IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 26 September 2022, accepted 8 November 2022, date of publication 18 November 2022, date of current version 29 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223365

RESEARCH ARTICLE

Marathi Speech Intelligibility Enhancement Using I-AMS Based Neuro-Fuzzy Classifier Approach for Hearing Aid Users

PRASHANT G. PATIL[1], TUSHAR H. JAWARE[1], SHEETAL P. PATIL[1], RAVINDRA D. BADGUJAR[1], FELIX ALBU[2], (Senior Member, IEEE), IBRAHIM MAHARIQ[3], BAHAA AL-SHEIKH[4], (Senior Member, IEEE), AND CHITTARANJAN NAYAK[5]

[1]R. C. Patel Institute of Technology, Shirpur, Maharashtra 425405, India
[2]Department of Electronics, Valahia University of Targoviste, 30004 Targoviste, Romania
[3]College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait
[4]Department of Biomedical Systems and Informatics Engineering, Yarmouk University, Irbid 21163, Jordan
[5]Department of Communication Engineering, School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

Corresponding author: Chittaranjan Nayak (83chittaranjan@gmail.com)

The work of Felix Albu was supported by the Romanian Ministry of Research, Innovation and Digitization, National Scientific Research Council (CNCS), the Executive Unit for Financing Higher Education, Research, Development and Innovation (UEFISCDI), through the National Research-Development and Innovation Plan (PNCDI) under Project PN-III-P4-PCE-2021-0780.

ABSTRACT Globally, 1.6 billion individuals suffered from hearing disability in 2019. According to the World Health Organization, by 2050, the number of people with hearing impairments will rise to 2.5 billion. Speech perception in noisy surroundings is a challenge for hearing aid users. This study aimed to design a novel methodology to improve the speech recognition ability of hearing aid users from various backgrounds. To improve speech enhancement, we propose a discrete cosine transform (DCT)-based improved amplitude-magnitude spectrogram (I-AMS) algorithm with a fuzzy classifier. First, the I-AMS approach disintegrates speech signals containing noise into time-frequency units and eliminates the noise present in the signal. Next, the time frequency units (t-f units), modulation frequency (f_m), and centre frequency (f_c) are extracted from the denoised signal. A neuro-fuzzy classifier was used to classify the background speech environment into three different classes. The proposed I-AMS algorithm was tested, achieved improvements in terms of sensitivity (+1.02%) and accuracy (+11.80%). Speech denoising revealed a 1.27% improvement in speech recognition performance.

INDEX TERMS Improved amplitude magnitude spectrogram, insertion gain, intelligibility, marathi speech, neuro-fuzzy classifier, time-frequency units.

I. INTRODUCTION

Speech is a form of human communication, and over the past 50 years, speech recognition has become a fascinating research field. Speech is the core of human activity because it helps humanity collaborate in a common and viable manner [1]. Approximately 13% population in developed countries suffers from hearing deficiencies. These factors influence communication abilities and prevent normal living [2]. In addition, approximately 25% users avoided the use of hearing aids owing to bothering and repulsive shrieks.

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Pau.

Most hearing aids (HA) are designed for single-background environments. In a noisy background, the signal and noise are amplified similarly [3]. Hearing aids are electroacoustic devices that improve the audibility of individuals with hearing impairment. The main objective is to increase speech intelligibility through amplification to achieve a better hearing aid performance [4]. However, this procedure typically increases the sound power levels in each frequency band, including the hearing thresholds of the user, which has no noticeable benefit [5]. To avoid this, a frequency-lowering technique was used to transfer the frequency band from the dead (impaired) to the audible band [6], [7]. Speech playback at a lower rate than sampling is a unique method of frequency lowering that degrades the speech quality [8]. Several approaches have been proposed to overcome the challenges of speech recognition. The speech recognition process was developed using the hidden Markov model [9] and the stereo vision neural network model [10]. Some known speech denoising methods include Wiener filtering [11], spectral subtraction algorithms [12], and subspace filtering [13], which have attracted substantial attention and exploration owing to their simple designs and implementations. The orthogonal-polynomial-based speech enhancement algorithm emphasizes the development of a minimum low-distortion estimator for speech and noise data signals. The observed signal was transformed into the transform domain using an orthogonal polynomial [14]. During speech processing, these linear approaches minimize noise while simultaneously enhancing the signal-to-noise ratio (SNR). The Support vector machine (SVM) approach [15] has been proven to improve the oversimplification capability of the classifier. Speech recognizers are generally calibrated to avoid mismatches during the recognition period, such as minimal distinction malfunctions [16].

This study proposes a novel combinational feature extraction and classification approach to increase the speech intelligence of hearing aid users in the Marathi language. Numerous commercial hearing aids cannot adapt to acoustic environmental changes or background conditions. The focus of this study was to design a speech background classifier that helps improve the auditory performance of HA users under different background conditions. First, we decomposed the input speech into discrete t-f units and reduced the signal noise. Next, we extracted useful features such as the time frequency unit (t-f), center frequency (fc), and modulation frequency (fm) from the Marathi speech using the improved amplitude magnitude spectrogram (IAMS) technique. The input speech feature values were categorized into corresponding classes based on the ratio [17]. This ratio was determined using a neuro-fuzzy classifier based on approximate and original spectral values. A window function was applied, followed by weighting and addition of the corresponding mask value to obtain an enhanced signal. Quality improvement of speech includes the recognition of syllables, monosyllables, vowels, consonants, words, short sentences, and phonemes by hearing aid users under different speech background conditions. Marathi speech samples were collected from participants of different sexes and speech backgrounds. A novel contribution of the proposed approach lies in feature selection for speech enhancement in HA. The features are selected using a discrete cosine transform (DCT)-based improved amplitude magnitude spectrogram (I-AMS) algorithm, which reduces speech processing time. The neuro-fuzzy classifier categorized the denoised speech signal into four classes: target, target-dominated, masker-dominated, and masker.

This paper is organized into six sections. In Section 2, state-of-the-art literature is provided. The proposed approach for improving speech intelligibility is described in detail in Section 3. Section 4 illustrates the collection of databases, experimentation process, and audiogram analysis. Section 5 focuses on the signal enhancement, classifier performance, and recognition results. Finally, in Section 6, conclusions and future work are presented.

II. LITERATURE REVIEW

Numerous researchers have assessed the hearing loss on certain frequencies.

Ching et al. [18] clarified the speech perception of people with hearing disabilities and calculated their speech intelligibility index (SII). Customization of the SII is considered to boost the accuracy [19]. The index scale was considered inadequate in a recursive recognition skill test, and alternative improvements were proposed [18]. Satisfactory outcomes were obtained using this system, in which the amount of distortion with the audible frequency capability of the user was merged. This approach has been assessed for syllables and sentences. Moreover, in [20], noise reduction techniques for speech quality improvement useful for hearing-impaired (HI) individuals were examined. Speech quality was improved using the shrinkage sparse coding (SSC) technique [21]. In this method, the examination is extended to contain speech quality evaluations using interrelated comparative ranking (ICR) [22].

In [23], improved high-frequency speech intelligibility in noise was proposed, and sound sources were located on a horizontal plane with high accuracy. First, speech frames are decomposed into three groups of speech models: amplitude, frequency, and phase [24]. The input speech frequency above the reference cut-off frequency (f_c) was reallocated towards a lower frequency range to improve high-frequency speech recognition ability [25]. Frequency compression ratio (CR) was set for various frequency ranges. To prevent spectral distortion of speech, the input spectrum was categorized into six octaves [26]. Furthermore, Matthias et al. [27] developed the F_0 modulation $F_{0\ (mod)}$ processing technique for the cochlear implant (CI). This approach offers F0 (mod), which enhances the spectral pitch cue by performing intensity modulation of multichannel electrical stimulation [28]. The input speech signal F0 (fundamental frequency) was used. This approach has been verified for recognition at word and sentence levels in various noise-level situations.

Table 1 depicts existing speech processing methods [1], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39] and their processing strategies, along with their highlights and limitations for vowel, consonant, word, and sentence recognition.

Cochlear implant (CI) speech perception systems, where acoustic listening was proposed to replicate the effects of speech-in-noise intelligibility, have been proposed in [40] and [41]. In relation to electrical and/or acoustic stimuli, a model has been used to simulate the neurons of the auditory system [42]. Positioning and spatial-temporal structured spiking variations were employed as inner illustrations of the noisy voices. For signals including stagnant noise, speech reception thresholds were predicted for a sentence [43] and

TABLE 1. Comparative study of existing speech processing algorithms.

Signal Processing Method/s	No. of Hearing Aid Users	Recognition Tests Conducted for	Outcomes, Results & Limitations
Channel vocoder [29,30]	8	word acknowledgement, vowel, and consonant perception,	More training needed to adopt the technology
AVR transonic (slow playback) [1]	4	Vowel and consonant recognition, Spondees, Monosyllabic words &sentence recognition.	2 participants enhanced their sentences using the device by an average of 22%.
Linear Frequency shifting [31]	15	Testing of consonant-vowel nonsense syllables in a closed set, Monosyllabic words	50% group revealed enhancements (7% on average for female speakers). There were no statistically significant improvements for the group in frequency shifted speech.
Nonlinear frequency compression [32] (With Speech Gain)	5	Sentences in Noise, Consonants and Monosyllabic words	Not useful improvement found for any participant.
Nonlinear frequency compression [33] (Without Speech Gain)	17	Monosyllabic words	Phoneme score is improved by 6% by using this device.
Linear Transposition [34]	13	Nonsense syllables, Monosyllabic words, sentences in quiet and in noise.	17% improvement was reported by 2 Hearing aid users.
Frequency Transposition & Compression [35]	30	Vowels Consonants Syllables Mono Syllables Words Short Sentences	FC reported 91% improvement while FT reported 70%.
Non-Linear Transposition [36]	8	Monosyllabic words, nonsense syllables	50% participants show higher frequency outcomes in positive direction
Electrical Amplifier used in HA [36]	07	Quiet & noisy Sentences, Words, Monosyllabic words.	Bulky, not comfortable to use
Self-Adaptive multilane Filter [37]	10	Alphabets recognition, Sound test and plural recognition	Gain depends on the level of speech and noise. Few participants have indicated improvements
Smart Noise Suppression Method [38]	23	Vowels Consonants Syllables Mono Syllables Words Short Sentences	Improves ability to hear under adverse listening conditions.
LMS Adaptive filter [39]	12	Vowels Consonants Syllables Mono Syllables Words Short Sentences	Depends on acoustic feedback and path more gain is needed to improve speech in the audible range.
Automatic gain control for HA Processing [39]	25	Speech Intelligibility testing for Aided thresholds.	For female speaker 50% HA user showed improvements. This system avoids Threshold in hearing, More Softer Sounds at HA end.

checked for an automatic method of speech recognition in [44].

