STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **14**, September 2025, pp 1481–1511.
Published online in International Academic Press (www.IAPress.org)

An Optimized Hybrid Approach for Reducing Computational Overheads and Evaluating Audio Signal Characteristics in Wireless Acoustic Sensor Networks

Utpal Ghosh [1,2], Uttam Kr. Mondal [1,*], Abdelmoty M. Ahmed [3], Ahmed A. Elngar [4,*]

[1]*Department of Computer Science, Vidyasagar University, Midnapore, 721102, West Bengal, India*
[2]*Dept. of Computer Science, Sarojini Naidu College for Women, Kolkata, 700028, West Bengal, India*
[3]*Computer Science Dept., Faculty of Information and Technology, Ajloun National University, P. O. Box 43, Ajloun 26810, Jordan*
[4]*Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, 62521, Egypt*

Abstract This paper presents a hybrid system designed to analyze multiple properties of audio signals while minimizing quality losses during transmission over Wireless Acoustic Sensor Networks (WASNs). The proposed system operates in two phases. In the first phase, audio signal quality is evaluated using key parameters such as packet loss ratio, signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR) and signal fidelity. The experimental results of proposed method reveal an increased packet loss ratio, reduced PSNR and lower signal fidelity degraded audio quality. An acceptable threshold is established to maintain quality, though network traffic exceeding this threshold negatively impacts performance. To address this, the system incorporates controls for packet loss, SNR, PSNR and fidelity, ensuring the transmitted audio maintains parity with the source. A WASN framework is introduced for distributed and efficient audio property analysis in the second phase. The framework employs feature extraction techniques, including Mel Frequency Cepstral Coefficient (MFCC) and Power Normalized Cepstral Coefficient (PNCC), alongside other existing methods, to extract comprehensive features from audio signals. Combining quality assessment and distributed analysis, this hybrid system provides a robust solution for enhancing audio signal processing within dynamic and resource-constrained network environments.

Keywords WASNs, Packet Loss Ratio, SNR, PSNR, MFCC, PNCC, Signal Fidelity

DOI: 10.19139/soic-2310-5070-2475

1. Introduction

Audio streaming is very popular nowadays for online radio stations those broadcast music and give access to various stations [1],[2]. The user wants to transmit high-quality audio from cloud sites or from other sources to destinations, and the organizations are also trying to fulfil these services according to the user's satisfaction level. Packet loss and packet delay decrease the quality of the audio signal at end stations [3],[4]. Several organizations have taken some measures to upgrade the quality of the audio signal by pretending the losses [5],[6],[7]. The user's perception, rating of the performance, and product observation are conducted by quality of experience (QoE) [8],[9]. Quality of experience is the blueprint of subjective and objective qualities those are needed when humans doing interact with technology [10]. The quality of experience is divided into two categories, one is subjective and the other is objective [11],[12]. Any organization is not able to know the acceptable level of signal fidelity, SNR, packet loss, PSNR, etc. in audio signal streaming. The acceptable level or threshold value for packet loss, PSNR, SNR, and signal fidelity has been analyzed. Different researchers performed their research on developing various

*Correspondence to: Uttam Kr. Mondal (Email: uttam_ku_82@yahoo.co.in). Department of Computer Science, Vidyasagar University, Midnapore, West Bengal, India (721102); Ahmed A. Elngar (Email:elngar_7@yahoo.co.uk).Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, Egypt (62521)

ISSN 2310-5070 (online) ISSN 2311-004X (print)

reliable systems that have the ability to sense audio or acoustic signals.We know that most of communications are done through speech or a vocal process, naturally people intended to communicate with computers in the same way [13]. Audio is capable of providing interaction between people and computers and this is the reason that people gain interest in developing such types of computers or devices that are able to recognize people's voices. The effort to recognize speech or voice globally and also achieve more computational power without consuming a huge amount of power could result in the development of suitable and user friendly applications for speech recognition [14]. There are different audio, or more specifically, speech recognition systems available that are language-specific [15]. The Research work based on the acoustic interaction between humans and computers tries to develop the system more accurately and also tries to achieve correctness [16].

1.1. Motivation

Wireless Acoustic Sensor Networks (WASNs) are increasingly being adopted as an effective tool for transmitting and processing real-time audio signals. Among various applications include environment monitoring, surveillance, and voice-controlled systems. In contrast, these networks have considerable challenges relating to transmitting high quality audio, mainly due to packet losses, noise interference and bandwidth restrictions. No existing QoE model stipulates reliable thresholds for the evaluation and control of quality-reduction metrics such as Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), packet loss ratio and signal fidelity. Usually, one traditional audio extraction feature method is employed for example, MFCC or PNCC, thus limiting the system's overall ability to fully grasp all audio characteristics. Hence an urgent demand exists to develop a generic, hybrid and optimized solution that will reduce the computational overhead and degradation of audio in wireless transmission while enhancing feature extraction accuracy through multiple audio descriptors, thereby supporting more robust classification and analysis in real-time environments.

1.2. Key Contributions

1.2.1. Two-Phase Hybrid WASN Framework: Here it is one system with two phases:

- Phase 1 specializes in quantitative evaluation and dynamic adjustment of audio signal quality based on considerations of parameters such as packet loss ratio, SNR, PSNR and signal fidelity. Hence, a threshold-based approach is used to consider and control the audio degradation during transmission.
- Phase 2 focuses on feature extraction using both MFCC and PNCC algorithms while SVM-based classification aids in further refining the identification of any audio signal characteristics.

1.2.2. Intermediate Station-Based Channel Allocation Algorithm: This algorithm is designed to install intermediate stations within the WASN, thereby increasing availability of free transmission channels. This reduces packet loss (especially voice signals) by dynamically scanning and hence dynamically allocating available channels for an uninterrupted and quality-preserving transmission.

1.2.3. Signal Reconstruction and Error Compensation: At the receiver end, the system employs adaptive filtering and error compensation schemes to reconstruct the transmitted audio signals, thereby providing a lossless or near-lossless output signal that retains fidelity to the original source.

1.2.4. Integration of Feature Extraction with DCT Sharing: The sharing of a common DCT block between MFCC and PNCC is considered such that computational redundancy is minimized, leading to energy savings and efficiency improvement in feature extraction within WASN.

1.2.5. Performance Evaluation and Comparison: The exhaustive simulation carried on MATLAB compares the proposed hybrid system with the traditional systems (i.e., DCT, STFT, etc.) by showing the improvement in:

- Energy consumption (being less by 12.5%)
- Packet Delivery Ratio (being more by 5.43%)

- Packet Loss Ratio (being less by 5.43%)
- End-to-End Delay (being less by 12%)

1.2.6. Novelty-Contributed in Multi-Parameter Quality Analysis: Unlike most previous works that concentrate either on one or two parameters, an elaborate multi-metric analysis and rectification are performed here to guarantee the ensured transmission and property analysis of audio signals among different WASN settings.

This research article is categorized into five sections. Section 2 describes the literature review. Section 3 provides the research methodology, Section 4 and 5 demonstrates the results and discussion respectively and lastly, Section 6 ends with the conclusion and future scope of the present technique.

2. Literature Review

The authors of [17] provides some assessment on subjective and objective quality of experience of audio quality and evaluation of audio in three-dimensional signal. The article of [18] presented a framework for measuring quality of experience, that predicts the quality of audio signal. This technique also demonstrates the downside of existing objective quality methods, as well as discussed how those drawbacks have been overcome. The article [19] investigated the effect of sound human being in the form of biological and physiological. Various types of application like body vibration, vibro-acoustic and focal vibration etc. are used for performing several mechanism of tests. The authors in [20] showed how sound has been implemented to real-world application. It also leads to the simulation activity of the new and other existing technologies. It addresses the issues for commercial noise surveillance like criteria of noise, risk of damage hearing, noise-assessment measurements, measuring apparatus and various sound-source types which incorporates the computation and assessment of the output. This paper [21] depicted a VOIP based QoE based scheduling algorithm for downlink, that has the ability to increase the involvement of the number of end users based on each cell in VOIP technology. This technique shows that the algorithm improves downlink scheduling and also enhances the size of cells by 75%. In article [22], the authors compute the quality of experience of VOIP applications like Google Talk, MSN Messenger, and Skype by applying the buffer playout dimension algorithm. A methodology is proposed by the authors of [23] that increases the quality of audio or video transmission on IP by managing the packet loss in video with framing error. In article [24], the authors described a technique that is used to measure the bandwidth of the receiver to produce a high quality audio signal in a multi-user method. This method calculates the available bandwidth of the network, which depends on the status of the buffer and throughout segments.

