR) [41]. Convolution of speech and interference with RIRs result in speech mixes collected by four microphones placed in a room with RT60s of 0.3s. The room size for this experiment was [6.1x5.3x2.7] m, with four source to microphone array distances ranging from 1m to 4m. The assessment measures for the proposed algorithm's performance are perceptual Evaluation of Speech Quality (PESQ) and Log Spectral Distance (LSD).

Experimental Results:

TABLE I. Performance comparison in of PESQ and LSD with White Noise

Distance/Noise(dB)		0	10	20	30	0	10	20	30
		PESQ				LSD			
1M	MCWF-WPE[21]	1.95	2.22	2.36	2.39	2.68	2.2	1.93	1.85
	WPE-MVDR[22]	2.17	2.45	2.65	2.76	2.36	1.87	1.62	1.55
	ISC-LP[23]	2.3	2.62	2.86	2.9	2.14	1.72	1.5	1.42
	BS-WPE[27]	2.37	2.76	2.9	2.97	2.21	1.76	1.5	1.43
	Proposed	**2.6**	**2.9**	**3.18**	**3.22**	**1.8**	**1.47**	**1.23**	**1.17**
2M	MCWF-WPE[21]	2.02	2.28	2.41	2.43	2.55	2.13	1.97	1.89
	WPE-MVDR[22]	2.22	2.5	2.64	2.82	2.16	1.75	1.61	1.57
	ISC-LP[23]	2.37	2.67	2.8	2.85	2.08	1.79	1.63	1.57
	BS-WPE[27]	2.43	2.93	3.12	3.14	1.87	1.56	1.39	1.33
	Proposed	**2.65**	**3**	**3.18**	**3.4**	**1.68**	**1.3**	**1.21**	**1.2**
3M	MCWF-WPE[21]	2.04	2.31	2.4	2.43	2.61	2.15	1.95	1.9
	WPE-MVDR[22]	2.3	2.6	2.74	2.82	2.29	1.88	1.7	1.67
	ISC-LP[23]	2.47	2.78	2.87	2.9	2.12	1.74	1.59	1.59
	BS-WPE[27]	2.48	2.77	2.86	3.16	1.96	1.52	1.42	1.38
	Proposed	**2.55**	**3.1**	**3.24**	**3.37**	**1.75**	**1.31**	**1.23**	**1.29**
4M	MCWF-WPE[21]	1.93	2.27	2.3	2.38	2.61	2.15	1.95	1.9
	WPE-MVDR[22]	2.16	2.54	2.6	2.77	2.29	1.88	1.7	1.67
	ISC-LP[23]	2.27	2.66	2.77	2.89	2.12	1.74	1.59	1.59
	BS-WPE[27]	2.37	2.73	2.85	2.94	1.96	1.52	1.42	1.38
	Proposed	**2.54**	**3**	**3.1**	**3.17**	**1.75**	**1.31**	**1.23**	**1.27**

TABLE II.Performance comparison of PESQ and LSD with Babble Noise.

Distance/Noise(dB)		0	10	20	30	0	10	20	30
		PESQ				LSD			
1M	MCWF-WPE[21]	1.95	2.22	2.36	2.39	2.68	2.2	1.93	1.85
	WPE-MVDR[22]	2.2	2.48	2.62	2.72	2.32	1.86	1.65	1.58
	ISC-LP[23]	2.34	2.71	2.87	2.94	2.14	1.68	1.44	1.51
	BS-WPE[27]	2.39	2.82	3.06	3.09	1.99	1.6	1.36	1.4
	Proposed	**2.47**	**2.85**	**3.16**	**3.3**	**1.77**	**1.32**	**1.23**	**1.25**
2M	MCWF-WPE[21]	2.02	2.28	2.41	2.43	2.55	2.13	1.97	1.89
	WPE-MVDR[22]	2.25	2.52	2.67	2.71	2.2	1.83	1.7	1.68
	ISC-LP[23]	2.34	2.69	2.83	2.88	2.08	1.71	1.63	1.57
	BS-WPE[27]	2.43	2.9	3.04	3.06	1.99	1.67	1.52	1.48