Furthermore, in [45], interaural differences in terms of intensity, time, and phase were introduced. The sound arrival time (S_{at}) and sound arrival level (S_{al}) allow the distinction of sounds in the horizontal planes, and these parameters are helpful for source separation and speech perception in environments with background noise [46], [47]. Sound localization behaviour was examined using 14 bimodal uses; all users adopted the same CI, the advanced Phonak hearing aid. The primary objective was to find binaural and monaural cues for horizontal sound-source localization [48].

In [49], speech intelligibility in terms of inconsistency in cochlear implant (CI) users was investigated. Speech comprehension in various environments with background noise was investigated in [46] and [50]. They utilized a speech amplification algorithm [51] based on a neural network to increase the speech perception in the presence of background noise. The noise was separated from the speech signal and converted into time-frequency units [52], [53]. A neural network was used for channel frequency estimation in [54].

There is still a need for improvement in terms of the gain frequency correlation, noise cancellation, acoustic feedback, and signal processing delay. Many algorithms that are designed and implemented for single-background environments are not useful for other speech backgrounds. Input speech denoising, insertion gain, and feature extraction required for classification are key parameters of the proposed method.

III. THE I-AMS BASED SPEECH SIGNAL ENHANCEMENT TECHNIQUE

The proposed approach focuses on four phases: preprocessing, feature extraction, training-testing rate for classifiers, and speech enhancement. Figure 1 shows the overall diagram of the proposed neuro-fuzzy classifier for speech enhancement. First, noise was removed from the input speech. Subsequently, an improved amplitude-magnitude spectrogram (I-AMS) technique is used. The extracted features are trained using a neuro-fuzzy classifier. During the training phase, the noise-masked signal t-f units [55] were classified into four classes: target class *(class_1),* target-dominated class *(class_2),* masker class *(class_3),* and masker-dominated class *(class_4).*

During the enhancement phase, the noise-masked signal has individual t-f units, which are multiplied by the equivalent class weight to obtain the enhanced speech waveform.

A. PRE-PROCESSING

Pre-processing in I-AMS is an important stage in speech enhancement and consists of four steps: pre-emphasis, Automatic Gain Control (AGC), FFT filter bank, and envelope extraction.

1) PRE-EMPHASIS

Pre-emphasis is the first stage involved in the pre-processing. The input signal may have different frequency components that fall between high and low frequencies. To avoid high frequencies and compensate for the high-frequency components [2], we use a pre-emphasis filter whose pre-emphasis factor 'α' is selected in the range of 0.9 1. In the pre-emphasis phase, the high-frequency speech components are amplified to a higher magnitude than the noise components, which helps improve the signal-to-noise ratio (SNR).

2) AUTOMATIC GAIN CONTROL

After the pre-emphasis stage, the filtered output signals were passed through automatic gain control (AGC) [56]. Amplification converts soft, moderate, and loud sounds to audible ranges. The AGC controls the gain during processing according to the background environment of the speaker and the HA user. It is improved using a dual-loop AGC that contains low-gain AGC for level deviation and high-gain AGC for severe deviation. In sentence assessments, the dual-loop AGC offered a better speech understanding ability than the fast AGC approach.

3) FFT FILTER BANK

The compressed signal was transformed using FFT. It computes the real and imaginary parts of the signal, where it decomposes the N' point time-domain signal into the frequency domain. Then it calculates corresponding 'N' point frequency spectra and convolves into a single frequency spectrum.

B. FEATURE EXTRACTION USING IMPROVED AMPLITUDE MAGNITUDE SPECTROGRAM (I-AMS)

Feature extraction is the most significant step in the I-AMS technique, as illustrated in Fig. 2. The signals were sampled, bandpass filtered, rectified, and segmented. In this method, we used the discrete cosine transform rather than the normal Fourier transform [57]. We also added delta functions to improve the feature vector values by using Equation (1).

$$\Delta F_T(\lambda, t) = F_S(\lambda, t) - F_S(\lambda, t-1) \tag{1}$$

where F_T is the transformed frequency and F_S is the speech frequency. Dataset (D) of the Marathi speech samples was partitioned into the training (D^{TR}) and testing $\left(D^{TS}\right)$ datasets.

The input signal $I\,(t)$ contains both the clean signal $C(t)$ and noise signal $N(t)$ as indicated in Equation 2.

$$I(t) = C\,(t) + N\,(t) \tag{2}$$

During the sampling process, the continuous time signal is transformed into a discrete-time domain at a sampling rate of 16 KHz [58], [59]. The input signal $I\,(t)$ in Equation (2) is sampled as $I\,(n)$ as shown in Equation (3).

$$I\,(t) = I\,(n.T)\,, \quad where\ n = 0, 1, 2, \ldots \tag{3}$$

The time duration of each frame with 320 samples was 20m-sec with an overlap of 50%. Rounding and truncation are widely used in quantization processes. We implemented a 6-bit quantization process in which quantization improvement was performed using the μ law. The quantized signal is processed through a pre-emphasis phase to enhance the power level, in which emphasis is placed on the higher frequency contents of the signal compared to the lower one. SNR improvement is achieved by limiting the undesirable effects of saturation and attenuation losses [60]. The SNR of the nth frequency band was calculated using Equation (4).

$$\bar{R}_n = R_n - \alpha(R_{n-1}) \tag{4}$$

where $\alpha = 0.95$ and $\bar{R}_n$ is n^{th} frequency. The pre-emphasized signal was passed through a band-pass filter with 25 channels. The processed signals were separated into different time-frequency units (t-f) using bandpass filters. The signal is converted into 25 different t-f units where each t-f unit related to corresponding channel which is represented by 'C_i' where$1 \leq i \leq 25$. Each channel has a corresponding upper and lower cut-off frequency. After complete wave rectification, the envelope of each band was decimated using 3. The decimation envelope was divided into 128 intertwining segments, which had 32 ms segments with 64 overlapping samples per frame. Each segmented signal is defined by 'S_{ij}' where $1 \leq i \leq 25$, $1 \leq j \leq N_i$, w here, N_i is the number of segments related to the ith channel. The sampled signals are windowed using the Hanning window with a 25ms window

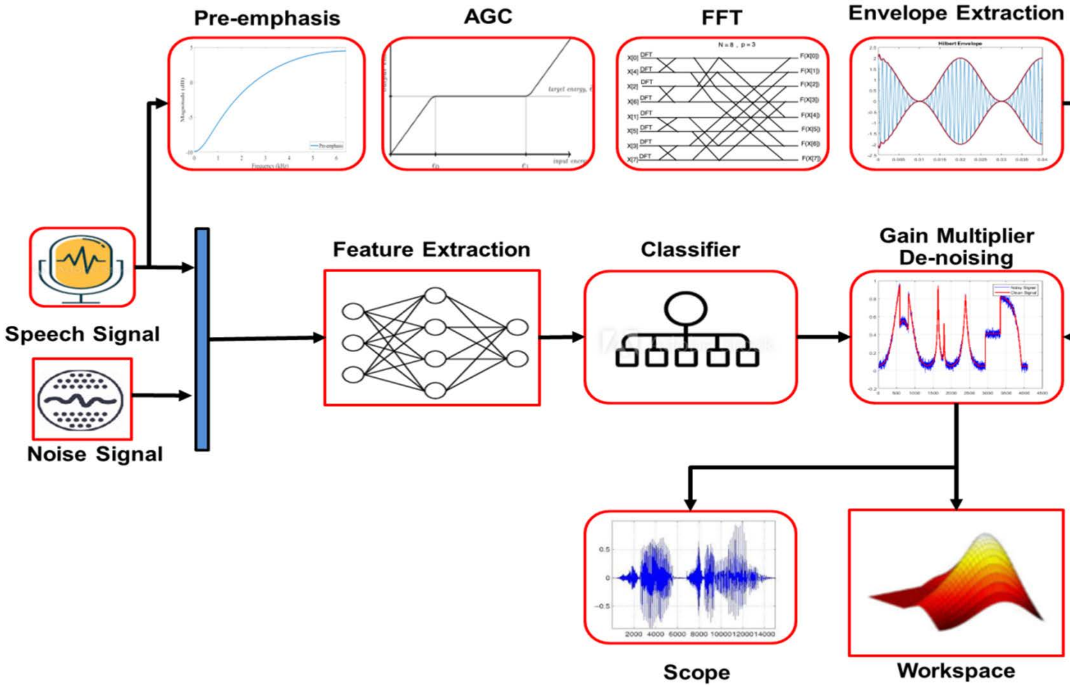

FIGURE 1. The proposed Neuro fuzzy classifier and I-AMS Speech enhancement system.

size that eliminates spectrum artifacts [61]. The window function is defined by Equation (5):

$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\Pi n}{N-1}\right)\right] \tag{5}$$

where N is the sample width, and n is an integer from 0 to (N-1). Zero padding and DCT have also been used [62]. In terms of the number of cosine functions at various frequencies [6], a discrete cosine transform (DCT) conveys a limited-magnitude sequence at different DCT data points. The DCT was applied to the input signal using (6).

$$Y_{\mathrm{t}} = \sum\nolimits_{t=0}^{N-1} y_n.\left[\cos\frac{\pi}{N}\left(n+\frac{1}{2}\right).t\right] \tag{6}$$

where $t = 0, \ldots, N-1$

The DCT computes the modulation spectrum for each of the 25 channels, and each channel is duplicated by 15 triangular windows [63] within a range of 15.6–400 Hz (Equation (7)).

$$w(n) = \frac{2}{N-1}\left[\left(\frac{N-1}{2}\right) - \left|n - \frac{N-1}{2}\right|\right] \tag{7}$$

These spectrum amplitudes are summed, and each describes the feature vector $F_s(\rho, t)$, where t is the time slot and ρ corresponds to the subband.