As technologies are developed day by day, various devices are enclosed to the audio recognition machine to gain a better outcome, like applications for voice recognition in mobile phones [25], automobiles [26], retrieving information and controlling access to audio in a centralised manner [27], emotion detection devices [28], some audio monitoring systems [29], and various voice assistance devices [30]. Researchers in [31] demonstrated a novel technique for analysing of emission based audio signals those are originated through phase resolved partial discharge. It represents the comparison analysis based on different measuring approaches, also various laboratory simulation based experiments have been configured. In article [32], the authors represented the analysis of spectrogram in the form of voice signal, based on the Cohen class technique. This technique also describes the Wigner-Ville'a distribution, the affine class of distribution and the reassigned distribution. In article [33], a vibro-acoustic identification method based on blind source separation approach was presented in order to advanced transformer monitoring technology. At first, the ordered local values of potential function have been employed for the extraction of vibro-acoustic signal. After that, the eigenvector representation of the signals are analysed to determine the operational state of transformer. This technique also depicts the simulation scenarios such as operations, amplitude increment, change in the components of the transformer. In a real world environment, audio signals consist of unusual noise. To improve the efficiency of the audio signal, have to reduce or remove noise from the audio. In article [34],[35],[36], [37], the researcher presents different methods to decrease noise from audio. When the audio channels vary, the efficiency of the audio recognition device degrades [38]. Most researchers

used MFCC and PNCC for the efficient extraction of features of audio signals [39],[40],[41],[42]. The article [43] proposed a technique to manage the desirability of support vector machines to achieve the best decision and classification, as well as using the concept of supervised learning. In paper [44], the researchers proposed several techniques, where using the classification processes of SVM various jobs are performed on an audio signal. Figure 1 presents the SmartArt chart view of literature review.

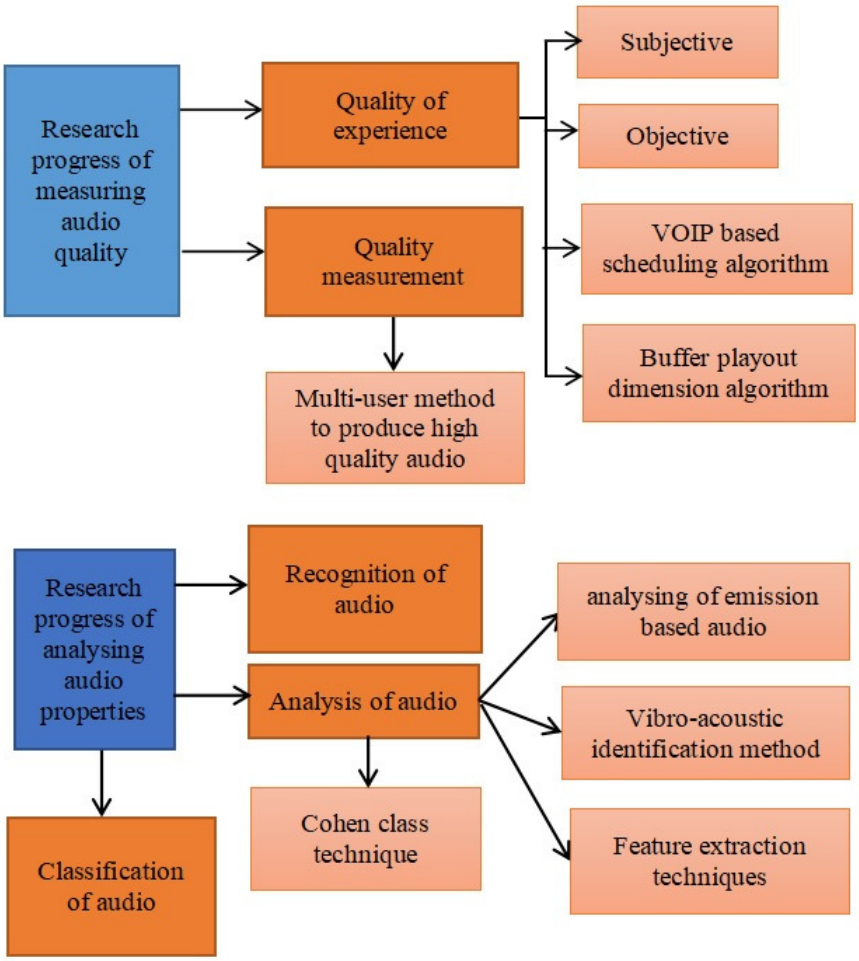

Figure 1. Workflow diagram of channel allocation algorithm

3. Proposed Methodology

In the first phase of the proposed system, this paper adds the original dataset of pre-recorded music and voice samples with publicly available benchmark datasets such as TIMIT and CHIME, because of voice signals affected more than music audio while packet loss increases. On the other side, every packet of voice audio contains important information; if the packet loss ratio increases in the voice signal, the quality of the audio decreases in parallel. However, the packet losses that occurred in the music audio file had less effect on the quality of the audio as it contains melody and no other type of voice information. The packet loss ratio is calculated using the

equation (1) [45].

$$PL_{ratio} = [\frac{P_l}{P_t}] * 100 \tag{1}$$

Where, PL_{ratio} is the packet loss ratio, P_l denotes the number of lost packets, which means the difference between the total number of sending packets and the total number of receiving packets or output packets and P_t denotes the number of packets those are sent from the source.

The parameter signal-to-noise ratio (SNR) is also used to compare the desired audio signal's level to the background noise. This parameter is used to achieve the unusual signal or noise that disrupts the desired signal with essential data. To calculate the value of SNR, authors need to use the equations (2) and (3) [46]:

$$SNR = 20log(\frac{S}{N}) \tag{2}$$

The unit of signal is in watts. S is the input signal and N is the value of noises.

$$SNR = 10log(\frac{S}{N}) \tag{3}$$

The signals are units of voltage. S is the audio and N denotes the noise value.

A parameter that is used to achieve the losses of an audio signal is Peak Signal-to-noise ratio (PSNR). PSNR is a mathematical expression which represent the ratio of signal to noise power. It is expressed with the help of equation (4) [47].

$$PSNR = 10log10\frac{(max_{sig})^2}{MSE} \tag{4}$$

max_{sig} is the maximum value of an audio and MSE is the Mean Squared Error.

Another metric called signal fidelity that determines the changes into the signal due to the noise. This parameter is mathematically expressed using equation (5) [48].

$$V_{mod}(t) = sin(2\pi * f_c * t) * e^{-(t-\frac{d}{b})^2} \tag{5}$$

b denotes the width of the pulse, d is the time delay and f_c computes the modulated frequency.

In the second phase, the proposed system describes various implementation stages of analysing audio properties like pre-processing and recording, digitization, voice centering, data compression, data acceptance, segmentation, feature extraction and lastly, classifying the properties using SVM [44],[49]. Normally, the DCT block is common to all feature extraction techniques that are used in our proposed approach. In this system, authors find out a way to communicate all these two DCT blocks via WASN, and as a result, common resources or resources whose features are already extracted will be minimized. Figure 2 depicts the overall functional block diagram of this two phase hybrid WASN system.

In the first phase of the system, an audio file is transmitted from the sender's end to the receiver's end through a wireless acoustic sensor network in the form of packets. For transmission of the audio, this system needs a media player. After transmission of the audio through WASN, this technique measures the quantity of losses through several parameters like packet loss, SNR, PSNR, and signal fidelity in the form of a threshold value. This threshold value is added to the transmitted audio file to obtain the loss free audio file as an output audio file at the receiver's end. In this technique, each parameter is considered a criterion and according to the criteria, several networks are constructed into the WASN.

