International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancement of Speech for Hearing Aid Applications Integrating Adaptive Compressive Sensing with Noise Estimation Based Adaptive Gain

Mr. Hrishikesh B. Vanjari [1], Dr. Sheetal U. Bhandari[2] Dr. Mahesh T. Kolte[3]

Submitted:23/03/2023 **Revised:**25/05/2023 **Accepted:**11/06/2023

Abstract: Hearing aids provide the necessary amplification for successful rehabilitation of hearing-impaired persons. It becomes very challenging for hearing aid devices to attain close to normal hearing. This research suggests a method for improving communication for hearing-impaired people utilizing a combination of three strategies: noise estimation based integrated based gain function, adapted compressive sensing, and listener preference-based customization gain function. Use of integrated gain function and adapted compressive sensing helps to reduce the noise distortion. Use of the customization gain function allows for enhancing the noise-removed speech to the comfort level of the listener. It is achieved by shaping the frequency and amplitude of signal. The overall objective is to enhance quality (noise suppression) and intelligibility (perception) of speech. Performance of proposed solution is tested against noise at various SNR. Results are compared with existing works to established speech quality metrics. The proposed solution is able to attain about 40% improvement in noise quality and 70% reduction in processing time compared to existing works.

Keywords: *Compressive sensing, customized hearing loss, hearing aid, speech enhancement.*

1. Introduction

Worldwide, among sensory deficits, hearing loss is most prevalent.. According to recent WHO estimates [1], about 450 million people worldwide have been disabled by hearing loss. For those with mild to severe hearing loss and various combinations of conductive and sensorineural impairments, hearing aids provide a solution. Frequency dependent elevation of hearing thresholds is done to mitigate conductive loss. Frequency selective amplification is done to reduce conductive losses.

The dynamic range of hearing in people with sensorineural deficits has been diminished, along with the loudness connection between speech components and frequency-dependent elevation of hearing thresholds [29]. Speech comprehension is very challenging for those with sensorineural loss, particularly in loud settings [30].

Many works on speech enhancement for sensorineural loss have been carried out in the last two decades. The existing works are in the categories of selective frequency amplification, range compression and suppression of tones [28]. Under the presence of noise of different types, the performance degrades in these approaches. These approaches fail to prevent temporal and spectral masking. They also fail to suppress wideband nonstationary noise. The computational complexity of hearing aids limits the use of sophisticated signal processing methods.

In this paper, we provide a speech enhancement solution combining noise estimation based integrated gain function, adaptive machine learning based compressive sensing, and customizable gain function for shaping the frequency and amplitude according to the comfort level of the hearing-impaired listeners. The suggested remedy solves the two challenges of enhancing voice quality by reducing noise and enhancing speech understandability. More and more voice processing applications are using compressed sensing (CS). It is based on exploiting the signal sparsity and offers better utilization of the resources. Speech enhancement via compressive sensing enables the reconstruction of sparse data, thereby allowing a simpler implementation and no need of voice activity detection. Different from previous compressive sensing works, this work uses machine learning based optimal sensing matrix selection based on features of noise speech signal. This allows for faster convergence of $\ell 1 - norm$ in compressive sensing. Proposed solution is tested for different noisy speech samples in NOIZEUS corpus dataset. Results are compared with existing solutions for various speech quality metrics and processing delay. Significance of this work's contributions are listed below.

1 Research Scholar, 2 Head and Professor, Electronics &Telecommunication Engineering Department, 3 Professor, Electronics & Telecommunication Engineering Department,
Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune,
Maharashtra, India
[1]hrishikesh@outlook.in [2]sheetalubhandari@gmail.com
[3] mtkolte@yahoo.com

1. Ensemble of three strategies is done in a pipeline for reducing the noise and improving the speech intelligibility
2. Compressive sensing is fine-tuned with the adaptation of sensing matrix generation using machine learning. The adaptation is able to decrease computational complexity, storage, and delay compared to use of Gaussian random sensing matrix.
3. The frequency and power level are customizable according to the comfort of the hearing impaired listeners. The customization is achieved using a frequency and amplitude shaping gain function.

The organization of paper is: in section 2 provides a detailed survey on existing works and summarizes the open issues. Section 3 provides further information on the suggested speech enhancement method. Part 4 presents experimental findings and a comparison analysis of the suggested work. Part 5 concludes the study and outlines its future scope.

2. Related Work:

Authors of [2] suggested a compressive sensing-based voice enhancement technique that is devoid of previous noise estimate. Compressive sensing allows for the recovery of speech signals, that are normally spare and nonspare signals are filtered. Solution was specific to Arabic speech with Gaussian noise. The solution takes considerably higher delay and performs poorly for real world noise. Convolutional neural networks with multiobjective learning were employed for voice augmentation by the authors in [3]. The noisy voice spectrum is cleaned up using a convolutional neural network. An algorithm is fast and targeted to be executed on smartphone. A review of some of the existing two-channel speech enhancement algorithms are presented in [4]. Most of the approaches need manual tuning by hearing aid users to adjust the noise suppression level. Deep learning algorithms for hearing loss and speech enhancement were experimented in [5]. A Mel-frequency feature along with LSTM is found to provide better PESQ scores. Authors in [6] used an integrated gain function for speech enhancement. One gain function is obtained using the coherence between speech and noise and another gain function is obtained using Super Gaussian Joint Maximum a Posteriori (SGJMAP). Background noise is suppressed using the coherence gain function and speech quality is improved using SGJMAP. Users with cochlear implants may read an evaluation of neural network-based speech improvement in [7]. Decomposition of a noisy voice signal into time-frequency units. Features that are extracted from time-frequency units and passed to the neural network. Frequency channel estimates are provided by the neural network. This estimation is used for noise suppression. Speech enhancement using multichannel Kalman filter is proposed in [8]. An optimal filter gain is achieved by joint exploitation of multichannel spatial information and temporal speech correlations. The approach is not suitable for hearing aid applications because of noise varsity in real environment. Authors in [9] proposed quantile noise estimation-based speech enhancement. Two noise estimation techniques – fixed and adaptive quantiles are proposed. Computational complexity is lower in this approach and it provides consistent performance for different noise types at different SNRs. Authors in [10] presented a 2 stage method to deal with speech corrupted by room reverberation and background noise. Authors used deep neural networks (DNN) to train a gain function to be applied to noisy speech signal to remove noise. However, the effectiveness of the method for hearing aid applications is not experimented. A supervised method for speech enhancement using deep learning is proposed in [11]. Deep learning model is trained for mapping between noisy signals to clear speech. Non-linear regression function based deep- Neural Network is used for modeling the relationship among noisy speech and clean speech. Over smoothing problem in nonlinear regression are resolved using global variance equalization. Noise aware training strategies are applied to improve the performance of DNN for unseen noise conditions. A method for enhancement of speech in a noisy and reverberant environment using DNN is proposed in [12]. A spectral mapping among corrupted speech in a reverberant environment and the clear speech is learnt using a DNN. Log magnitude spectrum of clear speech is output of DNN. The DNN has been conditioned for various loud speech types. Goal of training is to minimize mean square error (MSE). Authors in [13] proposed a mechanism to solve the generalization problem in supervised speech segregation although the use of large-scale training. A DNN is trained with 10,000 noises to predict ideal ratio mask and used to separate new noises from the sentences with several signal-to--to-noise ratio. The approach is found to be promising in new acoustic environments. For enhancing speech, [14] uses a log spectral amplitude estimation. Amount of speech distortion is controlled using the knowledge of frequency information. Scaling parameter is presented into gain function allowing the user to customize speech according to his listening preference. Authors in [37] proposed a hybrid speech estimator using two-stage filter. Filtering is based on Discrete Krawtchouk-Tchebichef transform. Linear estimation is done for noise prediction, which cannot be applied for hearing aid applications. Authors in [38] proposed an optimum low distortion estimator to estimate the noise. However, it requires the signal to be transformed using an orthogonal polynomial function which is computation intensive for hearing aid applications. Authors in [15] proposed a new speech enhancement gain function called as super-Gaussian joint maximum a posteriori. The gain function also has a trade-off parameter allowing customization according to listener preference. This