	Proposed	**2.55**	**2.94**	**3.18**	**3.25**	**1.59**	**1.21**	**1.12**	**1.11**
3M	MCWF-WPE[21]	2.04	2.31	2.4	2.43	2.61	2.15	1.95	1.9
	WPE-MVDR[22]	2.32	2.66	2.76	2.82	2.21	1.78	1.59	1.58
	ISC-LP[23]	2.36	2.69	2.83	2.88	2.06	1.64	1.5	1.6
	BS-WPE[27]	2.47	2.8	2.94	3.09	1.94	1.58	1.39	1.42
	Proposed	**2.61**	**2.95**	**3.12**	**3.19**	**1.66**	**1.23**	**1.17**	**1.15**
4M	MCWF-WPE[21]	1.93	2.27	2.3	2.38	2.5	2.12	1.94	1.93
	WPE-MVDR[22]	2.22	2.57	2.68	2.77	2.16	1.79	1.62	1.68
	ISC-LP[23]	2.29	2.73	2.77	2.92	2.02	1.65	1.48	1.61
	BS-WPE[27]	2.35	2.76	2.93	3.01	1.78	1.42	1.28	1.33
	Proposed	**2.54**	**2.9**	**3.14**	**3.3**	**1.54**	**1.33**	**1.38**	**1.22**

TABLE III. Performance comparison of PESQ and LSD with Restaurant Noise.

Distance/Noise(dB)		0	10	20	30	0	10	20	30
		PESQ				LSD			
1M	MCWF-WPE[21]	1.95	2.22	2.36	2.39	2.68	2.2	1.93	1.85
	WPE-MVDR[22]	2.16	2.49	2.7	2.75	2.31	1.83	1.64	1.61
	ISC-LP[23]	2.25	2.62	2.81	2.92	2.23	1.74	1.5	1.54
	BS-WPE[27]	2.5	2.78	2.92	3.05	2.12	1.68	1.47	1.43
	Proposed	**2.47**	**2.9**	**3.1**	**3.29**	**2.08**	**1.6**	**1.36**	**1.29**
2M	MCWF-WPE[21]	2.02	2.28	2.41	2.43	2.55	2.13	1.97	1.89
	WPE-MVDR[22]	2.25	2.52	2.7	2.77	2.23	1.83	1.71	1.69
	ISC-LP[23]	2.36	2.67	2.86	2.91	2.08	1.7	1.54	1.48
	BS-WPE[27]	2.5	2.77	3.07	3.13	1.86	1.45	1.4	1.45
	Proposed	**2.56**	**2.9**	**3.05**	**3.22**	**1.73**	**1.35**	**1.33**	**1.29**
3M	MCWF-WPE[21]	2.04	2.31	2.4	2.43	2.61	2.15	1.95	1.9
	WPE-MVDR[22]	2.32	2.62	2.73	2.76	2.22	1.87	1.71	1.68
	ISC-LP[23]	2.5	2.78	2.87	2.96	2.15	1.81	1.62	1.57
	BS-WPE[27]	2.48	2.78	2.88	3.06	1.9	1.51	1.36	1.48
	Proposed	**2.76**	**3**	**3.2**	**3.26**	**1.64**	**1.22**	**1.09**	**1.37**
4M	MCWF-WPE[21]	1.93	2.27	2.3	2.38	2.5	2.12	1.94	1.93
	WPE-MVDR[22]	2.2	2.55	2.64	2.73	2.11	1.75	1.58	1.63
	ISC-LP[23]	2.29	2.72	2.81	2.89	2.1	1.77	1.61	1.62
	BS-WPE[27]	2.41	2.85	2.9	3.07	1.79	1.61	1.44	1.43
	Proposed	**2.50**	**2.9**	**3.13**	**3.35**	**1.58**	**1.29**	**1.21**	**1.3**

Tables I, II, and III compare speech dereverberation performance with varying noise levels for PESQ and LSD parameters. Table I shows the PESQ and LSD values with white noise for proposed algorithm. Table II shows the PESQ and LSD values with babbling noise for proposed approach. Table III shows the PESQ and LSD values with restaurant noise for suggested algorithm. The bold figures in the preceding tables depict the performance of the proposed algorithm in presence of reverberations and noise and also demonstrate that the suggested method outperforms previous techniques. The distance between source to microphone array is adjusted from 1m to 4m, and the signal-to-noise ratio is varied from 0db to 10db, 20dB, and 30dB, with reverberation duration of 0.3s. Even when the distance between source to microphone array is extended from 1m to 4m, PESQ increases and LSD decreases.