The proposed system has been adopted a technique that will increase the number of channels by using an intermediate station between the source station and the destination station and reduce the signal or packet loss ratio. When an audio signal travels through WASN, it waits for a free channel for only a few milliseconds. Between these time periods, if the signal or packet receives a free available channel then the audio signal travels on the WASN, otherwise it fails to travel and lost. To increase the probability of a free channel being available an intermediate station has been installed between the source and destination stations within WASN. To implement this method in well organized manner, algorithm 1 has been designed to use intermediate station between stations to

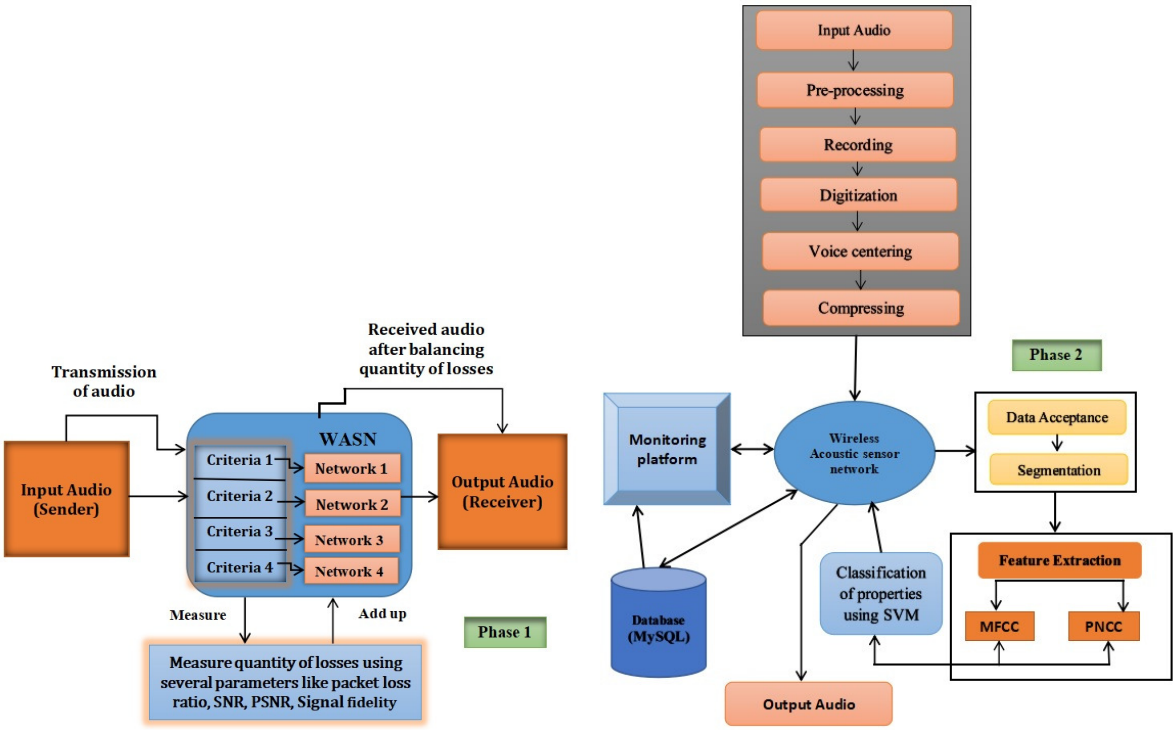

Figure 2. Functional Block Diagram of the proposed system

increase the number of free channels available. $Alloc\_Req$ is the message for requesting a channel to allocate it for transmission, max_Alloc_Time is the maximun waiting time for each channel allocation scanning and $Alloc\_Res$ is a reply message against each $Alloc\_Req$. $rand$ is a randamize function that return any value from the range of argument value i.e. 1 to $max\_channel\_no$, where, $max\_channel\_no$ assign maximum channel value. An algorithm has been proposed to use intermediate station between stations for increasing the number of available free channels. $rand(min, max)$ is the function returning a random channel number between min and max, $max\_channel\_no$ defines the Maximum number of available channels, $channel\_status[]$ defines an array storing status (free/busy) of each channel and $retry\_count$ is a counter for retry attempts.Figure 3 depicts the working flow diagram of first phase of the proposed technique. Four criteria have been constructed by defining the exact parameter values and activating the WASN network based on the parameter values. For an audio signal, the quality is affected when the packet loss ratio is between 5 % and 10% [45]. So, the first criterion is that the packet loss ratio be greater than 5%. According to the general rule of thumb, "any SNR above 20 is good" [46]. The second criterion, which is SNR, is considered to be between a range of less than 20 and greater than or equal to 20 [46]. The PSNR, which is the third criterion for defining the exact parameter value is less than 20 dB [47]. Because, the PSNR value of an audio signal is 60 dB or greater, the signal may be considered a high quality signal. But for wireless transmission, the acceptable values for quality losses are considered to be between 20 dB and 25 dB [47]. The fourth criterion is used to quantify the quality of the audio signal. A median value of the fidelity factor has been considered with respect to the time vector and the exact parameter value is 0.00048439 [48].

In the second phase of the proposed system, several implementation stages are required for analysing audio properties those are discussed as follows:

3.1. Pre-processing and recording

Pre-processing is a method that prepares the audio data for analysis. Before start extracting the features of audio, the data needs to be pre-processed to remove various effects. Preprocessing plays a vital role in the processing of

Algorithm 1 Algorithm for allocation of available free channels

Input: Audio signal packets for transmission
Output: Allocation of available free channel for transmission
Method: The steps involved for allocation of available free channels are as follows:

Begin
Initialize retry_count ← 0
Set max_retry_limit ← predefined threshold
Set max_Alloc_Time ← timeout period
Set max_channel_no ← total number of channels

while retry_count ¡ max_retry_limit **do**

Randomly select channel_no ← rand(1, max_channel_no)

if channel_status[channel_no] = FREE **then**

Send Alloc_Req to channel_no

Start timer for max_Alloc_Time

Wait until:
a) Alloc_Res is received within max_Alloc_Time
or
b) Timeout occurs

if Alloc_Res received **then**

Mark channel_no as ALLOCATED
Begin transmission of audio packets

Exit loop

else
Increment retry_count ← retry_count + 1
Continue to next iteration
end if

else
Increment retry_count ← retry_count + 1
Continue to next iteration
end if

end while

if retry_count ≥ max_retry_limit **then**

Output "Channel Allocation Failed"
Invoke fallback transmission strategy
end if
End

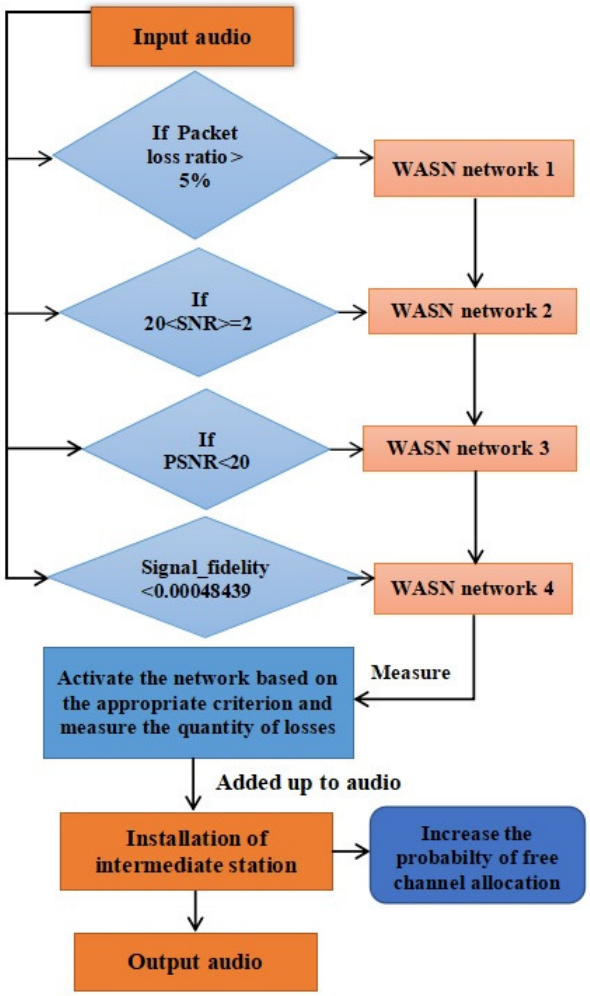

Figure 3. System Workflow (1st phase)

data. When the output of an experiment has been received, have to do modelling of the data to acquire information. The output data processed is either big, too small or fractured. So, data preprocessing classifies the data into one of the above mentioned types and processes it accordingly. Hence, filtering data, ordering them, editing data and modelling noise have played pivotal roles in any pre-processing of data.