We included delta functions in the extracted features to consider shifts in the time and frequency domains [64], [65], where the delta function is expressed in Equation (8).

$$\Delta F_T(\rho, t) = F_S(\rho, t) - F_S(\rho, t-1) \quad \text{For } t = 2, \ldots, T \tag{8}$$

The delta function [66] in terms of frequency is defined by Eq. (9).

$$\Delta F_B(\rho, t) = F_S(\rho, t)/F_S(\rho-1, t) \quad \text{For } \rho = 2, \ldots, B \tag{9}$$

For $t = 1$, Equation (9) is expressed as

$$\Delta F_T(\rho, 1) = F_S(\rho, 2) - F_S(\rho, 1)$$

For $\rho = 2$ Equation (9) becomes

$$\Delta F_B(1,t) = F_S(2,t)/F_S(1,t).$$

The overall feature vector [67] is expressed using the delta function:

$$A_s(\rho, t) = [F_s(\rho, t), \Delta F_T(\rho, t), \Delta F_B(\rho, t)] \tag{10}$$

We selected 25 sub-bands: B. Because $a_S(b, \tau)$, $\Delta a_T(b, \tau)$ and $\Delta a_B(b, \tau)$ have dimensions of 15, the total feature vector dimension of $A_S(b, \tau)$ is 45.

C. NEURO-FUZZY CLASSIFIER TRAINING

Each input t-f unit is classified into the corresponding class [68]. In the proposed method, the signal was classified into four classes: masker, masker-dominated, target-dominated, and target classes. The quality ratio classes are represented as Q1, Q2, Q3 and Q4, respectively. Let us consider the noisy speech spectrum $N(b, \tau)$ at time slot τ and sub-band$b$. The signal spectrum $\bar{Y} = (b, \tau)$ is estimated by multiplying the gain function [72], [73] with the noisy speech spectrum $N(b, \tau)$ at time slot τ and sub-band$b$ using Equation (11).

$$\bar{Y}(b, \tau) = G(b, \tau).|N(b, \tau)| \tag{11}$$

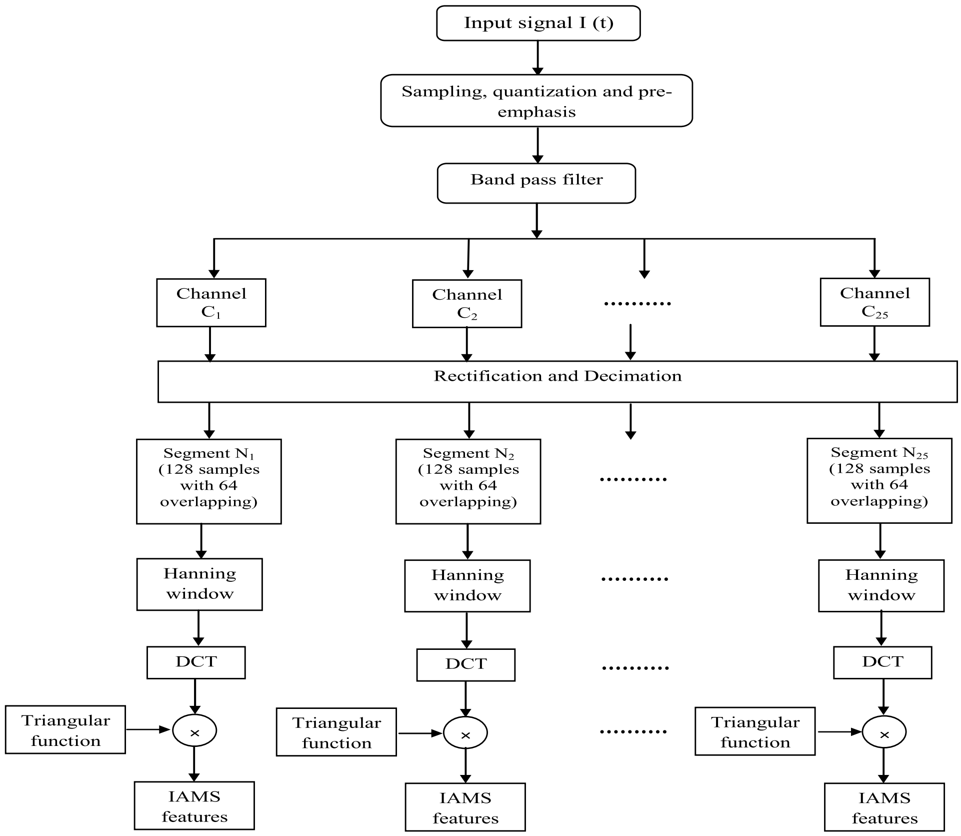

FIGURE 2. The block schematics of the I-AMS feature extraction.

The gain $G(b, \tau)$ is calculated using Equation (12).

$$G(b, \tau) = \sqrt{\frac{SNR_p(b, \tau)}{1 + SNR_p(b, \tau)}} \tag{12}$$

The prior signal-to-noise ratio [71] is SNR_p and is computed using Equation (13).

$$SNR_p(b, \tau) = \frac{\alpha.|\bar{Y} = (b, \tau - 1)|^2}{\lambda_D(b, \tau - 1)} + (1 - \alpha) \cdot \max\left[\frac{|N(b, \tau)|^2}{\lambda_D(b, \tau)} - 1, 0\right] \tag{13}$$

where the smoothing constant is $\alpha = 0.98$ and the background noise variance estimation is λ_D.

The estimated magnitude of speech was compared to the actual speech magnitude [72], and Equation (14) was used for the corresponding t-f unit.

$$Q = \frac{|\bar{X}(b, \tau)|}{|S(b, \tau)|} \tag{14}$$

During the training stage, the four different classes shown in Equation (15) were used:

$$G(b, \tau) = \begin{cases} Q1 & if\ t - f\ units\ belongs\ to\ class\ 1 \\ Q2 & if\ t - f\ units\ belongs\ to\ class2 \\ Q3 & if\ t - funits\ belongs\ to\ class3 \\ Q4 & if\ t - f\ units\ belongs\ to\ class4 \end{cases} \tag{15}$$

where Q_1 is the masker class, Q_2 is the masker-dominated class, Q_3 is the target-dominated class, and Q_4 is the masker, masker-dominated, target-dominated, and target classes, respectively. The DCT provides significantly higher energy compaction than the DFT. We collected 8100 Marathi speech samples from female and male (14 female, four male) speakers in different speech background conditions. For neuro-fuzzy classifier training and testing purposes we used 70-30%, 60-40% and 80-20% data from collected samples.

D. ENHANCEMENT MODULE

After classifier training, the pre-processing noise input signal is convolved with the calculated optimal binary value.

The proposed waveform synthesis technique is illustrated in Fig. 3. The predicted class [73] produces gain $G(b, \tau)$ of the mask represented in Equation (16).

$$G(b, \tau) = \begin{cases} 0.00, & if\ t - f\ units\ belongs\ to\ class\ 1 \\ 0.33, & if\ t - f\ units\ belongs\ to\ class\ 2 \\ 0.66, & if\ t - f\ units\ belongs\ to\ class\ 3 \\ 1.00, & if\ t - f\ units\ belongs\ to\ class\ 4 \end{cases} \tag{16}$$

IV. EXPERIMENTATION AND TESTING

This section presents a detailed analysis of the proposed method of experimentation and testing using hearing-aid

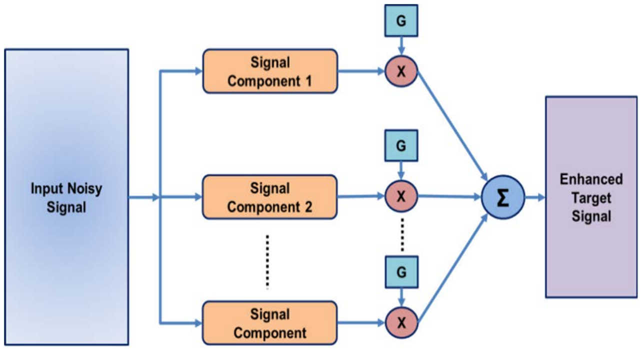

FIGURE 3. Block schematics of the I-AMS feature extraction.

TABLE 2. Statistics of recorded marathi speech dataset.

Recording Language	Marathi
Age group	7-25
Number of Speaker	18
Genders Count	14 Female 4 Male
Consonants	35
Vowels	15
Rhyming Word (Pair of Confusing words)	100
No of Speech Backgrounds	3
Samples Collected from each Speaker	450
Total No of Samples	8100

users. The selection of hearing aid users with relevant audiogram analysis is a key stage in the experimentation. Each participant was accurately examined to determine their audiogram response at a particular decibel frequency level.