Effectiveness of the suggested approach is also shown in the table above. In all quantitative criteria, whether visible or unseen, the proposed speech dereverberation approach outperforms MCWF-WPE [21], WPE-MVDR [22], ISC-LP [23], and BS-WPE [27]. In compared to the previous baseline, the proposed approach results in a significant gain of 25%, 13.43 percent, 9.37 percent, and 4.37 percent for the PESQ test, and a relative decrease of 31.21 percent, 20.24 percent, 16.12 percent, and 7.80 percent for LSD readings. The preceding research supports the effectiveness of our proposed approach. This is because our proposed method is a two-stage algorithm that employs collaborative approximation of the speech signal and an enhanced model. MCWF-WPE [21], which utilizes a multi-channel Wiener filter with weighted estimate error for integrated noise reduction and dereverberation, is compared to the suggested approach. The suggested technique is compared to WPE-MVDR [22], which performs denoising, dereverberation, and source separation simultaneously utilizing a cascade configuration of a Weighted Prediction Error minimization dereverberation filter followed by an MVDR beamformer. The proposed method is also compared to ISC-LP [23], in which the sidelobe-cancellation (SC) and linear prediction (LP) filters work together. Finally, the proposed technique is contrasted with BS-WPE [27], which provides blind signal-based dereverberation based on a linear prediction-based methodology known as weighted prediction error (WPE).

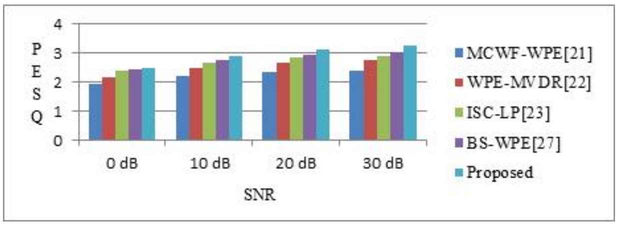

Fig.1 Plot of PESQ with average of all noise for distance of 1m.

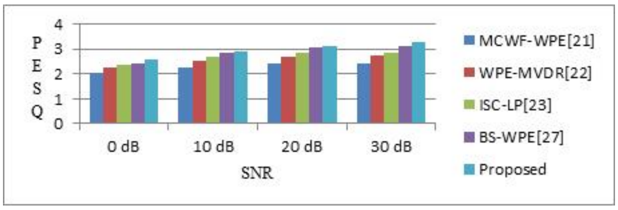

Fig.2 Plot of PESQ with average of all noise for distance of 2m.

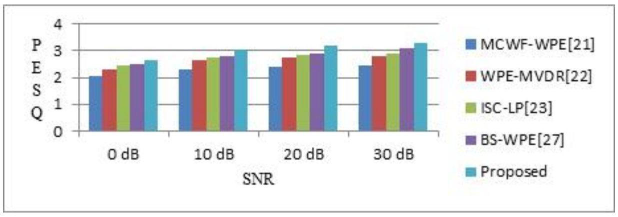

Fig.3 Plot of PESQ with average of all noise for distance of 3m.

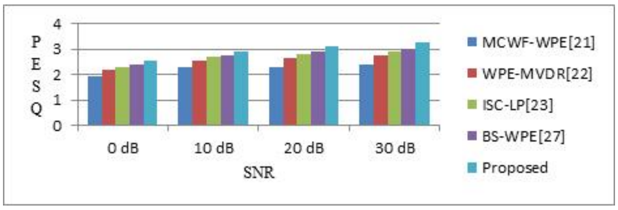

Fig.4 Plot of PESQ with average of all noise for distance of 4m.

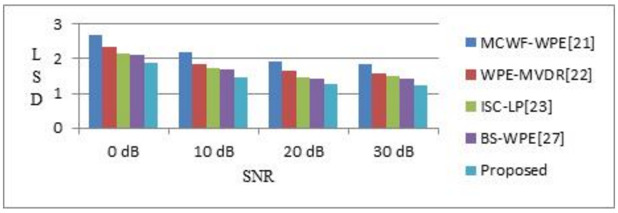

Fig.5 Plot of LSD with average of all noise for distance of 1m.