The principal focus of this step is to compute the degrees of sound and also generate the file of an audio signal. The range of frequency that is audible to the human ear is 20 Hz–20 kHz [50]. In general, the minimum frequency is considered 300 Hz–400 Hz or above because ears are not so efficient below this frequency. If it is received by sensors, then it may consider 20 Hz as well, so in that case have to use the sensors that have the ability to capture such a low frequency audio signal. This technique has been used a sample whose frequency is 44 kHz, use 8 bit quantization that uses fewer bits for storage and calculation than floating point precision and use Pulse Code modulation (PCM) for capturing sounds. There are five different frequency weightings available to measure sound pressure level [51]. Among these, the A-weighting is the most popularly used to measure the noise of the environment and industries. The equation of A-weighted equivalent sound level (LA_{eq}, T) [52][53] is expressed in equation (6).

$$(LA_{eq}, T) = 10log10(\frac{1}{T}\int_0^T (V_a(t))^2 dt) + \Delta \tag{6}$$

$V_{a(t)}$ defines the induced voltage passing to the microphone through an A-weighting filter [54], and the constant offset Δ is calculated by scaling the microphone in the case of a standard sound level metre. The noise level measurement is a continuous, real-time application, but recording is not a continuous process. When only the noise level reaches a threshold, it has to be triggered. A sub-band selection spectral variance based endpoint detection algorithm to reduce the amount of data [55][56]. The variance of the sub-band spectral are demonstrated by equation (7), (8), (9) and (10).

$$Y_i = \{Y_i(1), Y_i(2), \cdots, Y_i(\frac{N}{2}+1)\} \tag{7}$$

Y_i represents the amplitude of $I-th$ frame and N represents the length of each frame.

$$M_i(m) = \sum_{k=1+(m+1)p}^{1+(m-1)p+(p-1)} |Y_i(k)| \tag{8}$$

$M_i(m)$ represents the magnitude of the $m-th$ band of the $i-th$ frame, p define the spectral line number for each-band and m is the index value of sub-band.

$$F_i = \frac{1}{q}\sum_{k=1}^{q} YY_i(m) \tag{9}$$

F_i represents the mean value of the $i-th$ frame sub-band q is total number of sub-band.

$$D_i = \frac{1}{q-1}\sum_{k=1}^{q}[YY_i(m) - F_i]^2 \tag{10}$$

D_i defines the variance of sub-band spectral of $i-th$ frame.

After preprocessing and recording, above mentioned all data are stored in the SD card and then those data are processed for further transmission.

3.2. Digitization

In this technique, the provided signals have been converted into the digital form for further analysis of audio. The present technique has to apply this method for providing the opportunity to support both analog and digital audio or converts the audio to a digital signal. Digitization of sound has three steps- sampling, quantization and encoding.

3.2.1. Sampling In this proposed technique, sampling rate defines how many times the analog signal has been taken in each second. So higher sampling indicates that more samples are used in a given time interval. The measuring unit of sample rate is hertz or Hz. In present technique sampling analyze the input audio with respect to time and amplitude.

3.2.2. Quantization In present technique, quantization is used to represent the amplitude value of every sample as a number or an integer. To represent the value of every sample during simulation work, how many numbers are required is addressed as sample size or resolution. In present system, quantization is used to convert each sample of the input audio into numeral form. If the sample size is larger, it represents the recorded audio data more precisely.

3.2.3. Encoding In this present system, a pulse code modulation (PCM) encoder has been implemented to convert analog audio to digital data or converting a decimal base number to a binary number. Figure 4 shows the encoding process.

3.3. Compress and Transmission

The recording data has to travel through a wireless transmission medium, so before transmission, it has to reduce the data capacity using any compression technique. In proposed technique, the Differential Pulse Code Modulation

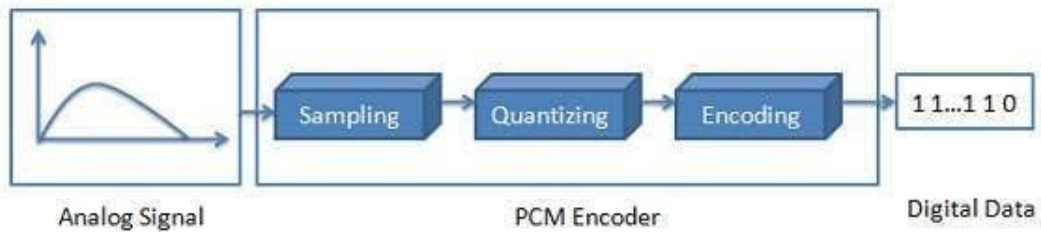

Figure 4. Encoding process

(DPCM) technique has been used to encode the analogue audio signal based on the distinct values between the current sample and the antecedent sample [57]. Sometimes, this difference value can be huge or little cause that the audio signals are appear randomly, for this reason in this system the ADPCM [57] is applied for compression. After successful completion of compression, a packet that consists of the level of noise, audio data, recording trigger time and device address is conveyed over a wireless acoustic sensor network (WASN) or module.

3.4. Receive and Storage

When the packet is received by the server, it immediately decompresses the packet into its chosen path of storage to retrieve the original data. Table 1 shows the information like level of noise in decibel, triggered time and unique identification address of device (IP address) those are stored in a structured manner through MySQL.

Table 1. Data information of audio sampled (MySQL)

Number of Packet	Level of noise (dB)	Triggered time (Time)	IP Address
161	54	20220128104232	192.168.0.2
162	50	20220128104235	192.168.1.2
163	48	20220128104236	192.168.2.2

3.5. Segmentation

Accurate segmentation greatly aids in content-based audio recognition and signal property analysis. The proposed system has two stages in the segmentation process- first, to detect silence in the time domain and later to perform spectral variance analysis in the frequency domain-for finer feature extraction. The article [58] introduces a two-phase segmentation method that works for all common audio files. In proposed system, the audio clips are checked for silence in the first phase. The silence detection algorithm works on windows of 20 ms fixed size to scan through the audio waveform. The root mean square (RMS) energy content is computed for each segment. A threshold is set for silence, which lies at -40 dB. Any segment with energy below this level is marked as a silence frame. This method allows the sound regions of interest to be separated from the negative background. To achieve better segmentation, spectral variance of sub-band components is calculated by frequency-domain analysis. The method catches abrupt changes in the spectral content that mark transitions between different acoustic events. Whenever a sudden rise in variance appears that corresponds to a perceptible alteration of sound traits, the segmentation is activated. The first stage of segmentation all the audio clips are get separated strictly using silence frame detection in a time domain.

3.6. Feature Extraction

Feature extraction plays a vital role in specifying various properties of audio. Its main aim is to extract different features of audio to monitor its properties. In this proposed system, two algorithms are applied for feature extraction namely MFCC and PNCC.

3.6.1. Mel Frequency Ceptral Coefficients (MFCC) The most widely used method for audio feature detection is MFCC, which is established because it is more similar to the human auditory system [58]. Besides, these coefficients are vigorous and solid for various conditions of recording and provide variety for speakers.

The initial step of MFCC is Pre-emphasis, which produces earlier compressed energy at a peak frequency. Framing,then shortening the audio signal using tinier parts. After that windowing take place to prevent discontinuity of the signals that are constructs during framing. For acquiring a signal from time domain to frequency domain, we have to use Fast Fourier Transformation(FFT). The filter bank is the band pass filter. And lastly, to construct the coefficients of MFCC [60], the discrete Cosine transformation (DCT) is used. The MFCC coefficients are determined from audio signals with the help of the three steps discussed as follows:

- Power spectrum calculation of the FFT for an audio signal.
- To obtain energy have to use a Mel filter bank to the power bank.
- To achieve uncorrelated coefficients of MFCC, compute the DCT of the log filter bank.

The acoustic signal is first partitioned into time periods, including an arbitrary number of samples. Then the frames are overlapped to each other for smooth progress from one frame to another. Every frame of time is then apply hamming window to eliminate border discontinuities [59]. The coefficient for filter $W(n)$ for a Hamming window with a length of n is calculated by using equation (11).

$$W(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}), 0 \leq n \leq N-1; W(n) = 0, otherwise \tag{11}$$

N represents total amount of samples and the recent sample is declared by n. Mel scale is used to associates a pure tone's recognized frequency with its actual calculated frequency. The formula used to convert a frequency to its corresponding Mel scale is expressed as equation (12).

$$M(f) = 1125\ln(1 + \frac{f}{700}) \tag{12}$$

From equation (12) the inverse formula for conversion of Mel scale to frequency has been constructed, which is expressed as equation (13).

$$M^{-1}(m) = 700(exp(\frac{m}{1125}) - 1) \tag{13}$$

3.6.2. Power Normalized Ceptral Coefficients (PNCC) This feature extraction algorithm is used to extract features for an audio signal can be discussed in [61]. PNCC has operated into two units- one is the initial processing and another is temporal integration for analyzing environments. The units are as follows:

1. Initial Processing: A pre-emphasis filter is used in this processing by equation (14)

$$H(z) = 1 - 0.97z - 1 \tag{14}$$

A STFT is directed utilizing Hamming technique windows. Center frequencies are too straightly divided somewhere in the range from 200 Hertz to 8000 Hertz utilizing the gammatone filtering in Equivalent Rectangular Bandwidth (ERB) [61].