A. RECORDING SPEECH DATASET

Marathi letters, words, short sentences, and rhyming words were recorded in three main situations: a silent room, speech with a musical background, and speech with fan noise. The speech dataset statistics and sample details are listed in Table 2.

B. EXPERIMENTATION FLOW

Figure 4 illustrates the experimental flow for the speech intelligibility measurements. Each speech-processing method was examined in terms of the performance parameter (scientific) and the recognition approach (developmental). The recognition scores of all hearing-aid users were measured using all the processing techniques.

C. PARTICIPANTS SELECTION PROCESS AND AUDIOGRAM ANALYSIS

Candidate selection and system design verification procedures were performed in accordance with the clinical test practice suggested by the hearing aid manufacturers' standards; 12 hearing aid users participated in testing: seven in the 7–14-year age group and five in the 14–17-year age group. The participants were selected from the NGO operated Priyadarshini deaf residential school, Shirpur (M.S) located in North Maharashtra region. All selected participants had mild to moderate hearing loss. The participants were categorized into two groups according to their sex. This categorization was useful for detecting the impact of words spoken by female speakers, sentence recognition, and intelligence ability. The audiologist's outcome and parameter fitting process revealed the dead region and patient's requirements. The use of audiologist outcomes helped us to select the essential set of performance parameters for the algorithm to avoid overfitting.

V. RESULTS AND DISCUSSIONS

In this study, the amplitude–frequency variation for noisy input and enhanced (de-noised) signal and the neuro fuzzy classifier performance parameter were calculated for different training and testing rates. After designing and implementing the proposed algorithm, recognition tests are performed under different conditions for a group of HA users.

A. SIGNAL ENHANCEMENT RESULTS

Denoising was the primary stage of this section. The proposed classifier was used to identify the class of incoming signals.

The signal-to-noise ratio (SNR) variation was plotted for various collected samples and compared with the existing pitch-intensity-based neural network approach [5], [7], [28] in terms of the SNR for various collected samples, as shown in Figs. 5 and 6. Both methods showed an SNR improvement over the existing neural-network approach for different backgrounds. The proposed denoising approach achieved a maximum SNR of 25 dB, whereas traditional neural-network-based speech intelligibility achieved a maximum SNR of 23 dB in a silent room situation.

The proposed I-AMS based Neuro fuzzy classifier approach achieves a maximum SNR of 13db while neural network-based speech intelligibility achieves a maximum SNR of 11 dB under a music background situation.

Insertion Gains for Speech (IGSPXX): Insertion gains (IG) are required to maintain the processed signal up to the requirements of a hearing aid user [76]. The insertion gain was estimated using an audiogram. The insertion gain (dB) response over the channel center frequency is shown in Fig. 7. Audiogram observations and frequency gain functions were incorporated into the proposed algorithm.

B. NEURO FUZZY CLASSIFIER PERFORMANCE

The neuro-fuzzy classifier performance results were measured using the following parameters: sensitivity [76], specificity, classification accuracy, false positive classification rate (FPCR), false negative classification rate (FNCR), false acceptance classification rate (FACR), false rejection classification rate (FRCR), positive estimation value (PEV), negative

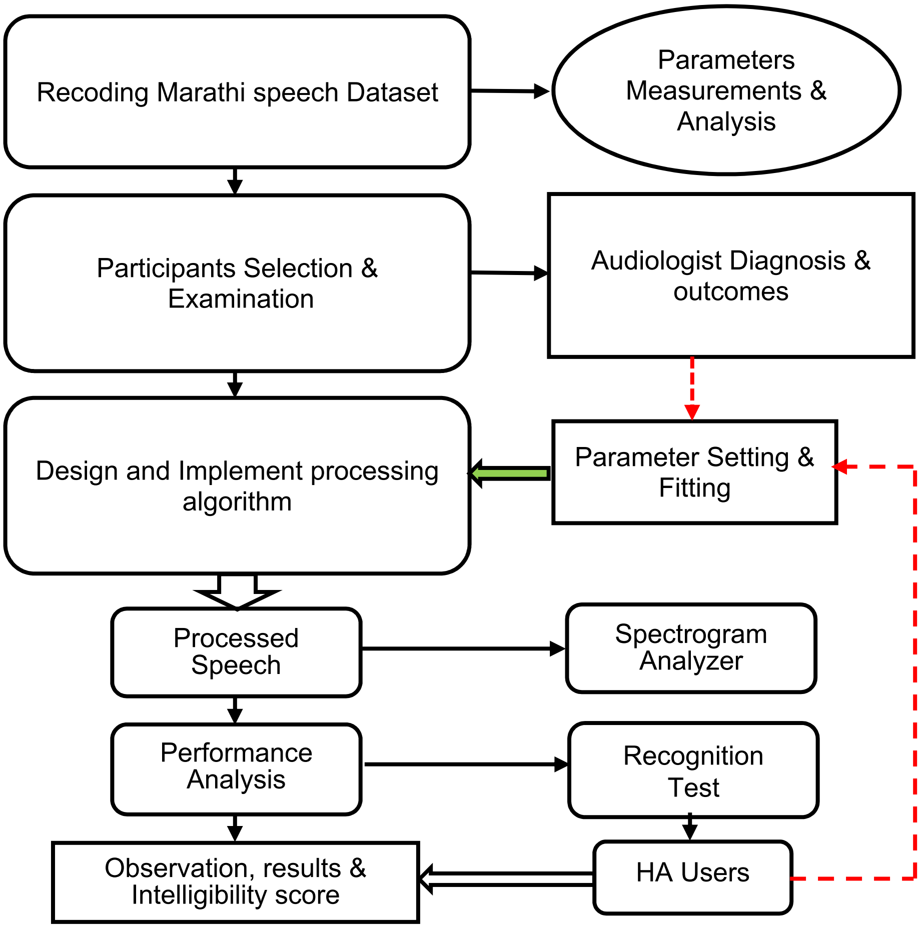

FIGURE 4. Speech intelligibility experimental flow and testing algorithm.

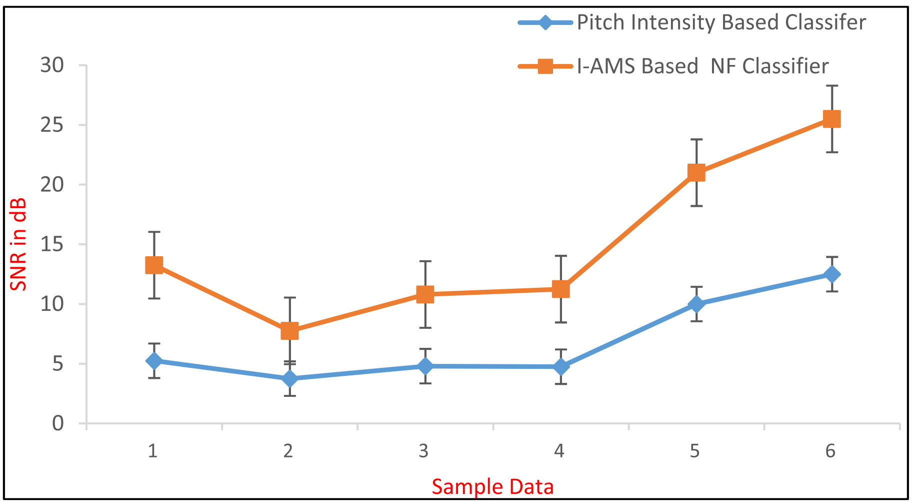

FIGURE 5. SNR comparison for the collected samples in a quiet room.

estimation value (NEV), and the Mathews correlation coefficient (MCC). These performance parameters were calculated for the different training and testing ratios of the neuro-fuzzy classifier. Sensitivity refers to the ability of the classifier to classify correctly; specificity [77] is a measure of the capability of the classifier to correctly classify negative signals [78]; and accuracy is given by Equation (17).

$$Accuracy = \left[\frac{\sum Truepositive + \sum Truenegative}{\sum no\ of\ samples}\right] \quad (17)$$

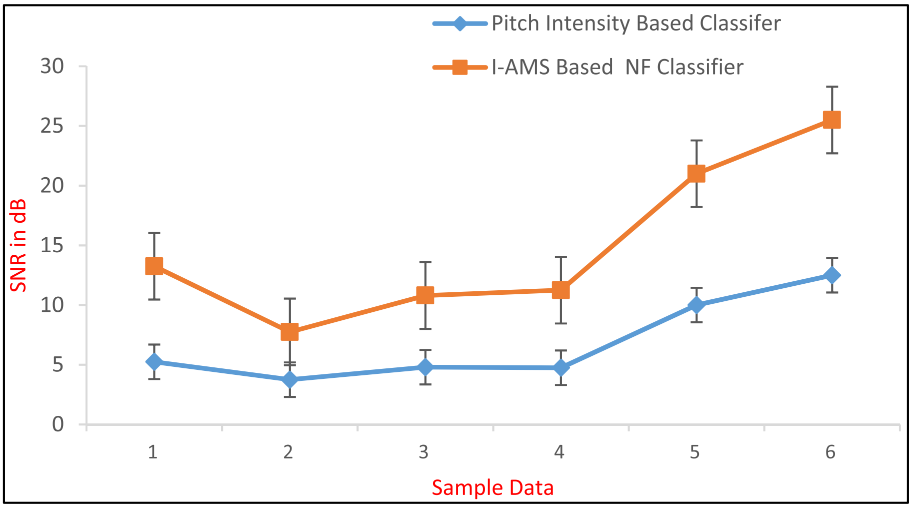

FIGURE 6. SNR comparison for the collected samples having a music background.

TABLE 3. Performance parameters measurement for different training and testing rates.