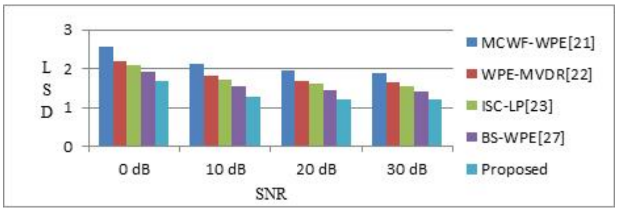

Fig.6 Plot of LSD with average of all noise for distance of 2m.

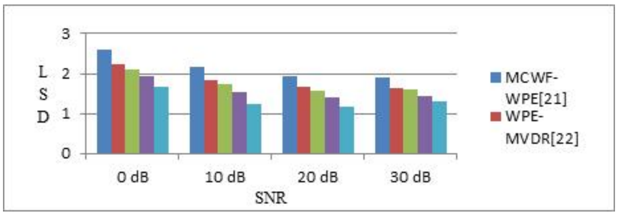

Fig.7 Plot of LSD with average of all noise for distance of 3m.

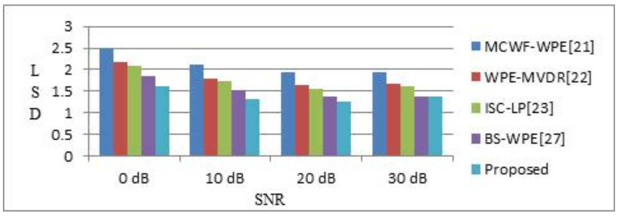

Fig.8 Plot of LSD with average of all noise for distance of 4m.

Figures 1, 2, 3 and 4 shows plot for average of all noise at 0dB, 10dB, 20dB and 30db. It demonstrates efficacy of proposed algorithm in terms of PESQ in presence of reverberations and noise. PESQ of proposed algorithm is improved indicating improvement in speech quality even though distance between source and microphone array is increased from 1m to 4m and signal to noise ratio is varied from 0dB to 30dB. Figures 5, 6, 7 and 8 shows plot for average of all noise at 0dB, 10dB, 20dB and 30db. It also shows effectiveness of proposed algorithm in terms of LSD. LSD values are of proposed algorithm are decreased in comparison with other compared techniques indicating better speech quality. Improvement in speech quality shows decrease in log spectral distance.

VI. CONCLUSIONS:

The measured microphone signals in many practical speech communication contexts include several echoes from walls and other artefacts in the room due to reverberation. Though early reflections do not necessarily affect speech perception, late reflections significantly reduce the consistency and intelligibility of the heard speech signal owing to interference masking effects. An Inverse Sparse approximation-based framework with GSC is proposed in this study to address the l1-norm regularised least square problem for sparse speech signal reconstruction. The suggested approach is compared to well-known state-of-the-art methods such as MCWF-WPE, WPE-MVDR, ISC-LP, and BS-WPE, among others. The suggested approach produced 3.3 PESQ, while previous methods achieved PESQ values ranging from 2.4 to 3.1. The suggested technique has LSD of 1.3, whereas others have 1.4 to 1.9, which is better than the studied methods. The suggested strategy outperforms the competing methods in the majority of circumstances.

DISCUSSION: A two-phase framework, GSC and Post Filter, is used to estimate the anechoic and denoised speech signal from noisy data. Beamforming is accomplished by filtering microphone signals and combining the outputs to generate the desired signal while removing interfering signals based on their spatial position. The present study's primary purpose is to develop NC. Inspired by robust estimation theory [35, 36] and sparse coding-based restoration strategies [37], we propose a unique noise removal methodology in this work that does not need a particular impulse step and so can effectively manage channel noise. Following GSC, the Wiener filter was employed as a post-processing filter. This study provides a GSC based on inverse sparse approximation for the issue of multichannel speech dereverberation in a noisy environment. Pre-conditioner and adaptive tolerance provides quicker convergence than the other approaches in this methodology. Test samples were obtained from the IEEE corpus dataset and tested in the identical room settings to assess the suggested approach. The effects of the suggested algorithm were comparable to those of other well-known techniques. In a variety of scenarios, the suggested method outperforms similar alternatives. The effectiveness of the suggested approach is shown by the outcomes section. In all quantitative criteria, whether visible or unseen, the proposed speech dereverberation approach outperforms MCWF-WPE [21], WPE-MVDR [22], ISC-LP [23], and BS-WPE [27]. In comparison to the aforesaid baseline, the suggested technique results in a relative improvement of 25%, 13.43 percent, 9.37 percent, and 4.37 percent for the PESQ test, and a relative fall of 31.21 percent, 20.24 percent, 16.12 percent, and 7.80 percent for LSD readings. These findings indicate the usefulness of the proposed paradigm.