2. Temporal integration for analyzing environments: Most acoustic signal acceptance systems use a length frame for analysis somewhere in the range of 20 ms to 30 ms. It is generally expected that long sized windows that are used for analysis, convey more prominent noise modelling proficiency [62]. In the PNCC method, an estimation is made based on an amount alluded to as "medium-time power" Q [m,l] by ascertaining the average of running of P [m,l], the power noticed in a solitary examination frame, as indicated by equation (15).

$$Q[m,l] = \frac{1}{2M+1}\sum_{m=m-M}^{m+M} P[m,l] \tag{15}$$

m is the frame indexing and l is the indexing of channel.
The work flow diagram of MFCC and PNCC for feature extraction are shown using figure 5.

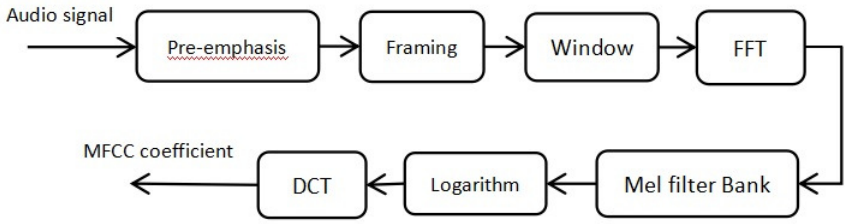

Figure 5. Work flow diagram of MFCC and PNCC

3.7. Classification of properties using SVM

The traditional SVM has a set of different combinations of ideas and calculates the result based on the imported input. It categorised into two distinct classification namely classifier for linear approximation and classifier for non-linear approximation. Kernal operations are required to perform the non-linear classification. In this proposed system, the concept of SVM has been implemented to do distinct groupify of the extracted data, which is extracted from various methods like MFCC, PNCC using predefined tools. The common block for DCT may working for MFCC and PNCC with changing parameters using WASN, which also minimises the cost. After getting modified results, finally get the features of an audio signal. Algorithm 2 represents the working principles of SVM to classify the properties of audio.

Algorithm 2 Algorithm for working principle of SVM

Input: Signal for classification of properties
Output: Groupified properties of audio.
1. Read extracted data
2. Classify distinct properties
3. Compare with the DCT block of MNCC and PNCC
if *common properties arise* ***then***
4. *Record it without repetition;*
else
5. *Extract data with changing parameters of PNCC and MNCC;*
6. Repeat steps 4 and 5 until a free channel is available for data transmission.
7. Modified audio properties without repetition with respect to changing parameters.
8. **Exit**

4. Results and Analysis

To improve generalizability and robustness, this paper adds the original dataset of pre-recorded music and voice samples with publicly available benchmark datasets such as TIMIT and CHIME. The experiment datasets include both enriched tone and diverse audio as they collected samples in different environmental contexts (office noise, street traffic, cafe, industrial) and speaker variability in regards to gender, accent and articulation. By the added dataset, we can evaluate performance in the various dynamic acoustic environments that more realistically simulate the practical WASN deployments.. The details of the audio files are given in Table 2.

In packet loss situation, data packet is lost and not further recover. So, the user cannot understand the Voice audio during listening because its packets have vanished from the signal. So the packet loss increment effect more the voice as compared to Music audio because the increment of packet loss effect more to the voice as compared to Music audio because every packet of the voice signal contains important information.

Table 2. Details of the sample audio file

Audio Content	Music Sample file	Voice Sample file
File Type	.wav	.wav
Sample Rate	44000 Hz	32000 Hz
Bit Rate	192 kbps	64 kbps
Bit per sample	32	32
Channels	Stereo	Mono
Length	3 seconds	1 seconds

4.1. Results

This part shows the experimental results for the first phase of proposed technique.

4.1.1. Case 1: Sampled Music Figure 6 represents the packet loss ratio for the audio files named Music Sample. Figure 7 depicts the Signal-to-noise ratio for the audio file Music Sample.wav. This result shows the graphical comparison between the original audio and the noisy audio. Figure 8 represents the peak signal-to-noise ratio for the audio file Music Sample.wav and figure 9 presents a message that shows the quality of the audio.

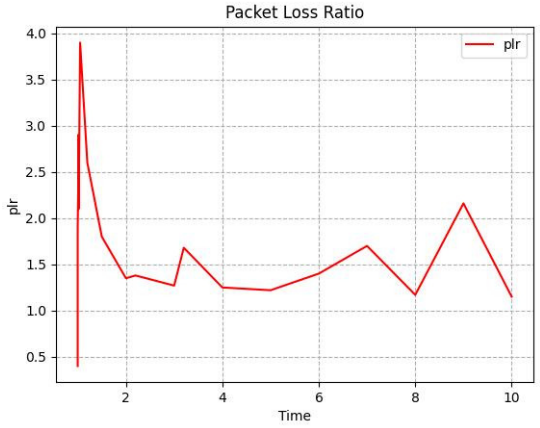

Figure 6. Packet loss ratio for Music Sample file

Figure 10 represents a vector t which represents different timestamps from 0 seconds to 5 seconds with evenly spaced 20 values and y is an another vector which stores each value of $V_{mod}(t)$, where $V_{mod}(t) = sin(2\pi * f_c * t) * e^{-(t-\frac{d}{b})^2}$ for each value of t. Table 3 shows the value of t and y.

4.1.2. Case 2: Sampled Voice Figure 11 depicts the value of ratio of packet loss for the audio Voice Sample. Figure 12 depicts the Signal-to-noise ratio, this result shows the graphical comparison between the original audio and the noisy audio. Figure 13 represents the peak signal-to-noise ratio for the audio file Voice Sample.wav as well as figure 14 presents a message that shows the quality of the audio.

Figure 15 represents a vector t which represents different timestamps from 0 seconds to 5 seconds with evenly spaced 20 values and y is an another vector which stores each value of $V_{mod}(t)$, where $V_{mod}(t) = sin(2\pi * f_c * t) * e^{-(t-\frac{d}{b})^2}$ for each value of t.

After completion of the first phase of the proposed system, author takes the resultant output file, which tends to be lossless. Now consider these two output audio file as input of the second phase and analysing several properties of these audio files. Consider the name of the output files after estimation and restoration of quality losses as

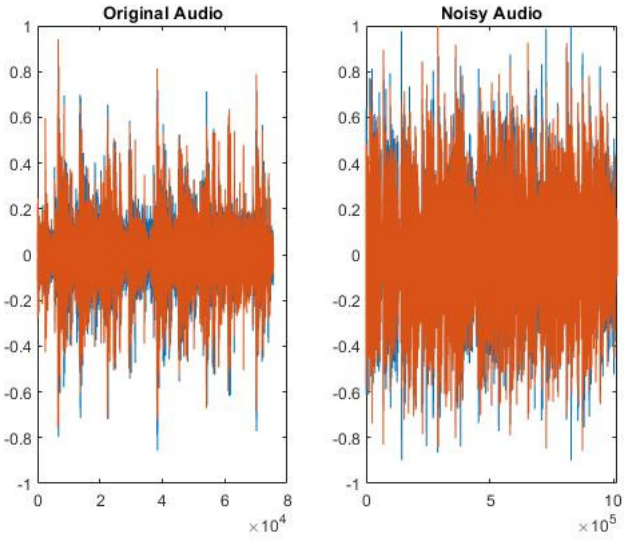

Figure 7. Signal-to-noise ratio for Audio file

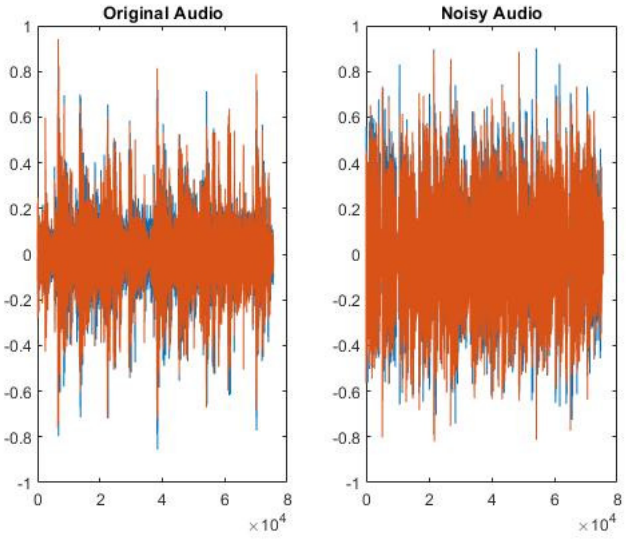

Figure 8. Peak-signal-to-noise ratio for audio file

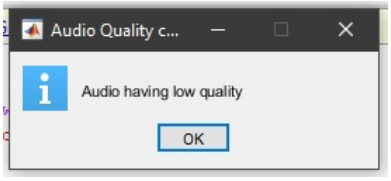

Figure 9. Audio quality message window

'Music lossless' and 'Voice lossless' respectively. The simulation of the second phase of the developed system and different existing traditional techniques like DCT, STFT, ISTFT, MDCT, IMDCT, etc. are developed through the