customization controlled the noise suppression and speech distortion in real time. Similar to [13], authors in [16] proposed a method to solve the generalization problem. Ideal ratio mask is predicted utilizing deep neural network. The method was able to provide substantial sentence intelligibility benefits for hearing impaired listeners. A solution to the mismatch issue in deep neural networks used for voice enhancement was put out by the authors in [17]. The methodology uses an ideal binary masked dynamic noise estimation strategy to include noise information in test utterances based on noise conscious training. To improve its generalization ability for unobserved and nonstationary noise settings, DNN is trained with 100 different forms of noise. To increase the quality of improved speech, global variance equalization is also employed. Multi objective framework for speech enhancement is proposed in [18]. Secondary features like categorical information like ideal binary mask and mel-frequency cepstral coefficients (MFCCs) are used in noise estimation. A novel ensemble deep neural network with two deep networks for speech enhancement in [19]. However, the computation complexity is very high in this approach.

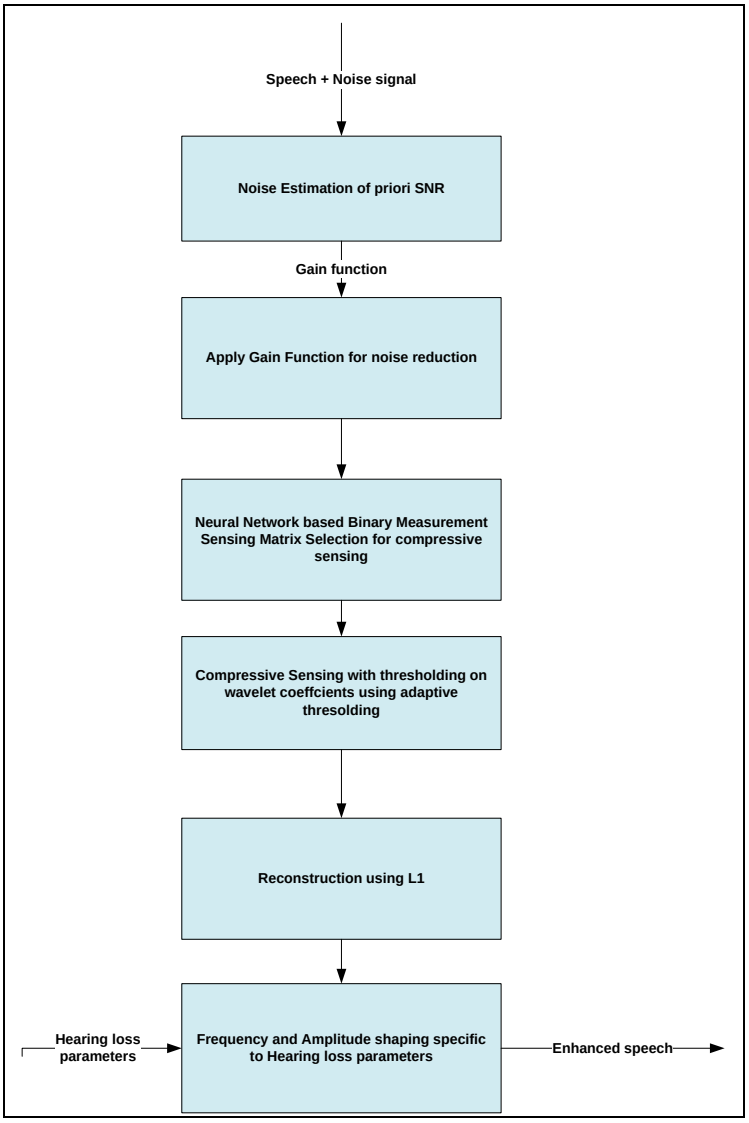

Fig 1 Proposed Speech Enhancement Solution

Coherance based speech enhancement function is proposed in [20]. Combining a gain function based on spectral subtraction with adaptive gain averaging with a gain function based on coherence. Background noise is suppressed effectively using a coherence based gain function. Gain function weighting factor is also introduced to customize the speech enhancement to the environment's noise level and hearing comfort. Technique is able to have higher performance compared to stand-alone spectral subtraction. Independent component analysis is used for speech enhancement with a single microphone in [21]. Speech spectrum estimation is needed prior to this approach and it can work for only one noise type.

3. Proposed Speech Enhancement Solution

Figure 1 shows the suggested voice enhancement solution's process flow. The noisy speech goes through three stages before it becomes noise suppressed and intelligible to the hearing-impaired listener. The first stage of filtering is

applying a noise estimation based gain function to suppress the noise. The second stage is adaptive compressive sensing and the third stage is a gain function tuned for the comfort level of the listener. Each of these stages are explained below