REFERENCES:

[1] Baraniuk, R.G. Candes, E. Elad, M. and Ma, Y. (2010). "Applications of sparse representation and compressive sensing". Proceedings of the IEEE. 98 (6): 906–909.

[2] M. Souden, J. Benesty, and S. Affes, "On optimal beamforming for noise reduction and interference rejection", Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics ASPAA’2009, New Paltz, New York, USA, October 18-21, 2009.

[3] Habets, Emanual & Benesty, Jacob & Cohen, Israel & Gannot, Sharon & Dmochowski, Jacek. (2010). New insights into the MVDR beamformer in room acoustics. IEEE Transactions on audio Speech and Language Processing. 18. 158 - 170.

[4] W. Herbordt and W. Kellermann, “Adaptive beamforming for audio signal acquisition,” in Adaptive signal processing: Application to real-world problems, J. Benesty and Y. Huang, Eds., pp. 155-194, Springer, Berlin, Jan. 2003.

[5] C. Pan J. Chen and J. Benesty "Performance study of the MVDR beamformer as a function of the source incidence angle" IEEE/ACM Trans. Audio Speech Lang. Process. vol. 22 no. 1 pp. 67-79 Jan. 2014.

[6] M. Souden J. Benesty and S. Affes "On optimal frequency-domain multichannel linear filtering for noise reduction" IEEE Trans. Audio Speech Lang. Process. vol. 18 no. 2 pp. 260-276 Feb. 2010.

[7] Cohen, Israel. (2004). Multichannel post-filtering in nonstationary noise environments. Signal Processing, IEEE Transactions on. 52. 1149 - 1160.

[8] R. Zelinski "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms" Proc. IEEE Int. Conf. Acoust.Speech Signal Process. pp. 2578-2581 Apr. 1988.

[9] I. McCowan and H. Bourlard "Microphone array post-filter for diffuse noise field" Proc. IEEE Int. Conf. Acoust. Speech Signal Process.vol. 1 pp. 905-908 2002.

[10] L. J. Griffiths and C. W. Jim "An alternative approach to linearly constrained adaptive beamforming" IEEE Trans. Antennas Propag. vol. AP-30 no. 1 pp. 27-34 Jan. 1982.

[11] Gannot, Sharon &Burshtein, David & Weinstein, Ehud. (2001). Signal enhancement using beamforming and nonstationarity with applications to speech. Signal Processing, IEEE Transactions on. 49. 1614 - 1626. 10.1109/78.934132.

[12] Doclo, Simon &Moonen, Marc. (2002). GSVD-based optimal filtering for single and multi-microphone speech enhancement. Signal Processing, IEEE Transactions on. 50. 2230 - 2244.

[13] Spriet, Ann &Moonen, Marc &Wouters, Jan. (2004). Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. Signal Processing. 84. 2367-2387.

[14] Doclo, Simon &Spriet, Ann &Moonen, “Efficient frequency-domain implementation of speechdistortion weighted multichannel wiener filtering for noise reduction”, Marc. (2022).

[15] A. Kuklasinski et.al, “Multichannel wiener filters in binaural and bilateral hearing aids-speech intelligibility improvement and robustness to DoA errors,” J. Acoust.Soc. Am., vol. 65, no. 1/2, pp. 8- 16, 2017.

[16] Kuklasinski, Adam & Jensen, Jesper. “Multichannel wiener filters in binaural and bilateral hearing aids - speech intelligibility improvement and robustness to DoA

errors". Journal of the Audio Engineering Society. 65. 8-16, (2017).

[17] Zhong-Qiu Wang and DeLiang Wang. 2020. Deep learning based target cancellation for speech

dereverberation. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 28 (2020), 941–950.

[18] M. Souden J. Benesty and S. Affes "A study of the LCMV and MVDR noise reduction filters" IEEE Trans. Signal Process. vol. 58 no. 9 pp. 4925-4935 Sep. 2010.