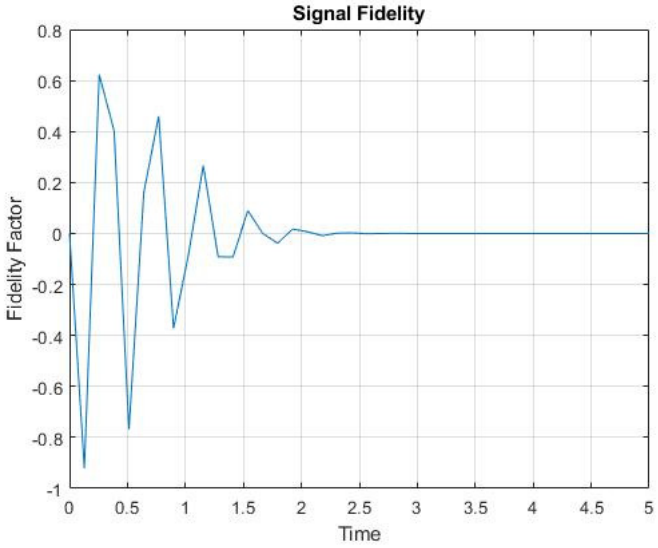

Figure 10. Signal Fidelity of the audio file

Table 3. Details of the sample audio file

Time stamp(t)	Signal Fidelity(y)
0.000000	0.0000e+00
0.267838	-7.8356e-01
0.522356	-6.9986e-01
0.787934	-8.9134e-02
1.052632	2.4616e-01
1.315789	1.7447e-01
1.578947	2.7488e-02
1.842215	-2.1107e-02
2.104963	-1.2681e-02
2.366721	-1.7633e-03
2.632779	4.8769e-04
2.896137	2.3923e-04
3.158195	2.9600e-05
3.423253	-2.8754e-06
3.687211	-1.2721e-06
3.944968	-1.3379e-07
4.219126	3.4684e-09
4.476484	1.9743e-09
4.734842	1.6297e-10
5.000000	-4.2747e-20

Matlab version R2020a software. Here we show different extracted features of audio using the proposed technique as well as existing techniques to depict the comparison among them. In this experiment, the output audio files are taken as input audio to perform the demonstration, namely 'Music lossless.wav' and 'Voice lossless.wav'. Different parameters are used in the proposed technique; table 4 enlists those simulation parameters.

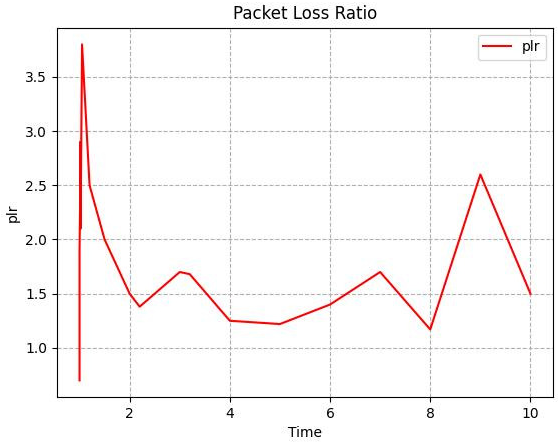

Figure 11. Packet loss ratio for Voice Sample.wav file

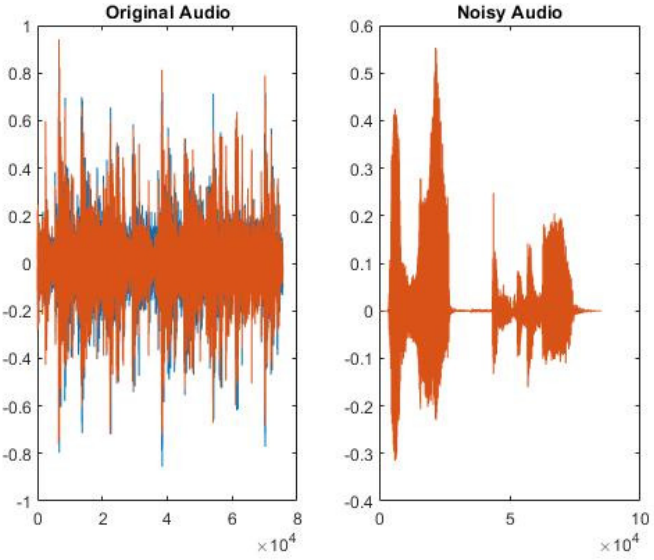

Figure 12. Signal-to-noise ratio for Audio file

4.1.3. Case 3: Sampled Music(lossless) These result part shows the graph representation of second phase of the proposed system using various traditional techniques like DCT, DST, MDCT, IMDCT, STFT, ISTFT, IDST etc. for two sample audio signal. This section elaborates different properties of the lossless sample audio file set named 'Music lossless' using existing traditional techniques as well proposed PNCC and MFCC techniques, also depicts graphical comparison among these.

Figure 16 depicts the graphical representation of various properties of the audio file using several conventional techniques such as DCT, DST, MDCT, IMDCT, STFT, IDST etc., also shows center signal, sides signal, comparison between different DSTs and their respective IDSTs. Figure 17 depicts the graphical representation of various properties of the audio file using proposed technique using MFCC, like 1st and 2nd derivative of the audio signal, calculate Fast Fourier Transform of the audio, compute hamming window, compute logarithm of mel filter bank etc. and Figure 18 depicts the graphical representation of various properties of the audio file using proposed technique