A. Gain function for Noise reduction

The noisy speech $x(t)$ for a additive noise is modeled as

$$\boldsymbol{x(t) = s(t) + n(t)} \qquad \textbf{Eq 1}$$

Where speech is given as $s(t)$ and $n(t)$ is noise signal. Signals can be represented in terms of spectral components as

$$\boldsymbol{x(t) = \sum_{p=1}^{M}\sum_{t=1}^{K} X(s,t)} \qquad \textbf{Eq 2}$$

Where M and K are number of frames and number of spectral components and X is spectral component. Similarly, noise and speech can be represented in terms of spectral components as

$$\boldsymbol{s(t) = \sum_{p=1}^{M}\sum_{t=1}^{K} S(p,t)} \qquad \textbf{Eq 3}$$

$$\boldsymbol{n(t) = \sum_{p=1}^{M}\sum_{t=1}^{K} N(p,t)} \qquad \textbf{Eq 4}$$

Finding an estimate for S that reduces predicted value of a specific distortion measure for a collection of spectral characteristics is the aim of noise reduction. An estimation is made in terms of SNR from the noise features. By applying gain to each of spectrum components of X, one may estimate S from the estimated SNR. Gain is a compromise between voice distortion and noise suppression. As a result, the predicted SNR for a certain noise power spectrum density (PSD) determines how effective speech augmentation is. The speech activity detection technique put out in [22] serves as the foundation for the proposed noise estimates. A posteriori and a priori SNR is calculated from PSD($\tilde{\gamma}_n$) given by [22] as

$$\boldsymbol{\widetilde{SNR_{post}(s,t)} = \frac{|X(s,t)|^2}{\tilde{\gamma}_n(s,t)}} \qquad \textbf{Eq 5}$$

$$\boldsymbol{\widetilde{SNR_{pri}(s,t)} = \beta\frac{|\bar{S}(s-1,t)|^2}{\tilde{\gamma}_n(s,t)} + (1-\beta)P[\overline{SNR}_{post}(s,t) - 1]} \qquad \textbf{Eq 6}$$

P is the half wave rectification function.$\bar{S}(s-1,t)$ is estimated speech spectrum for the previous frame. β is behavior control parameter with value from 0 to 1. Based on the SNR a priori estimate, the gain function is designed as

$$\boldsymbol{G(s,t) = \frac{\overline{SNR}_{pri}(s,t)}{1+\overline{SNR}_{pri}(s,t)}} \qquad \textbf{Eq 7}$$

The speech spectrum is obtained by applying the gain function to the noisy speech spectrum for each of the frames as

$$\boldsymbol{\bar{S}(s,t) = \sum_{p=1}^{M}\sum_{t=1}^{K} X(s,t)\, G(s,t)} \textbf{Eq 8}$$

B. Adaptive Compressive Sensing

Signals with sparse representation in certain bases may be compressed and recovered using the method of "compressive sensing." It is used to improve speech since it has the ability to handle sparse signals. The stages for CS-based speech augmentation are as follows:

1. Conversion to sparse representation
2. Sensing matrix construction
3. Acquisition of signal
4. Reconstruction

Wavelet basis is presented with the nonsparse noisy speech, and a hard threshold is used to replace any significant coefficients with a value of 0.

The creation of a random Gaussian sensing matrix.

Sparse vector coefficients are multiplied by the random Gaussian-sensing matrix. Inverse wavelet transform and l1 minimization are used to reconstruct speech signal. There are two issues with using compressive sensing to improve speech.

- Selection of a suitable thresholds for sparse representation to achieve a fine balance between sparseness and minimum distortion of speech is lacking
- The sensing matrix must be constructed in such a way to reduce loss of input signal
- Reduce number of iterations of $\ell 1$ minimization and speedup the recovery process.

1. Threshold Selection

An adaptive thresholding scheme is proposed in this work satisfying two goals of minimum degradation of the signal and reduction of background noise with uniformly distributed power spectral density. The proposed thresholding is done on the DFT spectrum of noised speech signal. Input noised speech signal is split into short segments as

$$x_i(n) = \begin{cases} w(n)x(n+iN(1-v), & \forall\, \mathrm{n} = 0,1,..\mathrm{N}-1 \\ 0, \forall\, \mathrm{n} \neq 0,1,..\mathrm{N}-1 & and\ i = 0,1,2 \ldots I-1 \end{cases}$$

Where N is length of segment $x_i(n)$, the segment index is denoted as i, I is total number of segments, $w(n)$ is weighting window, v is the overlap ratio. The length of segment N is chosen in a power of 2 in the case of easy application of DFT. The coefficient resulting from application of DFT on the $x_i(n)$ is given as

$$\boldsymbol{X_i(k) = DFT\left(x_i(n)\right) = \sum_{n=0}^{N-1} x_i(n) e^{-j\frac{2\pi}{N}kn}} \quad \textbf{Eq 9}$$

The threshold function T transforming $X_i(k)$ can be given as

$$\boldsymbol{T\left(X_i(k)\right) = \begin{cases} 0 & ,\forall\ |\mathbf{Xn_i}| < T_1 \\ X_i(k)\ \omega(|\mathbf{Xn_i(k)}|), & |\mathbf{Xn_i}| \geq T_1 \end{cases}} \quad \textbf{Eq 10}$$

$Xn_i$is calculated based on estimation of the mean and standard deviation of coefficients as

$$\mathbf{Xn_i} = \frac{X_i(k) - \mu_m(k)}{\sigma_m(k)} \quad \textbf{Eq 11}$$

The mean and standard deviation estimation is as follows

$$\boldsymbol{\mu_m(k) = \frac{\sum_{i=0}^{I-1} X_i(k)}{I}} \quad \textbf{Eq 12}$$

$$\boldsymbol{\sigma_m(k) = \sqrt{\frac{\sum_{i=0}^{I-1} X_i(k)^2}{I} - \mu_m(k)^2}} \quad \textbf{Eq 13}$$

I is the number of segments.

2. Sensing Matrix Construction

Sensing matrix used by majority of compressive sensing techniques is Gaussian random matrix. The presence of nonzero noninteger values makes the Gaussian random matrix denser. More computational and storage complexity are the results. Gaussian random matrix also raises hardware implementation costs. In this paper, a sparse binary matrix is suggested in replacement of a Gaussian random matrix.