[19] S. Hashemgeloogerdi and S. Braun, "Joint beamforming and reverberation cancellation using a constrained kalman filter with multichannel linear prediction," ICASSP 2020 - 2020 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 481-485 [20] Li, Yi & Sun, Yang &Naqvi, Syed. (2021). Single-channel dereverberation and denoising based on lower band trained SA-LSTMs. IET Signal Processing,(2021).

[21] Siyuan, Song & Cheng, Longbiao& Luan, Shuming& Yao, Dingding& Li, Junfeng& Yan, Yonghong.An integrated multi-channel approach for joint noise reduction and dereverberation.Applied Acoustics.(2021).

[22] Tomohiro Nakatani, ChristophBoeddeker, Keisuke Kinoshita, Rintaro Ikeshita, Marc Delcroix, and Reinhold Haeb-Umbach. 2020. Jointly optimal denoising, dereverberation, and source separation. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 28 (2020), 2267–2282.

[23] Thomas Dietzen, Simon Doclo, Marc Moonen, and Toon van Waterschoot. 2020. Integrated sidelobe cancellation and linear prediction kalmanf ilter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 28 (2020), 740–754.

[24] Zhong-Qiu Wang and DeLiang Wang. 2020. Deep learning based target cancellation for speech dereverberation. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 28 (2020), 941–950.

[25] Wei Xue, Alastair H. Moore, Mike Brookes, and Patrick A. Naylor. 2021. Speech enhancement based on modulation-domain parametric multichannel kalman filtering. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 29 (2021), 393–405.

[26] Martín-Doñas, Juan M. ; Jensen, Jesper ; Tan, Zheng-Hua ; Gomez, Angel ; Peinado, Antonio. / Online multichannel speech enhancement based on recursive EM and DNN-based speech presence estimation. I: IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2020 ;3080-3094.

[27] R. Ikeshita, N. Kamo and T. Nakatani, "Blind signal dereverberation based on mixture of weighted prediction error models," in IEEE Signal Processing Letters, vol. 28, pp. 399-403, 2021.

[28] Wung, Jason et al. "Robust multichannel linear prediction for online speech dereverberation using weighted householder least squares lattice adaptive filter." IEEE Transactions on Signal Processing 68 (2020): 3559-3574.

[29] Ofer Schwartz, Sharon Gannot, and Emanuël A. P. Habets. 2015. Multi-microphone speech dereverberation and noise reduction using relative early transfer functions. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 23, 2 (February 2015), 240–251.

[30] Yuan, J., He, J. Blocking sparse method for image denoising. Pattern Anal Applic 24, 1125–1133 (2021).

[31] Huo, F., Zhang, W., Wang, Q. et al. Two-stage image denoising algorithm based on noise localization. Multimed Tools Appl 80, 14101–14122 (2021).

[32] Li, Menghang et al. "Blind motion deblurring via L0 sparse representation." Comput.Graph. 97 (2021): 248-257.

[33]] Li, Xuesong& Zhang, Youqiang&Ge, Zixian& Shi, Hao& Cao, Guo& Fu, Peng. (2021). Adaptive nonnegative sparse representation for hyperspectral image super-resolution. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. PP. 1-1.

[34] Zhang, Mingqiang& Yuan, Dongfeng& Zhang, Minggao. (2021). Learning-based sparse data reconstruction for compressed

data aggregation in IoT networks. IEEE Internet of Things Journal.PP. 2327-4662.

[35] Peter J. Huber. "Robust Regression: Asymptotics, Conjectures and Monte Carlo." Ann. Statist. 1(5) 799 - 821, September, 1973.

[36] P. J. Huber, “Robust Statistics,” John Wiley & Sons, Inc., New York, USA, 1981

[37] Dong, Weisheng et al. “Nonlocally centralized sparse representation for image restoration.” IEEE Transactions on Image Processing 22 (2013): 1620-1630.

[38] B. P. Ng and H. Duan, "Designing amplitude/phase response of the Frost beamformer," in IEEE Transactions on Signal Processing, vol. 55, no. 5, pp. 1944-1949, May 2007.

[39] X. Huang *et al.*, "An interior point method for nonnegative sparse signal reconstruction," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 1193-1197.

[40]"IEEE Recommended Practice for Speech Quality Measurements," in IEEE Transactions on Audio andElectroacoustics, vol. 17, no. 3, pp. 225-246, September 1969

[41] Allen, J.B., Berkley, D.A.: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Amer. 4(65), 943–950 (1979).