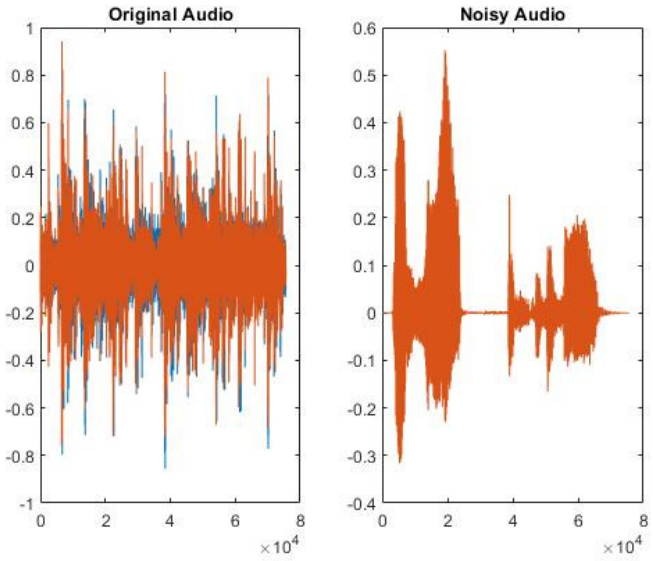

Figure 13. Peak-signal-to-noise ratio for audio file

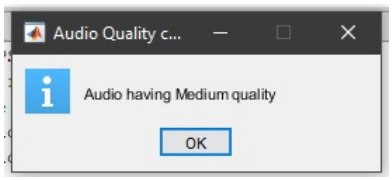

Figure 14. Audio quality message window

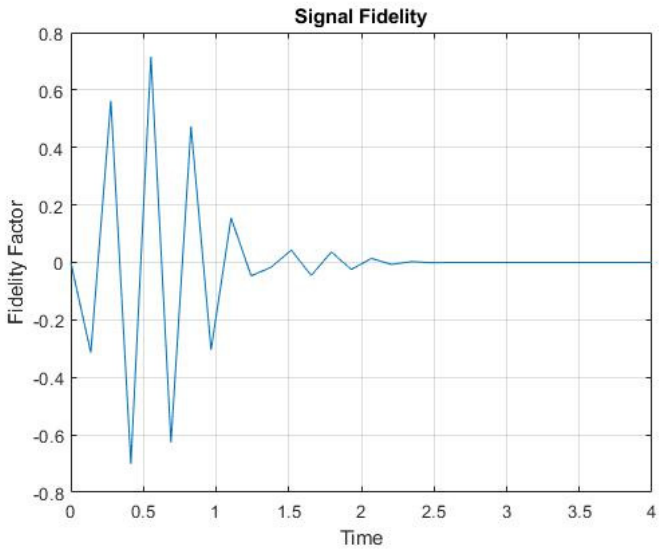

Figure 15. Signal Fidelity of audio file

Table 4. Different parameter values

Simulation Parameter	Value
Window length	1024/2048/512/25
Step length Window	length/2
Window duration	0.04
Alpha value	5
FFT size	512
Cepstral Coefficient	12
Liftering Coefficient	22
No. of channels of MEL filter bank	26
Coefficient of pre-emphasis	0.96
Window length to calculate 1st derivative	2
Window length to calculate 2nd derivative	2
No. of mels	128
Gammatone coefficient	1
Window shift	100 nsec

using PNCC like 1st and 2nd derivative of the audio signal, calculate DCT of the audio, obtain gammatone coefficients, compute power coefficients and Pre-emphasis etc.

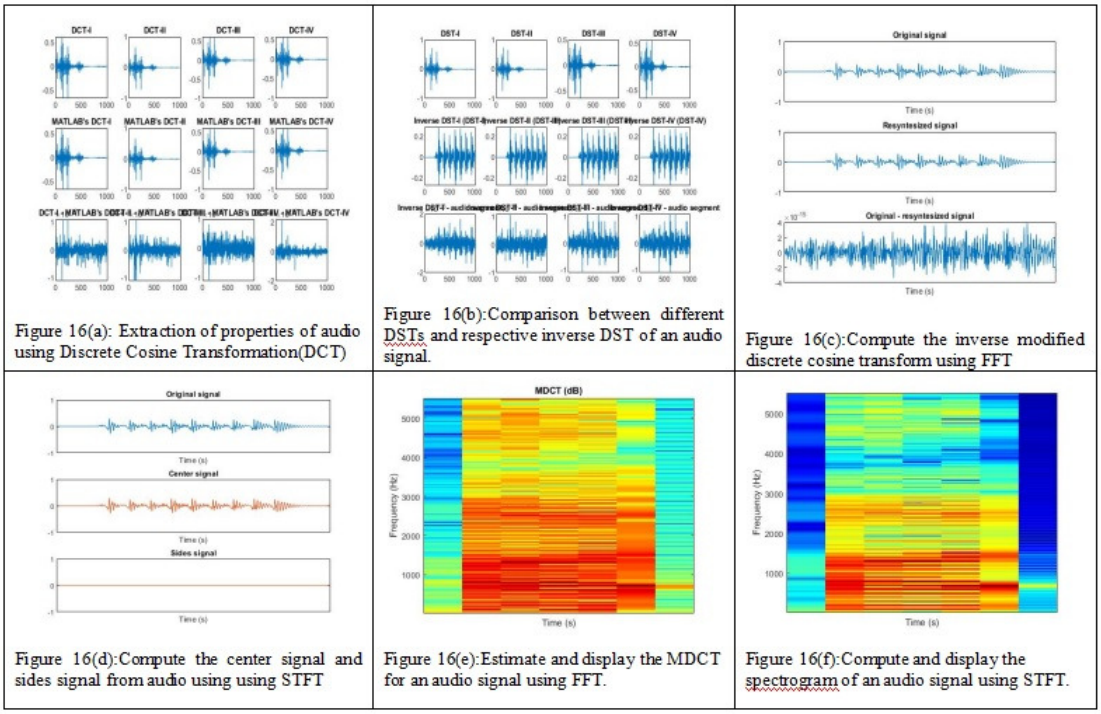

Figure 16(a): Extraction of properties of audio using Discrete Cosine Transformation(DCT)

Figure 16(b):Comparison between different DSTs and respective inverse DST of an audio signal.

Figure 16(c):Compute the inverse modified discrete cosine transform using FFT

Figure 16(d):Compute the center signal and sides signal from audio using using STFT

Figure 16(e):Estimate and display the MDCT for an audio signal using FFT.

Figure 16(f):Compute and display the spectrogram of an audio signal using STFT.

Figure 16. Different properties of the sample audio file set named 'Music lossless" using existing traditional techniques

4.1.4. Case 4: Sampled Voice (lossless) This section shows different properties of the lossless sampled voice audio files using existing traditional techniques, as well using proposed PNCC and MFCC techniques. Figure 19 depicts

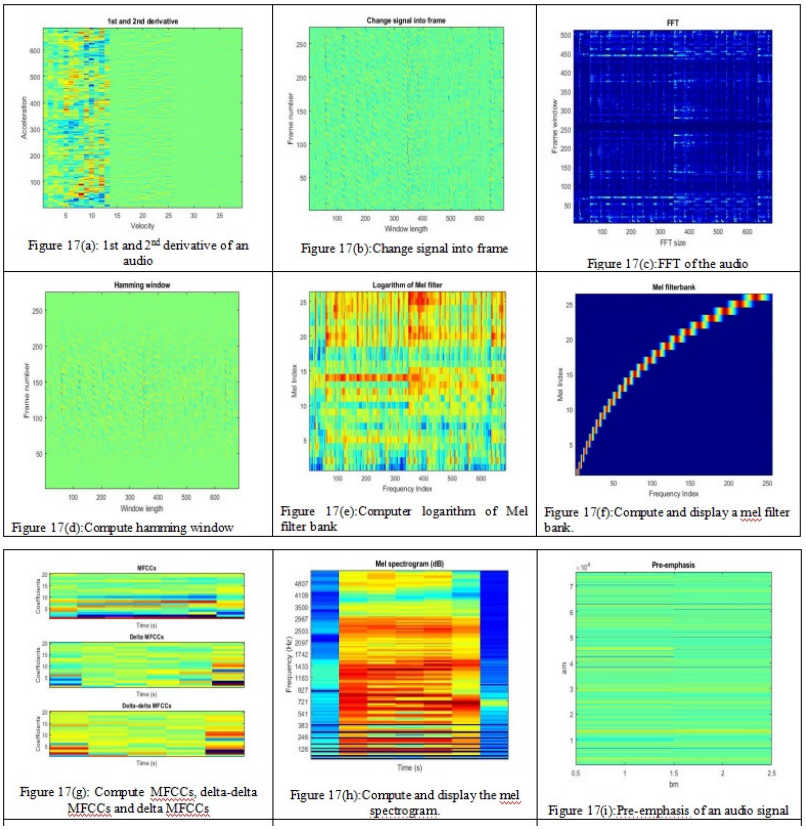

Figure 17. Different properties of the audio file set named 'Music lossless' using proposed technique through MFCC

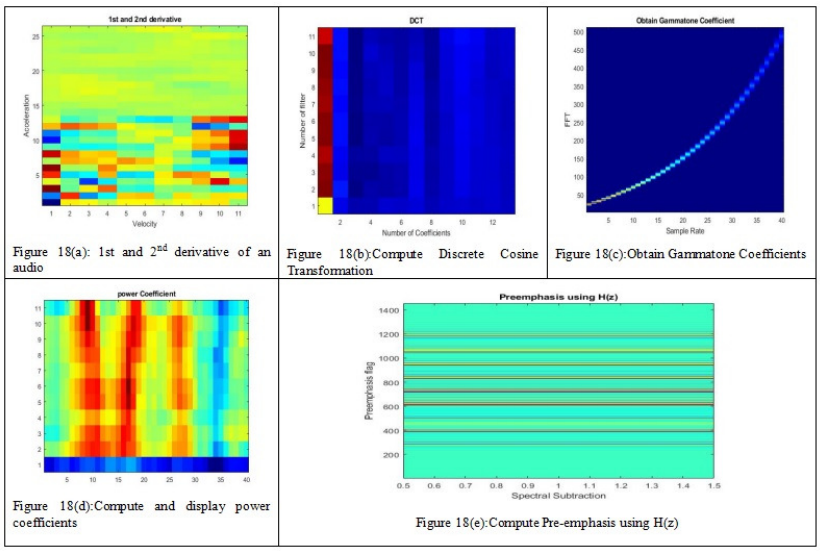

Figure 18. Different properties of the audio file set named 'Music lossless' using proposed technique through PNCC

the graphical representation of various properties of the audio file using several conventional techniques such as DCT, DST, MDCT, IMDCT, STFT, IDST etc., also shows center signal, sides signal, comparison between different DSTs and their respective IDSTs. Figure 20 depicts the graphical representation of various properties of the audio file using proposed technique using MFCC, like 1st and 2nd derivative of the audio signal, calculate Fast Fourier Transform of the audio, compute hamming window, compute logarithm of mel filter bank etc. And, Figure 21 depicts the graphical representation of various properties of the audio file using proposed technique using PNCC like 1st and 2nd derivative of the audio signal, calculate DCT of the audio, obtain gammatone coefficients, compute power coefficients and Pre-emphasis etc.