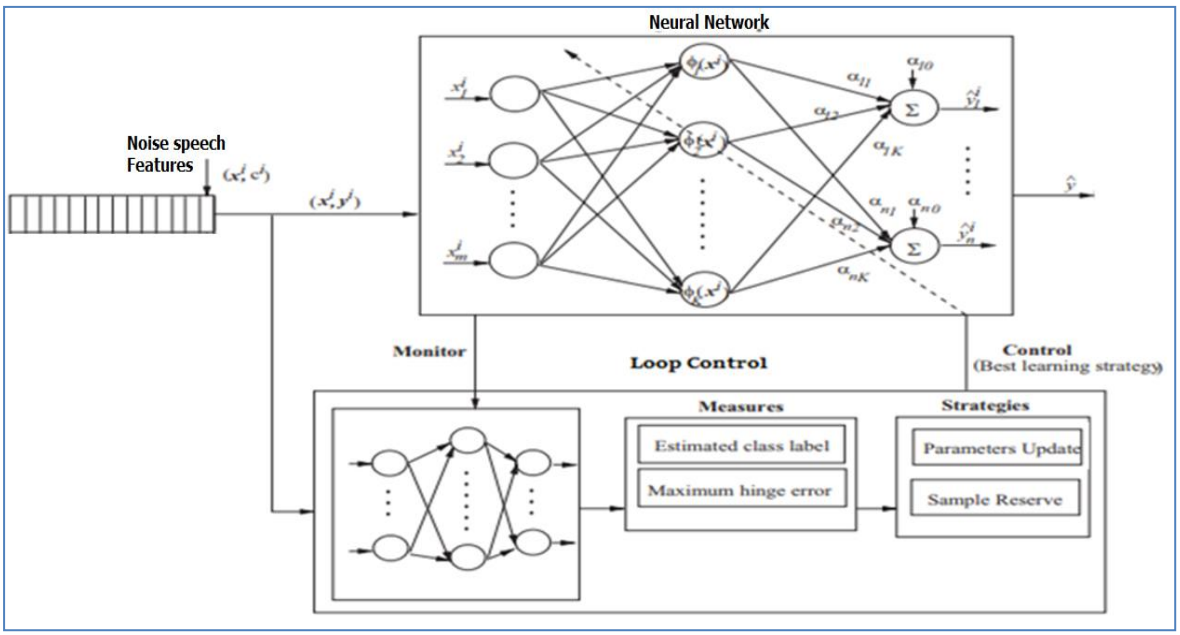

Fig 2 Neural Network with control

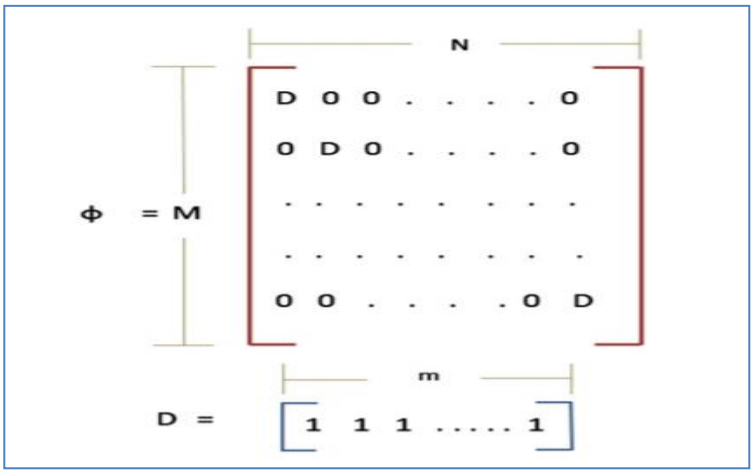

Fig 3 Sensing Matrix

Algorithm 1: Sensing Matrix generator

Input: Number of measurements M and the length of the signal N

Output: Array of Sensing Matrix (M*N)

$m \cong Q/P$

m_1 = integer less than m

m_2 = integer less than m

$Nm_2 = Q - P \times m_1$

$Nm_1 = P - Nm_2$

$Rpm_2 = r_1 \; and \; r_M$

$Rpm_1 = [r_1, r_2, r_3, \ldots r_M] - Rpm_2$

$rowt_1 = \{1_1, 1_2, 1_3, \ldots 1_{m1}, 0_1, 1_2, 1_3, \ldots 1_{N-m1}\}$ // $m1$ ones and $N - m1$ zeros

$rowt_2 = \{1_1, 1_2, 1_3, \ldots 1_{m2}, 0_1, 1_2, 1_3, \ldots 1_{N-m2}\}$ // $m2$ ones and $N - m2$ zeros

For k=1 to M do

If $r_k \in Rpm_1$ then

$row_k = rowt_1$

$rowt_1 = circular \; shift \; rowt_1 \; right \; by \; m_1 \; times$

$rowt_2 = circular \; shift \; rowt_2 \; right \; by \; m_1 \; times$

Else

$row_k = rowt_2$

$rowt_1 = circular \; shift \; rowt_1 \; right \; by \; m_2 \; times$

$rowt_2 = circular \; shift \; rowt_2 \; right \; by \; m_2 \; times$

End if

End for

Diagonalblock, $D_b = \{row_1^T, row_2^T, \ldots . row_M^T\}^T$

$M_s = matrix\ with\ U\ number\ of\ D_b s$

End if

allM = []

L= $m1$

For i=1: $m1$

$M_{temp} = M_s$

For j=1:rows in M_{temp}

$M_{temp}[j][L] = 0$

End

L=L-1;

allM[i] = M_{temp}

End for

Return allM

The suggested sparse binary matrix contains fewer nonzero values than the Gaussian random matrix, which lowers the complexity, time, and storage needs of computing. The proposed solution for sensing matrix generation uses machine learning for fine-tuning the sensing matrix selection. It is done based on features extracted from noisy speech signals.

With sensing matrix serving as outcome and the characteristics of noisy speech signals serving as the input, a multilayer feed-forward neural network [36] is trained. Up until a suitable accuracy level is reached, the feed-forward neural network is loop regulated and retrained. The feed-forward neural network shown in Figure 2 is trained using following characteristics that are retrieved from noisy voice samples.

1. Zero crossing rate (ZCR) [34]
2. Average power (AP)
3. MFCC [35]

A minor modification to the sensing matrix synthesis technique suggested in [23] produces sparse binary sensing matrix for a noisy voice sample. M*N sensing matrix is created for M measurements of a signal of length N. The produced sensing matrix is shown in Figure 3. Dimension of diagonal block is given as$\frac{Q}{P} \cong m$. Less and more than m, m1 and m2, an integer value is assigned. Entire number of rows in D is split between rows with m1 ones and rows with m2 ones. Expression for how many rows include m2 ones is

$$Nm_2 = Q - P \times m_1$$

Where $D = Q \times P$. Number of rows with m_1 is given as

$$Nm_1 = P - Nm_2$$

By stacking and rotating rows with m 1 and m 2 values, the diagonal block D is created. The sensing matrix is made by inserting the diagonal block into zero matrix of size M*N. This sensing matrix is utilized as a foundation matrix, and a new sensing matrix is created for each iteration equivalent to continuous ones by setting each one in all rows to zero. An algorithm for sensing matrix generation is given in Algorithm 1.

A training dataset for training neural network for sensor selection is given in Algorithm 2. The characteristics of zero crossing rate, AP, and MFCC are recovered for each of the noised speech samples.

ZCR provides the noised speech signal's rate of sign change over time. It may be used to locate the frequencies around the area of energy concentration. It is determined as

$$ZCR = \frac{1}{L-1}\sum_{n=1}^{L-1} I_{R<0}(x(n)x(n+1)) \quad \textbf{Eq 14}$$

$I_{R<0}$ is indicator function. An AP of noised speech signal is computed as

$$AP = \frac{1}{L}\sum_{j=1}^{L} |x_j(n)|^2 \quad \textbf{Eq 15}$$

Using Algorithm 1, each of the noised speech samples receives a sensing matrix, which is subsequently applied to the noised speech sample. L1 minimization is used for reconstruction. Several sensing matrices are used to compare the PESQ quality metric between reconstructed signal and noised speech samples, and sensing matrix with the highest PESQ is chosen as associated class label for input noised speech sample. Outcome sensing matrix for each of the noised speech samples is created using the aforementioned technique, and it is then utilized to train a neural network. The ZCR, AP, and MFCC characteristics are recovered from a fresh noised speech signal and sent to the trained neural network to create sensing matrix. The spare matrix produced after adaptive thresholding is then multiplied using this sensing matrix.