Parameters	t-f unit	Fc	Fm	t-f unit	Fc	Fm	t-f unit	Fc	Fm
	Case-I 80-20% Training-Testing Data			**Case-II 70-30% Training-Testing Data**			**Case-III 60-40% Training-Testing Data**		
Sensitivity	0.9843	0.9761	0.94444	0.9946	1	0.98404	0.98406	1	0.94422
Specificity	0.5925	0.2777	0.24528	0.225	0.1125	0.0625	0.19626	0.1215	0.08490
Accuracy	0.8666	0.7666	0.73743	0.76493	0.7350	0.70896	0.7486	0.7374	0.68908
FPCR	0.4074	0.7222	0.75472	0.775	0.8875	0.9375	0.80374	0.8785	0.91509
FNCR	0.0158	0.0238	0.05555	0.00531	0	0.01595	0.01593	0	0.05577
FACR	0.4074	0.7222	0.75472	0.775	0.8875	0.9375	0.80374	0.8785	0.91509
FRCR	0.0158	0.0238	0.05555	0.00531	0	0.01595	0.01593	0	0.05577
PEV	0.8493	0.7592	0.74843	0.751	0.7258	0.71154	0.74174	0.7275	0.70958
NEV	0.9411	0.8333	0.65	0.94737	1	0.625	0.84	1	0.3913
MCC	0.6752	0.3879	0.27494	0.39169	0.2857	0.12515	0.32389	0.2973	0.05420

False positive classification Rate (FPCR) is defined by Equation (18)

$$FPCR = \frac{False\ Positive}{Actual\ Negative} = \frac{FP}{TN + FP} \tag{18}$$

False negative classification Rate (FNCR) is defined by Equation (19)

$$FNCR = \frac{False\ Negative}{Actual\ Negative} = \frac{FN}{TN + FP} \tag{19}$$

False Acceptance classification Rate (FACR) is defined by Equation (20)

$$FACR = \frac{False\ Positive}{Total\ No\ of\ Attempts} = \frac{FP}{N} \tag{20}$$

False Rejection classification Rate (FRCR) is defined by Equation (21)

$$FRCR = \frac{False\ Negative}{Total\ No\ of\ Attempts} = \frac{FN}{N} \tag{21}$$

Positive Estimation Value (PEV) is defined by Equation (22)

$$PEV = \frac{True\ Positives}{True\ Positives + False\ Positive} = \frac{TP}{TP + FP} \tag{22}$$

Negative Estimation Value (PEV) is defined by Equation (23)

$$NEV = \frac{True\ Negative}{True\ Negative + False\ Negative} = \frac{TN}{TN + FN} \tag{23}$$

TABLE 4. Short sentence correct recognition rate comparison.

HA User Count	Proposed Techniques with Neuro Fuzzy Approach					Existing Techniques [7]	
	Without Enhancement	With Enhancement	With Enhancement and Extracting			FC	FT
			fm	fc	t-f unit		
L1	50.60	61.30	65.30	65.30	67.30	53.60	60.50
L2	50.00	50.60	57.30	56.00	53.30	48.30	58.50
L3	41.30	35.30	64.00	68.00	68.80	52.70	60.50
L4	49.30	36.66	68.60	62.00	64.30	60.50	63.80
L5	40.60	59.30	60.00	76.60	77.00	65.80	70.80
L6	50.00	48.60	53.33	58.60	74.60	72.90	75.20
% Average recognition Score	47.33	48.60	61.40	64.40	66.50	58.96	64.88

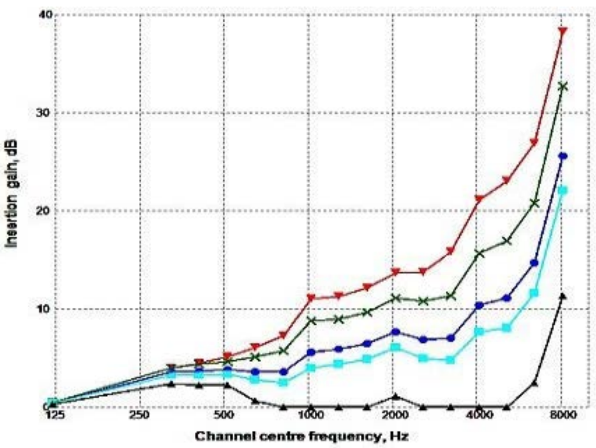

FIGURE 7. Insertion gain (dB) response over the channel center frequency.

The Matthews correlation coefficient (MCC) is widely adopted in machine learning to calculate the excellence of binary (two-class) classification. MCC is the relationship coefficient between experiential and forecasted two-stage classifications [79]. It has a value between -1 and $+1$.

$$\begin{aligned} Where; \quad & coefficient\ of\ +1 = Correct\ estimation \\ & coefficient\ of\ -1 = Incorrect\ estimation. \\ & coefficient\ of\ 0 = Randome\ stimation \end{aligned}$$

Table 3 shows the performance parameters of the I-AMS-based classifier by extracting different features as t-f units, center frequency, and modulation frequency for different training and testing rates. The target (higher) frequency band is linearly compressed to a lower (audible) frequency This approach is designed using critical bark-band scaling, which reduces spectral loss in the lower speech frequency range [47].

C. MARATHI LANGUAGE RECOGNITION TEST RESULTS

The 7- to 12-year-old participants were randomly divided into two groups.

Figures 8 and 9 show the average vowel and consonant recognition scores calculated for the proposed method by extracting the t-f unit, center frequency, modulation frequency, and individual hearing aid. In these experiments, each vowel and consonant were played randomly multiple times with different speaker and listener backgrounds. For the short-sentence recognition test, six listeners were selected from existing 12 users. These participants were selected based on their perception of the highest individual recognition rate during the vowel and consonant tests. A short-sentence recognition test was conducted with different backgrounds of the speakers and listeners.

The overall recognition score was measured for speakers and listeners in the quiet-quiet, quiet-crowded, crowded-quiet, and crowded-crowded rooms. Table 4 shows that the individual recognition score calculated for five different cases, which indicates that the highest rate of approximately 67% was achieved after denoising and t-f unit extraction.

In Fig. 10, the SNR variation comparison for the proposed technique indicates that the processing method retains the SNR level and reduces the insertion gain requirement.

In Table 5, the processed speech after t-f unit extraction retains a SNR level between minimum 50.6 dB to maximum 67.3 dB which yields a better speech quality for HA users, while other speech features as centre frequency (fc)

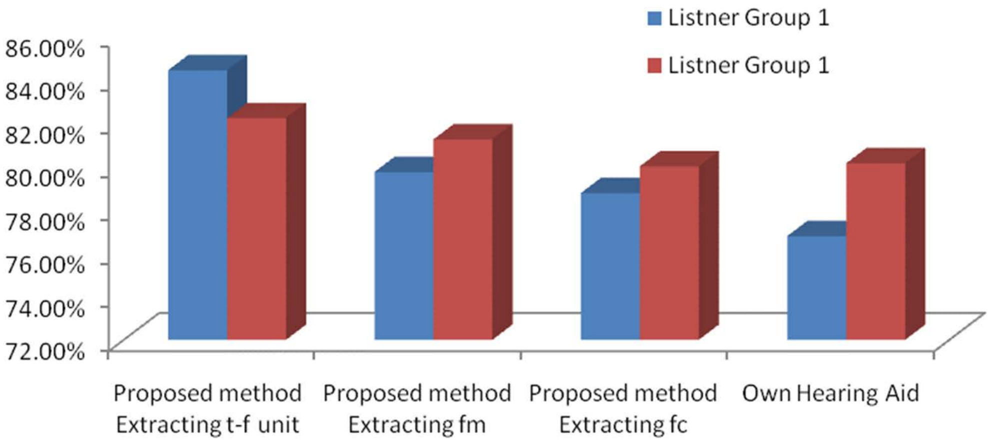

FIGURE 8. Average recognition score for the Marathi vowels.

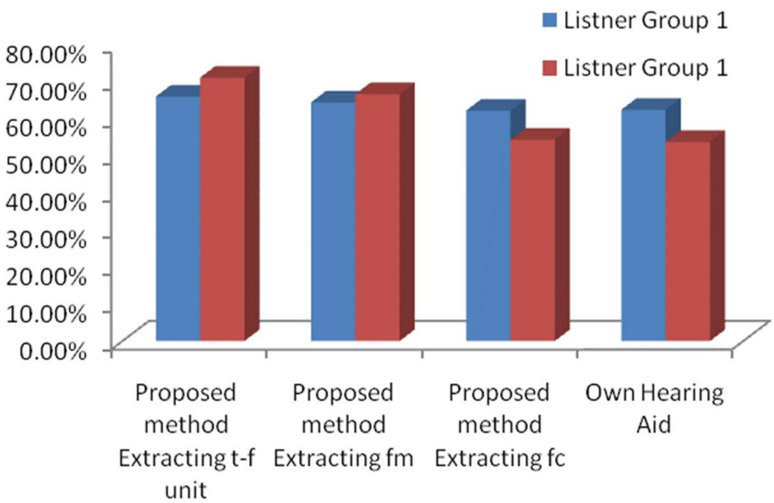

FIGURE 9. Average recognition score for the Marathi consonants.