C. Customization gain function

The frequency and amplitude of the signal is shaped by the customized gain function considering the comfort level of the listener.

The frequency shaping is achieved by applying an adaptive gain function based on the frequencies listeners has hearing discomfort. The difficult to hear frequencies are raised using the gain function.

The adaptive gain function is given below in Figure 4.

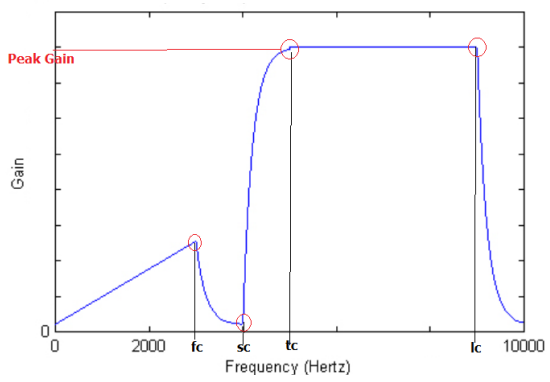

Fig 4 Frequency Gain function

Users can select two frequencies (fc and lc). The sc and tc in the above response is set as

$$sc = (fc - 2000)$$

$$tc = sc + (sc - fc)$$

Amplitude shaper ensures that the output power of the signal does not exceed a saturation level. Another advantage of amplitude shaping is that noises concentrated in low power levels are also filtered. The gain function for amplitude shaping is given as

$$AG = \begin{cases} P_{out}, \forall\, P_{out} < P_{sat} \\ P_{sat}, \forall\, P_{out} \geq P_{sat} \end{cases}$$

4. Results

Performance of proposed solution is compared with compressed sensing-based speech enhancement presented in [2], noise adaptable speech enhancement proposed in [26], and noise estimation-based speech enhancement proposed in [9]. Noisy speech corpus [25] was utilized for evaluating performance. The performance was measured in terms of

1. Perceptual Evaluation of Speech Quality (PESQ) [31]
2. Mean Opinion Score on a five-point scale for 20 listeners (Table 1)

Quality	Excellent	Good	Fair	Poor	Bad
Rating	5	4	3	2	1

Table 1 MOS Ratings

3. Processing time
4. Segmental SNR (seqSNR) computed as

$$seqSNR = \frac{1}{M}\sum_{m=1}^{M} 10\log\left[\frac{\sum_{n=R(m-1)+1}^{Rm} x^2(n,m)}{\sum_{n=R(m-1)+1}^{Rm}\left(x(n.m) - y(n.m)\right)^2}\right]$$

Where M is the frame count, x is the noise-added signal, and y is the suggested solution's output. X(n,m) is the m^{th} frame's n^{th} sample.

5. Perceptual Evaluation of Audio Quality (PEAQ) [32]
6. Perceptual Speech Quality Measure (PSQM) MOS Score [33]
7. Output SNR calculated as

$$SNR(dB) = 20\log\frac{||X||_2}{||X - \bar{X}||_2}$$

D. Impact on output SNR

The performance was tested for different noise types in three different SNR of 0, 5 and 10 and the average output SNR is given in Table 2. The proposed solution has better output SNR compared to [2], [26], and [9] even at low input SNR of 0 dB.

Noise Type	Solution	0 dB	5 dB	10 dB
Airport	**Proposed**	3.34	4.96	8.14
	[2]	6.25	8.14	11.4
	[9]	7.24	8.67	11.86
	[26]	4.78	6.94	9.57
Babble	**Proposed**	3.32	4.86	7.12
	[2]	6.12	7.96	10.56
	[9]	7.43	8.21	11.24
	[26]	5.92	8.67	9.12
Car	**Proposed**	4.12	5.24	8.50
	[2]	7.13	8.12	11.78
	[9]	7.96	8.45	11.96
	[26]	6.50	7.12	10.1
Restaurant	**Proposed**	4.25	6.63	9.60
	[2]	7.64	8.13	12.86
	[9]	7.89	8.86	13.12
	[26]	5.96	7.74	10.56

Station	**Proposed**	3.46	5.67	8.89
	[2]	7.83	8.14	11.67
	[9]	7.96	8.67	11.93
	[26]	4.86	6.24	10.21
Street	**Proposed**	4.46	6.20	8.95
	[2]	7.65	11.65	12.12
	[9]	7.96	11.97	12.54
	[26]	5.65	9.12	10.24

Table 2 Comparison of Output SNR

Noise Type	**Solution**	**0 dB**	**5 dB**	**10 dB**
Airport	**Proposed**	3.12	3.32	4.19
	[2]	2.0	2.17	2.34
	[9]	2.12	2.54	2.87
	[26]	2.9	3.1	3.5
Babble	**Proposed**	3.5	3.67	4.21
	[2]	2.65	2.73	2.88
	[9]	2.73	2.86	2.94

	[26]	3.1	3.2	3.5
Car	**Proposed**	3.67	3.87	4.43
	[2]	2.43	2.68	3.1
	[9]	2.56	2.83	3.2
	[26]	3.1	3.2	3.5
Restaurant	**Proposed**	3.89	4.1	4.32
	[2]	2.54	2.67	2.94
	[9]	2.63	2.87	2.95
	[26]	3.1	3.4	3.7
Station	**Proposed**	3.12	3.57	3.89
	[2]	2.56	2.64	2.84
	[9]	2.62	2.84	2.96
	[26]	2.9	3.0	3.1
Street	**Proposed**	3.31	3.56	3.83
	[2]	2.17	2.67	2.87
	[9]	2.22	2.71	2.93

	[26]	3.1	3.2	3.3

Table 3 Comparison of PESQ

Noise Type	**Solution**	**0 dB**	**5 dB**	**10 dB**
Airport	**Proposed**	4.12	5.23	6.15
	[2]	3.34	3.87	4.45
	[9]	3.67	4.13	4.67
	[26]	3.8	4.1	5.4
Babble	**Proposed**	4.32	5.34	6.26
	[2]	3.36	3.91	4.41
	[9]	3.71	4.21	4.81
	[26]	3.7	4.1	4.9
Car	**Proposed**	4.43	5.51	6.35
	[2]	3.47	4	4.47
	[9]	3.8	4.34	4.68
	[26]	3.9	4.4	4.6
Restaurant	**Proposed**	4.32	5.24	6.14
	[2]	3.67	3.89	4.12

	[9]	3.91	4.21	4.43
	[26]	4.0	4.4	5.4
Station	**Proposed**	4.45	5.12	5.97
	[2]	3.63	3.92	4.31
	[9]	3.68	3.96	4.41
	[26]	4.0	4.5	4.8
Street	**Proposed**	4.82	5.12	5.57
	[2]	3.31	3.98	4.14
	[9]	3.42	4.24	4.35
	[26]	4.1	4.5	4.7

Table 4 Segmental SNR

Noise Type	**Solution**	**0 dB**	**5 dB**	**10 dB**
Airport	**Proposed**	3.52	3.81	4.59
	[2]	2.51	2.67	2.54
	[9]	2.52	3.04	3.37
	[26]	3.1	3.2	3.5
Babble	**Proposed**	3.8	4.1	4.71
	[2]	3.15	3.33	3.38

	[9]	3.23	3.16	3.44
	[26]	3.2	3.8	4.1
Car	**Proposed**	4.17	4.37	4.93
	[2]	3.93	3.18	3.6
	[9]	3.16	3.33	3.7
	[26]	4.1	4.2	4.3
Restaurant	**Proposed**	4.39	4.6	4.82
	[2]	3.14	3.17	3.44
	[9]	3.13	3.37	3.45
	[26]	3.8	4.0	4.2
Station	**Proposed**	3.52	4.17	4.39
	[2]	3.16	3.14	3.34
	[9]	3.12	3.24	3.38
	[26]	3.2	3.4	3.8
Street	**Proposed**	3.81	4.16	4.33
	[2]	2.87	3.17	3.37
	[9]	2.72	3.21	3.33

	[26]	3.2	3.5	3.8

Table 5 Comparison of PEAQ

Noise Type	**Solution**	**0 dB**	**5 dB**	**10 dB**
Airport	**Proposed**	3.63	3.93	4.64
	[2]	2.64	2.80	2.67
	[9]	2.65	3.17	3.40
	[26]	3.1	3.3	3.6
Babble	**Proposed**	3.93	4.24	4.84
	[2]	3.28	3.46	3.41
	[9]	3.36	3.29	3.57
	[26]	3.3	3.6	3.8
Car	**Proposed**	4.20	4.41	4.6
	[2]	2.9	3.3	3.7
	[9]	3.1	3.4	3.8
	[26]	3.5	3.7	3.9
Restaurant	**Proposed**	4.1	4.4	4.62
	[2]	3.2	3.4	3.6

	[9]	3.1	3.4	3.7
	[26]	3.5	3.8	4.1
Station	**Proposed**	3.5	4.1	4.3
	[2]	3.1	3.1	3.3
	[9]	3.2	3.4	3.6
	[26]	3.2	3.5	3.7
Street	**Proposed**	3.9	4.2	4.3
	[2]	2.8	3.1	3.3
	[9]	2.7	3.2	3.3
	[26]	3.1	3.4	3.7

Table 6 Comparison of PSQM

E. Impact on PESQ

PESQ was tested for several forms of noise at different input SNRs, and the results are shown in Table 3. PESQ in proposed is 47% higher than [2], 16.46% higher than [26] and 39% higher than [9]. Even at a low SNR of 0 db, PESQ is above 3 in proposed solution.

F. Impact on SEGSNR

Table 4 provides the segmental SNR data for various noise types for various input SNRs. When input SNR rises consistently across all noise types, segmental SNR rises as well. The segmental SNR exceeds [2], [26], and [9] by 38%, 12.19%, and 31%, respectively.

G. Impact on processing time

The processing time for speech enhancement is compared with three different solutions for different samples in NOIZEUS corpus and the average processing time is given in Figure 5.

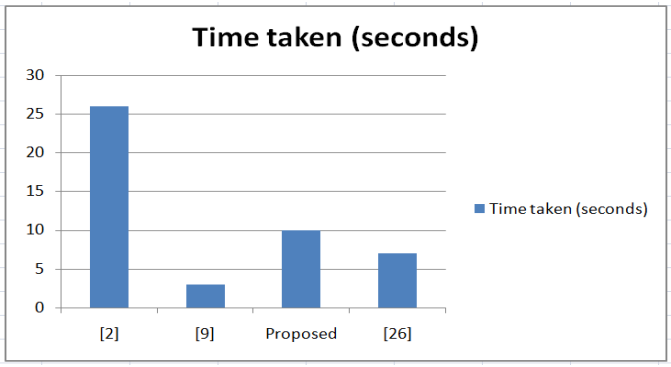

Fig. 5 Comparison of processing time

Time taken in the proposed solution is very much reduced compared to the compressive sensing solution used in [2]. It is 30% lower compared to [26]. It has reduced 63% compared to [2], but still it is higher than the noise estimation based solution proposed in [9].

H. Impact on PEAQ

The results of measuring PEAQ for various noise types at various input SNR are shown in Table 5. PEAQ in proposed is 40% higher than [2] , 23.17% higher than [26] and 34% higher than [9]. Even at a low SNR of 0 db, PEAQ is above 3.5 on an average in the proposed solution.

I. Impact on PSQM MOS Score

The results of measuring PSQM for various noise types at various input SNR are shown in Table 6. PSQM in proposed is 2 scales higher than [2] and 1 scale higher than [9] and [26].

J. Impact on MOS Score

For a total of 20 participants, the MOS score is calculated for three separate noise sources (street, car, and restaurant). Figure 6 displays the average subjective rating for each of the three forms of noise from the 20 participants.

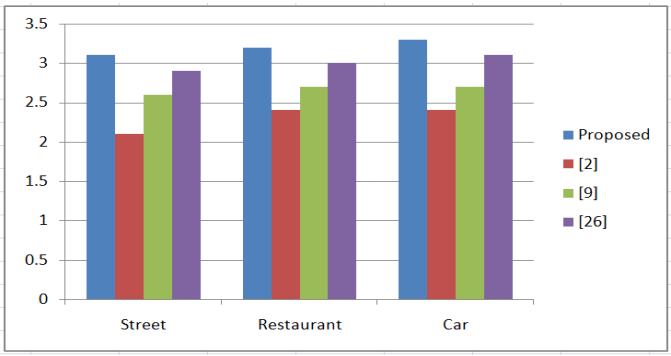

Fig. 6 Comparison of MOS

MOS score of proposed solution is 40% higher than [2], 6.25% higher than [26], and 19% higher than [9].

K. Machine Learning for Sensing Matrix Selection

The following parameters are used to train neural network for the sensing matrix and are listed in Table 7.

Input Layer Neurons	12 (10 MFCC coeff + 1 ZCF + 1 Average power)
Hidden Layer Neurons	25
Output Layer Neurons	80 (Number of measurements was set as 80)
Input and Hidden layer activation	Relu
Output layer activation	Sigmoid
Optimizer	Rmsprop

Table 7 Neural Network Parameters

The accuracy achieved over different epocs of training is given in Figure 7.