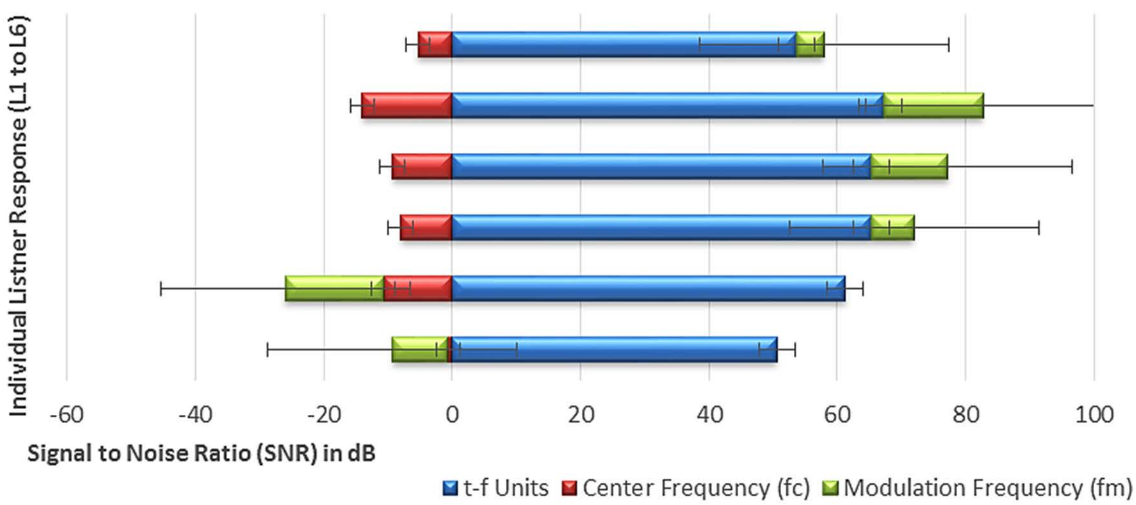

FIGURE 10. Signal to Noise Ratio (SNR) variation comparison for the proposed technique.

and modulation frequency (fm) reduces processed speech SNR and demands more insertion gain to meet patient's requirement.

Table 6 presents the spoken and listened confusion matrices for the Marathi consonants. The confusion matrix diagonals specified the correct identification of each consonant.

TABLE 5. Feature vector extraction and processed speech SNR correlation.

Processed speech SNR in dB	L1	L2	L3	L4	L5	L6
t-f Units	50.6	61.3	65.3	65.3	67.3	53.6
Center Frequency (fc)	-0.6	-10.7	-8	-9.3	-14	-5.3
Modulation Frequency (fm)	-0.87	-15.3	6.7	12	15.5	4.4

TABLE 6. Spoken and listened confusion matrix test for marathi consonants.

Spoken Alphabet (Column) Vs. Recognized Alphabets by HA user (Row)												
	क	ख	ग	घ	ज	झ	द	ध	ब	भ	श	स
क	12	7	4	2	0	1	0	2	2	3	2	0
ख	3	5	3	7	4	2	1	0	2	3	4	1
ग	4	4	6	2	3	4	5	0	1	2	3	1
घ	0	2	5	8	2	1	1	0	4	7	4	1
ज	3	1	7	1	9	6	2	0	2	3	1	0
झ	4	5	0	2	0	6	0	4	8	0	1	5
द	5	4	2	5	0	6	4	0	3	1	2	3
ध	0	0	3	7	0	5	7	11	0	1	1	0
ब	2	2	1	0	5	1	8	6	8	0	0	2
भ	0	2	0	0	6	0	3	6	0	9	4	5
श	2	0	4	0	4	0	2	2	2	3	8	8
स	0	3	0	1	2	3	2	4	3	3	5	9

The Marathi confusing consonants group was responsible for the reduced recognition rate.

VI. CONCLUSION

The proposed speech enhancement based on I-AMS processing was designed to improve hearing precision for the hearing-disabled under different speaker and listener background conditions. It makes several contributions to signal denoising (enhancement), insertion gains at different frequency levels, and feature extraction, training, and testing of neuro-fuzzy classifiers.

Current signal processing techniques in hearing aids process speech signals regardless of the speech background, which may maintain the SPL below the hearing threshold level. In the proposed technique, the minimum insertion gain (IGSPxx) is added according to the speech background to satisfy the hearing aid user requirements. The proposed method achieves an SNR of 25 dB in contrast to the existing technique, providing 23 dB in quiet room conditions, which is similar to a noisy background SNR of 13 dB as compared to the existing 11 dB. The recognition results showed the importance of denoising; the short sentence correct recognition rate increased from 47.33% to 48.60%. Denoising has a greater impact on recognition results in a noisy background than in a quiet background situation. The performance of the neuro-fuzzy classifier varies according to the training and testing rates. Sensitivity variation was found in the range 98.44%- 99.46%, 97.61-100 % and 94.44%-98.40% respectively, after extracting t-f unit, centre frequency, and modulation frequency. The classification accuracy ranged from 74.86% to 86.66% for 80-20% training and testing conditions. Finally, the t-f unit and training testing rate played a vital role in improving classifier performance. Speech enhancement with t-f unit extraction had a positive impact on short sentences recognition, and 66.70% accuracy was obtained for the t-f unit.

In addition, the proposed AMS-based classification method showed a significant improvement for female speakers when compared to male speakers. This research can be extended by extracting additional speech features and incorporating appropriate modifications during the training, classification, and testing phases. Additionally, implementations on different hardware platforms, such as complex programmable logic devices (CPLD) and field-programmable gate arrays (FPGA), are envisaged.

ACKNOWLEDGMENT
This research work was demonstrated and verified at the Priyadarshini Deaf Residential School, Shirpur, Dhule, India. Ethical approval was obtained from the relevant ethics committee of the Institutional Oversight Board. The researchers granted permission to exploit the experimental results and outcomes. The institute has provided written permission to publish research outcomes in conferences and journals.

REFERENCES
[1] R. Liang, J. Xi, J. Zhou, C. Zou, and L. Zhao, "An improved method to enhance high-frequency speech intelligibility in noise," *Appl. Acoust.*, vol. 74, no. 1, pp. 71–78, Jan. 2013.
[2] E. Alexandre, L. Cuadra, M. Rosa, and F. Lopez-Ferreras, "Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 8, pp. 2249–2256, Nov. 2007.
[3] J. M. Alexander, "Nonlinear frequency compression: Balancing start frequency and compression ratio," *Amer. Auditory Soc.*, vol. 12, no. 3, pp. 349–362, 2012.
[4] J. Zhou, R. Liang, L. Zhao, and C. Zou, "Whisper intelligibility enhancement using a supervised learning approach," *Circuits, Syst., Signal Process.*, vol. 31, no. 6, pp. 2061–2074, Dec. 2012.
[5] A. Nurettin, O. Ozdamar, and G. Cuneyt, "Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection," *Eng. Appl. Artif. Intell.*, vol. 19, pp. 209–218, Mar. 2006.
[6] T. Goossens, C. Vercammen, J. Wouters, and A. van Wieringen, "Masked speech perception across the adult lifespan: Impact of age and hearing impairment," *Hearing Res.*, vol. 344, pp. 109–124, Feb. 2017.
[7] X. Xiao, H. Zhang, G. Hu, C. Liu, and J. Liu, "Evaluation of frequency-lowering algorithms for intelligibility of Chinese speech in hearing-aid users," *Prog. Natural Sci.*, vol. 19, no. 6, pp. 741–749, Jun. 2009.
[8] S. Nasir, M. I. Khatik, G. Witjaksono, and G. Ahmad, "Variance based time-frequency mask estimation for unsupervised speech enhancement," *Multimedia Tools Appl.*, vol. 20, no. 12, pp. 3389–3398, 2019.
[9] B. Singh, N. Kapur, and P. Kaur, "Speech recognition with hidden Markov model: A review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 2, no. 3, pp. 1–4, 2012.
[10] T. Kitazoe, S.-I. Kim, and T. Ichiki, "Speech recognition using a stereo vision neural network model," *Artif. Life Robot.*, vol. 4, no. 1, pp. 37–41, Mar. 2000.
[11] M. A. A. El-Fattah, M. I. Dessouky, S. M. Diab, and F. E. A. El-sami, "Adaptive Wiener filtering approach for speech enhancement," *Ubiquitous Comput. Commun. J.*, vol. 3, no. 2, pp. 23–31, 2008.
[12] N. Kaladharan, "Speech enhancement by spectral subtraction method," *Int. J. Comput. Appl.*, vol. 96, no. 13, pp. 76–85, 2014.
[13] K. Ramalakshmi and R. N. D. Kumar, "Speech enhancement with signal subspace filter based on perceptual post filtering," *Int. J. Latest Trends Eng. Technol.*, vol. 2, no. 1, pp. 792–801, 2013.
[14] 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," *IEEE Access*, vol. 7, pp. 103485–103504, 2019, doi: 10.1109/ACCESS.2019.2929864.
[15] J. Padrell-Sendra, M.-I. Dario, and D. M. Fernando, "Support vector machines for continuous speech recognition," in *Proc. 14th Eur. Signal Process. Conf.*, 2006, pp. 1–4.
[16] Istikmal, A. Kurniawan, and Hendrawan, "Throughput performance of transmission control protocols on multipath fading environment in mobile ad-hoc network," in *Proc. 11th Int. Conf. Telecommun. Syst. Services Appl. (TSSA)*, Oct. 2017, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8272939/authors#authors
[17] S. Preethi and P. Aishwarya, "Combining wavelet texture features and deep neural network for tumor detection and segmentation over MRI," *J. Intell. Syst.*, vol. 28, no. 4, pp. 571–588, Sep. 2019.
[18] T. Y. C. Ching, H. Dillon, and D. Byrne, "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," *J. Acoust. Soc. Amer.*, vol. 103, no. 2, pp. 1128–1140, Feb. 1998.
[19] A. Garg and O. P. Sahu, "Cuckoo search based optimal mask generation for noise suppression and enhancement of speech signal," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 21, no. 3, pp. 24–31, 2015.
[20] J. Sang, H. Hu, C. Zheng, G. Li, M. E. Lutman, and S. Bleeck, "Speech quality evaluation of a sparse coding shrinkage noise reduction algorithm with normal hearing and hearing impaired listeners," *Hearing Res.*, vol. 327, pp. 175–185, Sep. 2015.
[21] R. Paluch, M. Krueger, M. Hendrikse, G. Grimm, V. Hohmann, and M. Meis, "Ethnographic research: The interrelation of spatial awareness, everyday life, laboratory environments, and effects of hearing aids," in *Proc. 6th Int. Symp. Auditory Audiol. Res.*, 2018, pp. 39–46.
[22] Y. K. Han and K. Lee, "A study on the Korean conversation speech level and spectrum in sound-treated environment," *Audiol. Speech Res.*, vol. 16, no. 2, pp. 133–139, 2018.
[23] M. W. Skinner, "Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation," *J. Acoust. Soc. Amer.*, vol. 67, no. 1, pp. 306–317, Jan. 1980.
[24] L. Ruiyu, L. Zhao, X. Zhang, J. Xi, and Q. Wang, "Acoustic source localization based on compressed sensing and auditory bionics for hearing aids," *Chin. J. Sci. Instrum.*, vol. 6, pp. 1390–1395, Jun. 2011.
[25] V. Anoop, P. V. Rao, and M. Nidhin, "Performance analysis of speech enhancement methods using adaptive algorithms and optimization techniques," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2015, pp. 732–741.
[26] T. Toda, M. Nakagiri, and K. Shikano, "Statistical voice conversion techniques for body-conducted unvoiced speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 9, pp. 2505–2517, Nov. 2012.
[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Jul. 2017, pp. 2261–2269.
[28] M. Milczynski, J. E. Chang, J. Wouters, and A. van Wieringen, "Perception of Mandarin Chinese with cochlear implants using enhanced temporal pitch cues," *Hearing Res.*, vol. 285, nos. 1–2, pp. 1–12, Mar. 2012.
[29] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *Proc. ICASSP*, Mar. 2016, pp. 31–35.
[30] D. Glista, S. Scollie, M. Bagatto, R. Seewald, V. Parsa, and A. Johnson, "Evaluation of nonlinear frequency compression: Clinical outcomes," *Int. J. Audiol.*, vol. 48, no. 9, pp. 632–644, Jan. 2009.
[31] K. L. Helm, "The effect of linear frequency compression and linear frequency transposition on speech perception in school-aged children," *J. Audiol. Commun. Sci.*, vol. 13, no. 9, pp. 289–305, 2011.
[32] D.-L. Hsieh, W.-T. Shih, and T.-C. Liu, "Extended bandwidth nonlinear frequency compression in Mandarin-speaking hearing-aid users," *J. Formosan Med. Assoc.*, vol. 117, no. 2, pp. 109–116, 2018.
[33] J. M. Alexander, "Nonlinear frequency compression: Balancing start frequency and compression ratio," in *Proc. 39th Annu. Meeting Amer. Auditory Soc.*, 2016, vol. 139, no. 2, pp. 938–957.
[34] J. Gou, J. Smith, J. Valero, and I. Rubio, "The effect of frequency transposition on speech perception in adolescents and young adults with profound hearing loss," *Deafness Educ. Int.*, vol. 13, no. 1, pp. 17–33, Mar. 2011.
[35] T. C. Parent and R. Chmiel, "Comparison of performance with frequency transposition hearing aids and conventional hearing aids," *J. Amer. Acad. Audiol.*, vol. 9, no. 1, pp. 67–77, 1998.
[36] F. Kuk, D. Keenan, J. Auriemmo, P. Korhonen, H. Peeters, C. Lau, and B. Crose, "Interpreting the efficacy of frequency-lowering algorithms," *Hearing J.*, vol. 63, no. 4, pp. 30–40, 2010.
[37] G. Deepika and M. Vidya, "Background noise reduction using FFBPNNLM network and adaptive filter," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 3, pp. 107–117, Apr. 2017.
[38] M. Kleinschmidt and J. Tcharz, "Noise suppression based on neurophysiological-motivated SNR estimation for robust speech processing," *J. Theor. Appl. Inf. Technol.*, vol. 67, no. 3, pp. 664–674, 2014.
[39] A. Pandey and V. J. Mathews, "Low-delay signal processing for digital hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 4, pp. 699–710, May 2011.