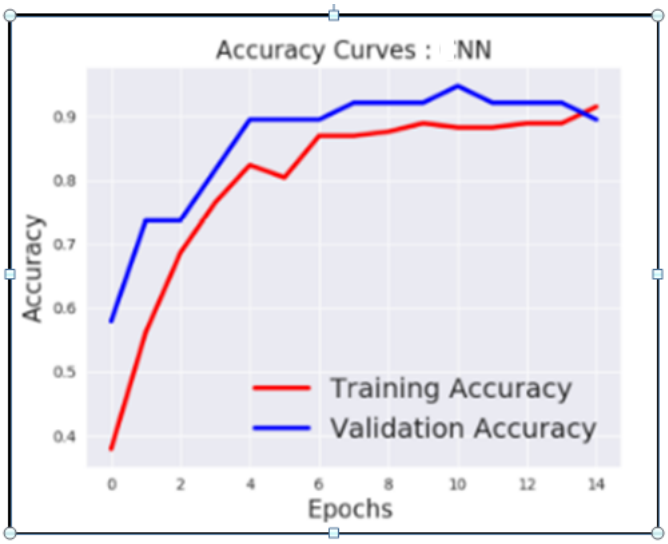

Fig. 7 Accuracy in Neural Network

An overall accuracy of 95% was achieved for sensing matrix selection by neural network.

Accuracy of neural network is tested with three different hinge loss functions.

Hinge Loss function	Objective function
Crammer and Singer	$\max(0,1+\max_{y\neq t} w_y x - w_t x)$ where t is the class label w_t and w_y are the model parameters
Weston and Watkins	$\sum_{y\neq t} \max(0,1+\max_{y\neq t} w_y x - w_t x)$
Zhang quadratically smoothed	$\begin{cases} \frac{1}{2\gamma}\max(0,1-ty)^2 \; \forall \; ty \geq 1-\gamma \\ 1-\frac{\gamma}{2}-ty \; , otherwise \end{cases}$

Fig. 8 Hinge loss functions

Figure 9 displays the neural network's accuracy for three distinct hinge loss functions.

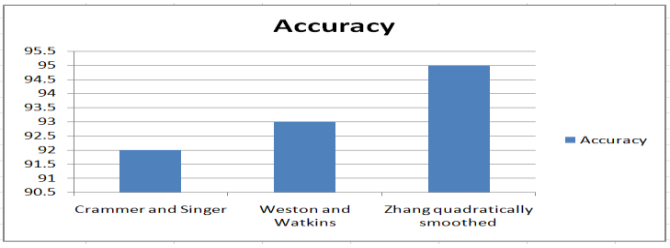

Fig. 9 Comparison of accuracy across hinge loss functions

Using Zhang quadratically smoothed hinge loss function, maximum precision is attained.

5. Conclusion

This paper proposes a way for hearing-impaired individuals to improve their speech using a combination of three distinct strategies. A gain function based on noise estimation using a priori estimation of SNR is used at the first level and noise reduction is done with it. After that, adaptive compressive sensing is carried out. Compression sensing is optimized for computation, storage, and delay by suitable binary sensor matrix selection. The sensor matrix selection was done based on the input signal features. Finally, a frequency – amplitude shaping customization gain function is applied to adjust outcome signal according to listeners comfort. The approach was able to obtain an average PESQ that was 40% higher than that of noise estimation and other compressive sensing-based speech

enhancement techniques, according to the performance findings, which also indicate a 63% decrease in processing time when compared to those other methods.

Acknowledgment

The authors are very grateful to colleagues for their important suggestions and cooperation .Also; we would like to thank our college and staff members for the constant support and facilities that are provided to us

References

[1] World Health Organization. Report of the International Workshop on Primary Ear & Hearing Care; 14 March, 2018. p. 1-19. Available from: http://www.who.int/pbd/deafness/activities/en/capetown_final_report.pdf. [Last accessed on 2018 Apr 15; Last updated on 2018 Apr 15].

[2] HANECHE, Houria&Boudraa, Bachir& Ouahabi, A. (2018). Speech Enhancement Using Compressed Sensing-based method. 1-6. 10.1109/CISTEM.2018.8613609.

[3] G. S. Bhat, N. Shankar, C. K. A. Reddy and I. M. S. Panahi, "A Real-Time Convolutional Neural Network Based Speech Enhancement for Hearing Impaired Listeners Using Smartphone," in IEEE Access, vol. 7, pp. 78421-78433, 2019, doi: 10.1109/ACCESS.2019.2922370.

[4] I. M. Panahi, C. K. A. Reddy and L. Thibodeau, "Noise suppression and speech enhancement for hearing aid applications using smartphones," *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 2017, pp. 1890-1894, doi: 10.1109/ACSSC.2017.8335692.

[5] Z. Zhang, Y. Shen and D. S. Williamson, "Objective Comparison of Speech Enhancement Algorithms with Hearing Loss Simulation," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 6845-6849, doi: 10.1109/ICASSP.2019.8683040.

[6] Shankar, Nikhil &Shreedhar Bhat, Gautam & Reddy, Chandan & Panahi, Issa. (2018). Noise dependent coherence-super Gaussian based dual microphone speech enhancement for hearing aid application using smartphone. The Journal of the Acoustical Society of America. 143. 1806-1807. 10.1121/1.5035916.

[7] Goehring, Tobias &Bolner, Federico & Monaghan, Jessica & van Dijk, Bas &Zarowski, Andrzej &Bleeck, Stefan. (2016). Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. Hearing Research. 344. 10.1016/j.heares.2016.11.012.

[8] W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor. Multichannel Kalman filtering for speech enhancement. In Proc. IEEE Intl Conf. Acoustics, Speech and Signal Processing, Calgary, Canada, Apr. 2018.

[9] N. Tiwari and P. C. Pandey, "Speech Enhancement Using Noise Estimation With Dynamic Quantile Tracking," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2301-2312, Dec. 2019, doi: 10.1109/TASLP.2019.2945485.

[10] Y. Zhao, Z. Wang and D. Wang, "A two-stage algorithm for noisy and reverberant speech enhancement," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 5580-5584, doi: 10.1109/ICASSP.2017.7953224.

[11] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, pp. 7–19, 2015.

[12] Y. Zhao, D. L. Wang, I. Merks, and T. Zhang, "DNN-based enhancement of noisy and reverberant speech," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 6525–6529

[13] J. Chen, Y. Wang, S. E. Yoho, D. L. Wang, and E. W. Healy, "Largescale training to increase speech intelligibility for hearing-impaired listeners in novel noises," The Journal of the Acoustical Society of America, vol. 139, pp. 2604–2612, 2016

[14] G. S. Bhat, N. Shankar, C. K. A. Reddy, and I. M. S. Panahi, ''Formant frequency-based speech enhancement technique to improve intelligibility for hearing aid users with smartphone as an assistive device,'' in Proc. IEEE Healthcare Innov. Point Care Technol. (HI-POCT), Bethesda, MD, USA, Nov. 2017, pp. 32–35

[15] C. K. A. Reddy, N. Shankar, G. S. Bhat, R. Charan, and I. Panahi, ''An individualized super-Gaussian single microphone speech enhancement for hearing aid users with smartphone as an assistive device,'' in IEEE Signal Process. Lett., vol. 24, no. 11, pp. 1601–1605, Nov. 2017.

[16] E. W. Healy, S. E. Yoho, J. Chen, Y. Wang, and D. Wang, ''An algorithm to increase speech intelligibility for hearing-impaired listeners in novel segments of the same noise type,'' J. Acoust. Soc. Amer., vol. 138, no. 3, pp. 1660–1669, 2015

[17] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, ''Dynamic noise aware training for speech enhancement based on deep neural networks,'' in Proc. 15th Annu. Conf. Int. Speech Commun. Assoc., 2014, pp. 1–5

[18] Y. Xu, J. Du, Z. Huang, L. R. Dai, and C. H. Lee, ''Multi-objective learning and mask-based post-processing

for deep neural network based speech enhancement,'' in Proc. Interspeech, 2015, pp. 1508–1512

[19] Ranjan, A. ., Yadav, R. K. ., & Tewari, G. . (2023). Study And Modeling of Question Answer System Using Deep Learning Technique of AI. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 01–04. https://doi.org/10.17762/ijritcc.v11i2.6103

[20] Q. Wang, J. Du, L.-R. Dai, and C.-H. Lee, ''A multiobjective learning and ensembling approach to high-performance speech enhancement with compact neural network architectures,'' in IEEE/ACM Trans. Audio, Speech, Language Process., vol. 26, no. 7, pp. 1185–1197, Jul. 2018.

[21] C. K. A. Reddy, Y. Hao, and I. Panahi, ''Two microphones spectralcoherence based speech enhancement for hearing aids using smartphone as an assistive device,'' in Proc. IEEE Int. Conf. Eng. Med. Biol. Soc., Oct. 2016, pp. 3670–3673

[22] Reddy, C.K.A, Ganguly A., Panahi, I., "ICA based single microphone blind speech separation technique using non-linear estimation of speech", IEEE Inter. Conf. Acous. Speech,, Sig. Proc. (ICASSP), 2017, pp: 5570 – 5574.

[23] P. Scalart and J. Vieira Filho, "Speech enhancement based on a priori signal to noise estimation," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Atlanta, GA, May 1996, vol. 2, pp. 629–632.

[24] M.S. ArunSankar,P.S.Sathidevi,"A scalable speech coding scheme using compressive sensing and orthogonal mapping based vector quantization",Heliyon,May,2019

[25] A. Ravelomanantsoa, H. Rabah, A. Rouane, Compressed sensing: a simple deterministic measurement matrix and a fast recovery algorithm, IEEE Trans. Instrum. Meas. 64 (12) (2015) 3405–3413

[26] https://ecs.utdallas.edu/loizou/speech/noizeus/

[27] K. Zaman, S. S. Maghdid, H. Afridi, S. Ullah and M. Zohaib, "Enhancement of Speech Signals for Hearing Aid Devices using Digital Signal Processing," *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2020, pp. 1-7, doi: 10.1109/ISMSIT50672.2020.9255299.

[28] Johnson, M., Williams, P., González, M., Hernandez, M., & Muñoz, S. Applying Machine Learning in Engineering Management: Challenges and Opportunities. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/90

[29] Garg S, Chadha S, Malhotra S, Agarwal AK. Deafness: burden, prevention and control in India. Natl Med J India. 2009 Mar-Apr;22(2):79-81. PMID: 19852345

[30] Hassager, Henrik & Wiinberg, Alan & Dau, Torsten. (2017). Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. The Journal of the Acoustical Society of America. 141. 2556-2568. 10.1121/1.4979783.

[31] Dobie RA, Van Hemel S, editors. Hearing Loss: Determining Eligibility for Social Security Benefits. Washington (DC): National Academies Press (US); 2004. 2, Basics of Sound, the Ear, and Hearing.

[32] Andrade, Adriana Neves de, Iorio, Maria Cecilia Martinelli, & Gil, Daniela. (2016). Speech recognition in individuals with sensorineural hearing loss. Brazilian Journal of Otorhinolaryngology, 82(3), 334-340.

[33] A. W. Rix, J. G. Beerends, M. P. Hollier and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221), 2001, pp. 749-752 vol.2,

[34] Salovarda, M. & Bolkovac, I. & Domitrović, Hrvoje. (2005). Estimating Perceptual Audio System Quality Using PEAQ Algorithm. 1 - 4. 10.1109/ICECOM.2005.205017.

[35] Anis Ben Aicha and Sofia Ben Jebara. 2012. Perceptual speech quality measures separating speech distortion and additive noise degradations. Speech Commun. 54, 4 (May, 2012), 517–528.

[36] Gouyon F., Pachet F., Delerue O. (2000),On the Use of Zero-crossing Rate for an Application of Classification of Percussive Sounds, in Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00 - DAFX-06), Verona, Italy, December 7–9, 2000.

[37] de Lara JRC. A method of automatic speaker recognition using cepstral features and vectorial quantization. In: Iberoamerican Congress on Pattern Recognition. Berlin, Heidelberg: Springer; 2005. pp. 146-153

[38] C. Z. Tang and H. K. Kwan, "Multilayer feedforward neural networks with single powers-of-two weights," in *IEEE Transactions on Signal Processing*, vol. 41, no. 8, pp. 2724-2727, Aug. 1993, doi: 10.1109/78.229903.

[39] Mahmmod, Basheera & H. Abdulhussain, Sadiq & Naser, Marwah & Abdulhasan, Muntadher & Mustafina, Jamila. (2021). Speech Enhancement Algorithm Based on a Hybrid Estimator. IOP Conference Series: Materials Science and Engineering. 1090. 10.1088/1757-899X/1090/1/012102.

[40] B. M. Mahmmod, A. R. Ramli, T. Baker, F. Al-Obeidat, S. H. Abdulhussain and W. A. Jassim, "Speech Enhancement Algorithm Based on Super-Gaussian Modeling and Orthogonal Polynomials," in *IEEE Access*,

vol. 7, pp. 103485-103504, 2019

Authors

First Author – Mr. Hrishikesh B Vanjari, Department of Electronic & Telecommunication, Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University , Pune, India. His area of Interest is Signal Processing (Speech Processing).

Second Author – Dr. Mahesh T. Kolte, Department of Electronic & Telecommunication, Pimpri Chinchawad College of Engineering, Savitribai Phule Pune University , Pune, India.His area of Interest is Signal Processing (Speech Processing).