[40] L. Zamaninezhad, H. Volker, A. Buchner, M. R. Schadler, and T. Jurgens, "A physiologically-inspired model reproducing the speech intelligibility benefit in cochlear implant listeners with residual acoustic hearing," *Hearing Res.*, vol. 65, no. 3, pp. 228–236, 2016.

[41] M. Keshavarzi, T. Goehring, R. E. Turner, and B. C. J. Moore, "Comparison of effects on subjective intelligibility and quality of speech in babble for two algorithms: A deep recurrent neural network and spectral subtraction," *J. Acoust. Soc. Amer.*, vol. 145, no. 3, pp. 1493–1503, Mar. 2019.

[42] E. W. Healy, S. E. Yoho, Y. Wang, F. Apoux, and D. L. Wang, "Speech-cue transmission by an algorithm to increase consonant recognition in noise for hearing-impaired listeners," *J. Acoust. Soc. Amer.*, vol. 136, no. 6, pp. 3325–3336, 2019.

[43] W. Jesteadt, D. L. Valente, S. N. Joshi, and K. K. Schmid, "Perceptual weights for loudness judgments of six-tone complexes," *J. Acoust. Soc. Amer.*, vol. 136, no. 2, pp. 728–735, Aug. 2014.

[44] X. Xianbo, Z. Hui, H. Guangshu, L. Chunhong, and L. Jia, "Evaluation of frequency-lowering algorithms for intelligibility of Chinese speech in hearing-aid users," *Prog. Natural Sci.*, vol. 104, no. 1, pp. 432–441, 2009.

[45] L. C. E. Veugen, M. M. E. Hendrikse, M. M. van Wanrooij, M. J. H. Agterberg, J. Chalupper, L. H. M. Mens, A. F. M. Snik, and A. J. van Opstal, "Horizontal sound localization in cochlear implant users with a contralateral hearing aid," *Hearing Res.*, vol. 336, pp. 72–82, Jun. 2016.

[46] T. Goehring, B. Federico, J. J. M. Monaghan, B. van Dijk, A. Zarowski, and S. Bleeck, "Speech enhancement based on 27 neural networks improves speech intelligibility in noise for cochlear implant users," *Hearing Res.*, vol. 344, pp. 183–194, Feb. 2017.

[47] S. Gaikwad, B. Gawali, and S. C. Mehrotra, "Novel approach based feature extraction for Marathi continuous speech recognition," *Proc. Int. Conf. Adv. Comput., Commun., Informat. (ICACCI)*, vol. 12, 2012, pp. 26–35.

[48] A. R. Kavitha and C. Chellamuthu, "Detection of brain tumour from MRI image using modified region growing and neural network," *Imag. Sci. J.*, vol. 61, no. 7, pp. 556–567, Sep. 2013.

[49] M. Finke, A. Büchner, E. Ruigendijk, M. Meyer, and P. Sandmann, "On the relationship between auditory cognition and speech intelligibility in cochlear implant users: An ERP study," *Neuropsychologia*, vol. 87, pp. 169–181, Jul. 2016.

[50] G. Kim and P. C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 2080–2090, Nov. 2010.

[51] P. Padmavathy and S. P. Mohideen, "An efficient two-pass classifier system for patient opinion mining to analyze drugs satisfaction," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101755.

[52] I. Hadhami and B. Aicha, "Speech signal enhancement using empirical mode decomposition and adaptive method based on the signal for noise ratio objective evaluation," *Int. Rev. Comput. Softw.*, vol. 9, no. 8, p. 1461, Aug. 2014.

[53] G. Kim and P. C. Loizou, "Improving speech intelligibility in noise using a binary mask that is based on magnitude spectrum constraints," *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 1010–1013, Dec. 2010.

[54] K. Muthuvel, S. Anto, and T. J. Alexander, "GABC based Neuro-fuzzy classifier with hybrid features for ECG beat classification," *Multimedia Tools Appl.*, vol. 78, no. 24, pp. 35351–35372, Sep. 2019.

[55] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016, pp. 47–54.

[56] Y. Liu and D. Wang, "Divide and conquer: A deep CASA approach to talker-independent monaural speaker separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 12, pp. 2092–2102, Dec. 2019.

[57] D. Wang, "Deep learning reinvents the hearing aid," *IEEE Spectr.*, vol. 54, no. 3, pp. 32–37, Mar. 2017.

[58] J. J. M. Monaghan, T. Goehring, X. Yang, F. Bolner, S. Wang, M. C. M. Wright, and S. Bleeck, "Auditory inspired machine learning techniques can improve speech intelligibility and quality for hearing-impaired listeners," *J. Acoust. Soc. Amer.*, vol. 141, no. 3, pp. 1985–1998, 2017.

[59] N. Tye-Murray, B. Spehar, M. Sommers, and J. Barcroft, "Auditory training with frequent communication partners," *J. Speech, Lang., Hearing Res.*, vol. 59, no. 4, pp. 871–875, Aug. 2016.

[60] D. Yu, M. Kolbaek, Z.-H. Tan, and J. Jensen, "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," in *Proc. ICASSP*, Mar. 2017, pp. 241–245.

[61] Y. Zhao, D. L. Wang, and E. M. Johnson, "A deep learning based segregation algorithm to increase speech intelligibility for hearing-impaired listeners in reverberant-noisy conditions," *J. Acoust. Soc. Amer.*, vol. 144, no. 3, pp. 1627–1637, 2018.

[62] L. Eipert, A. Selle, and G. M. Klump, "Uncertainty in location, level and fundamental frequency results in informational masking in a vowel discrimination task for young and elderly subjects," *Hearing Res.*, vol. 377, pp. 142–152, Jun. 2019.

[63] Z. Zhang, Y. Shen, and D. S. Williamson, "Objective comparison of speech enhancement algorithms with hearing loss simulation," in *Proc. ICASSP*, May 2019, pp. 6845–6849.

[64] E. Roverud, V. Best, C. R. Mason, and G. Kidd, "Selective and divided listening in normal-hearing and hearing-impaired listeners measured in a nonspeech pattern identification task," in *Proc. Meetings Acoust.*, vol. 23, 2015, pp. 50–62.

[65] S. A. Nossier, M. R. M. Rizk, and N. D. Moussa, "Enhanced smart hearing aid using deep neural networks," *Alexandria Eng. J.*, vol. 58, no. 2, pp. 539–550, 2019.

[66] A. M. C. Martinez, L. Gerlach, G. Payá-Vayá, H. Hermansky, J. Ooster, and B. T. Meyer, "DNN-based performance measures for predicting error rates in automatic speech recognition and optimizing hearing aid parameters," *Speech Commun.*, vol. 106, pp. 44–56, Jan. 2019.

[67] V. Best and J. Swaminathan, "Revisiting the detection of interaural time differences in listeners with hearing loss," *J. Acoust. Soc. Amer.*, vol. 145, no. 6, pp. 508–513, 2019.

[68] L. Varnet, C. Langlet, C. Lorenzi, D. S. Lazard, and C. Micheyl, "High-frequency sensorineural hearing loss alters cue-weighting strategies for discriminating stop consonants in noise," *Trends Hearing*, vol. 23, pp. 1–18, Nov. 2019.

[69] K. H. Kim and J. H. Lee, "Evaluation of the Korean matrix sentence test: Verification of the list equivalence and the effect of word position," *Audiol. Speech Res.*, vol. 14, no. 2, pp. 100–107, Apr. 2018.

[70] J. Cubick, J. M. Buchholz, V. Best, M. Lavandier, and T. Dau, "Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners," *J. Acoust. Soc. Amer.*, vol. 144, no. 5, pp. 2896–2905, Nov. 2018.

[71] S. M. K. Madsen, K. L. Whiteford, and A. J. Oxenham, "Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds," *Sci. Rep.*, vol. 7, no. 1, pp. 126–134, Dec. 2017.

[72] M. R. Wirtzfeld, N. Pourmand, V. Parsa, and I. C. Bruce, "Predicting the quality of enhanced wideband speech with a cochlear model," *J. Acoust. Soc. Amer.*, vol. 142, no. 3, pp. 319–325, 2017.

[73] A. Josupeit, E. Schoenmaker, S. van de Par, and V. Hohmann, "Sparse periodicity-based auditory features explain human performance in a spatial multitalker auditory scene analysis task," *Eur. J. Neurosci.*, vol. 51, no. 5, pp. 1–11, 2019.

[74] J. I. Marin-Hurtado, D. N. Parikh, and D. V. Anderson, "Perceptually inspired noise-reduction method for binaural hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 4, pp. 1372–1387, May 2012.

[75] S. Kortlang, M. Mauermann, and S. D. Ewert, "Suprathreshold auditory processing deficits in noise: Effects of hearing loss and age," *Hearing Res.*, vol. 331, pp. 27–40, Jan. 2016.

[76] A. U. Rajendra, O. S. Lih, H. Yuki, J. H. Tan, A. Muhammad, G. Arkadiusz, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, no. 1, pp. 389–396, May 2017.

[77] P. F. K. Rahiman, V. S. Jayanthi, and A. N. Jayanthi, "Deep convolutional neural network-based speech enhancement to improve speech intelligibility and quality for hearing-impaired listeners," *Med. Biol. Eng. Comput.*, vol. 57, no. 4, pp. 757–759, 2017.

[78] D. S. Brungart, J. I. Cohen, D. Zion, and G. Romigh, "The localization of non-individualized virtual sounds by hearing impaired listeners," *J. Acoust. Soc. Amer.*, vol. 141, no. 4, pp. 2870–2881, Apr. 2017.

[79] T. R. McRackan, J. B. Ahlstrom, W. B. Clinkscales, T. A. Meyer, and J. R. Dubno, "Clinical implications of word recognition differences in earphone and aided conditions," *Otol. Neurol.*, vol. 37, no. 10, pp. 1475–1481, 2016.

PRASHANT G. PATIL received the B.E. degree from North Maharashtra University, Jalgaon, in 2004, the M.E. degree from the RKDF Institute of Science and Technology, in 2011, and the Ph.D. degree from Rashtrasant Tukdoji Maharaj Nagpur University, in 2020. He is currently working as an Associate Professor with the Department of Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur, Maharashtra, India. He has more than 19 years of teaching experience. His current research interests include biomedical signal and speech processing, natural language processing, and image processing.

TUSHAR H. JAWARE received the bachelor's degree in electronics and telecommunication engineering from North Maharashtra University, Jalgaon, and the master's degree in digital electronics and the Ph.D. degree in medical image processing from Sant Gadge Baba Amravati University, Amravati. He is currently working as the Dean of research and development with the R. C. Patel Institute of Technology, Shirpur, Maharashtra, India. He has more than 17 years of teaching experience. He is a recognized Ph.D. Supervisor in electronics engineering at North Maharashtra University, Jalgaon, Dr. Babasaheb Ambedkar Technological University, Lonere. He is working as a Board Member of Studies in electronics and telecommunication engineering at North Maharashtra University, Jalgaon. He received the Loksatta Tarun Tejankit Award, in 2019, the GIS Young Innovator and Researchers Award (Central India), in 2018, from JSR Laboratory, Pune, in Association with the Asian Society for Scientific Research and "Bright Researcher Award, in 2017," from the International Institute of Organized Research. He received 12 awards for research and academic work from different societies.

SHEETAL P. PATIL received the B.E. degree from North Maharashtra University, Jalgaon, in 2003, and the M.Tech. degree from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, in 2011. She is currently pursuing the Ph.D. degree from North Maharashtra University, Jalgaon. She is currently working as an Assistant Professor with the Department of Computer Engineering, R. C. Patel Institute of Technology, Shirpur, Maharashtra, India. She has more than 14 years of teaching experience. Her current research interests include image processing, machine learning, and deep learning.

RAVINDRA D. BADGUJAR received the B.E. degree from North Maharashtra University, Jalgaon, in 2005, the M.E. degree from Shri Sant Gajanan Maharaj College of Engineering, Shegaon, in 2010, and the Ph.D. degree from KBC, North Maharashtra University, Jalgaon, in 2021. He has more than 17 years of teaching experience. He is currently working as an Associate Professor with the Department of Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur, Maharashtra, India. His current research interests include computer vision, biomedical image processing, microcontroller, and embedded systems.

FELIX ALBU (Senior Member, IEEE) received the M.S. degree in applied electronics and the Ph.D. degree in telecommunications from the University Politehnica of Bucharest, Romania, in 1993 and May 1999, respectively, and the Dr.Habil. degree in electronics and telecommunications, in 2014. He worked as a Teaching Assistant at the University Politehnica of Bucharest, from 1993 to 1999. During his Ph.D. study, he has been a Visiting Researcher for about two years at the National Institute of Telecommunications, Evry, France, and LAAS-CNRS, Toulouse, France. He has obtained extensive research experience as a Postdoctoral Researcher at University College Dublin, Ireland, from1999 to 2002, and the Aristotle University of Thessaloniki, Greece, from2004 to 2005. Also, he has got industrial research experience at Lake Communications, Dublin, from 2002 to 2003, and at Fotonation Romania, from 2006 to 2011. He is currently a Full Professor with the Valahia University of Targoviste, Romania. His research interests include various signal processing and machine learning areas. Since 2013, he has been a Senior Member of IEEE SPS.

IBRAHIM MAHARIQ received the B.Sc. degree from the Department of Electrical and Computer Engineering, Palestine Polytechnic University, Hebron, Palestine, in 2003, the M.Sc. and Ph.D. degrees from the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey, in 2009 and 2014, respectively, and the Ph.D. degree from the Department of Electrical and Electronics Engineering, TOBB Economic and Technology University, Ankara, in 2017. In 2018, he was promoted as an Associate Professor. He currently serves at the American University of the Middle East, Kuwait. He has over 85 journal publications on engineering and computational methods.

BAHAA AL-SHEIKH (Senior Member, IEEE) received the B.Sc. degree in electronics engineering from Yarmouk University, Jordan, the M.Sc. degree in electrical engineering from Colorado State University, CO, USA, and the Ph.D. degree in biomedical engineering from the University of Denver, CO, in 2000, 2005, and 2009, respectively. From 2009 to 2015, he worked at Yarmouk University as an Assistant Professor with the Department of Biomedical Systems and Medical Informatics Engineering and served as the Department Chairman, from 2010 to 2012. He served as a part-time Consultant at Sand-Hill Scientific Inc., Highlands Ranch, CO, in the biomedical signal-processing field, from 2009 to 2014. He served as an Associate Professor with the Department of Electrical Engineering, American University of the Middle East, Kuwait, from 2015 to 2022. He currently working as an Associate Professor with the Department of Biomedical Systems and Informatics Engineering, Yarmouk University, Jordan. His research interests include digital signal and image processing, biomedical system modeling, medical instrumentation, and sound-source localization systems.

CHITTARANJAN NAYAK received the Ph.D. degree in engineering from the National Institute of Technology, Agartala, India, in 2017. He is currently working as an Associate Professor with the Department of Communication Engineering, Vellore Institute of Technology, Vellore, India. His current research interests include soft computing, 1-D photonic multilayers, and the formation of photonic nanojets for different optoelectronic applications.