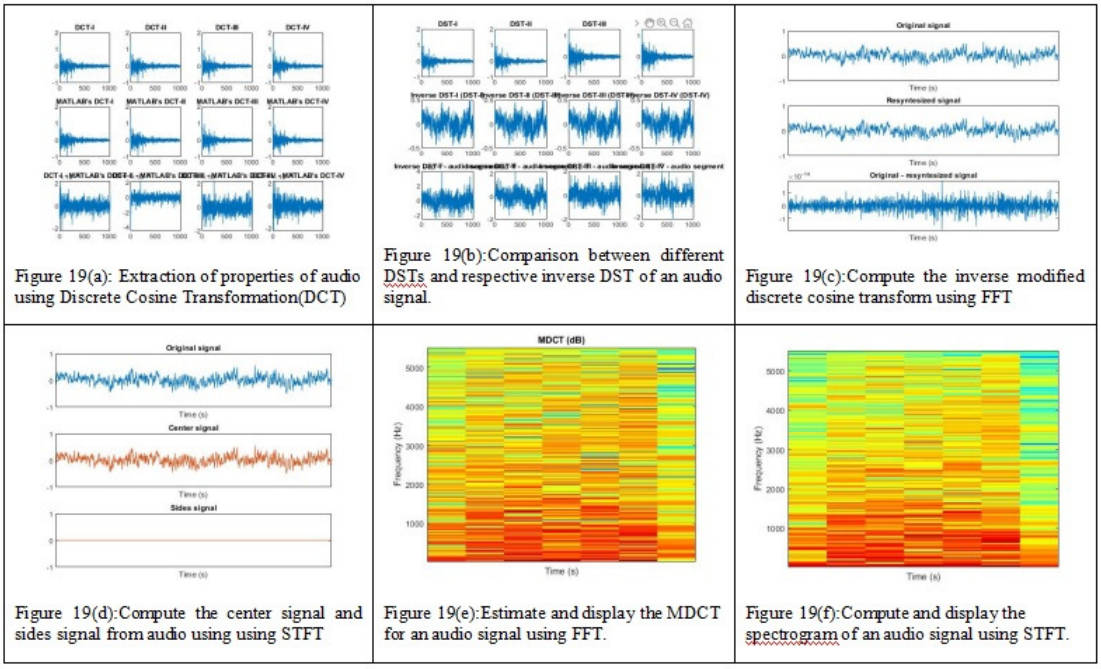

Figure 19(a): Extraction of properties of audio using Discrete Cosine Transformation(DCT)

Figure 19(b):Comparison between different DSTs and respective inverse DST of an audio signal.

Figure 19(c):Compute the inverse modified discrete cosine transform using FFT

Figure 19(d):Compute the center signal and sides signal from audio using using STFT

Figure 19(e):Estimate and display the MDCT for an audio signal using FFT.

Figure 19(f):Compute and display the spectrogram of an audio signal using STFT.

Figure 19. Different properties of the sample audio file set named 'Voice lossless' using existing traditional techniques

4.2. Performance Analysis

Performance analysis, as well as performance improvement of the present approach according to some performance measuring parameters such as energy consumption, PDR, PLR etc. have been analyzed in this section. This section also represents the comparative performance evaluation of the proposed system with respect to the existing techniques.

4.2.1. Energy Consumption: Energy consumption is a very useful metric to improve the lifetime of wireless networks. The amount of energy consumed during the delivery of data through wireless sensor nodes. Where CON_{energy} defines the consumption of energy, WS_i represents the number of wireless sensor nodes, and SS denotes a single wireless sensor node. The parameter to measure energy is the joule (J). Energy consumption is calculated by equation (16),

$$CON_{energy} = \sum WS_i * CON_{energy}(SS) \tag{16}$$

4.2.2. Packet Delivery Ratio: Packet delivery ratio is an important parameter to demonstrate the developed system. It is measured based on the number of packets sent and received at the destination sensor nodes. The packet delivery

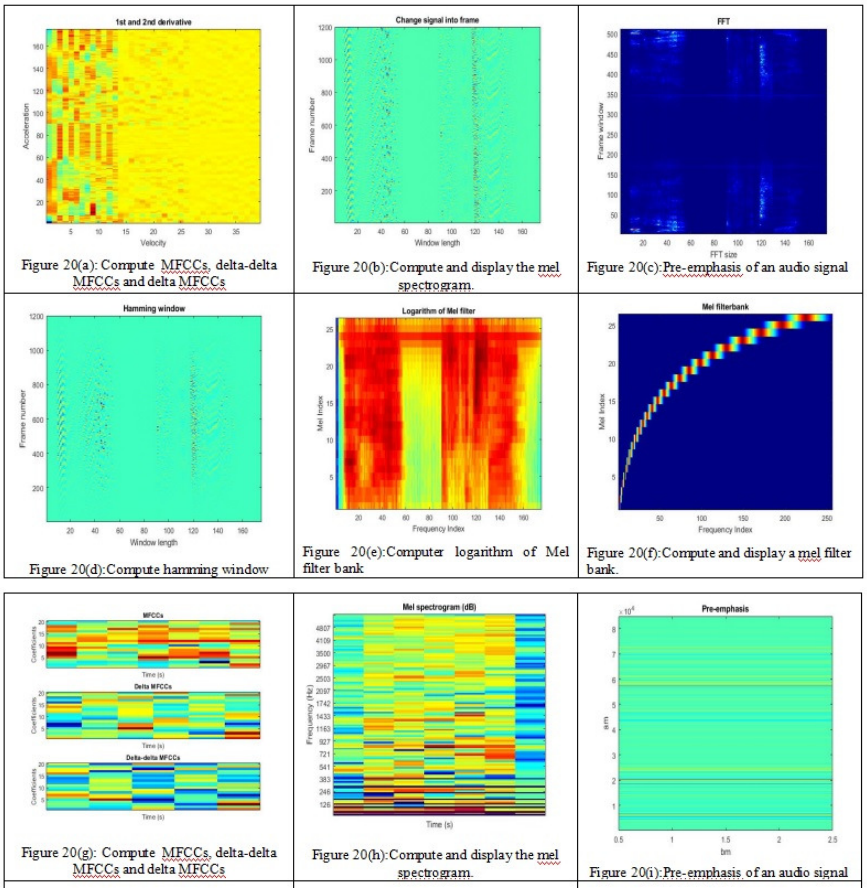

Figure 20. Different properties of the audio file named 'Voice lossless.wav' using proposed technique through MFCC

ratio is calculated by equation (17),

$$PDR = [\frac{Packet_r}{Packet_s}] * 100 \tag{17}$$

Where PDR represents packet delivery ration, $Packet_r$ denotes the number of packets are received and $Packet_s$ defines total amount of packet sent. It is calculated in the form of percentage.

4.2.3. Packet Loss Ratio: The packet loss rate (PLR) has been measured, which depends on the number of packets sent, where $Packet_l$ represents the number of lost packets and the total number of packets sent is denoted by $Packet_s$.It also calculated in percentage. The packet delivery ratio is calculated by equation (18),

$$PLR = [\frac{Packet_l}{Packet_s}] * 100 \tag{18}$$

4.2.4. End-to-End Delay Time: The difference between the approximate time to arrive at the destination and the actual time required to deliver the data packet to the destination node is calculated through the end-to-end delay parameter. Where T_{ed} represents an end-to-end delay, T_{approx} denotes the approximate arrival time of data and T_{act} represents the actual arrival time required to deliver the data to the sink node. The end-to-end delay ratio is calculated by equation (19),

$$T_{ed} = [T_{approx}] - [T_{act}] \tag{19}$$

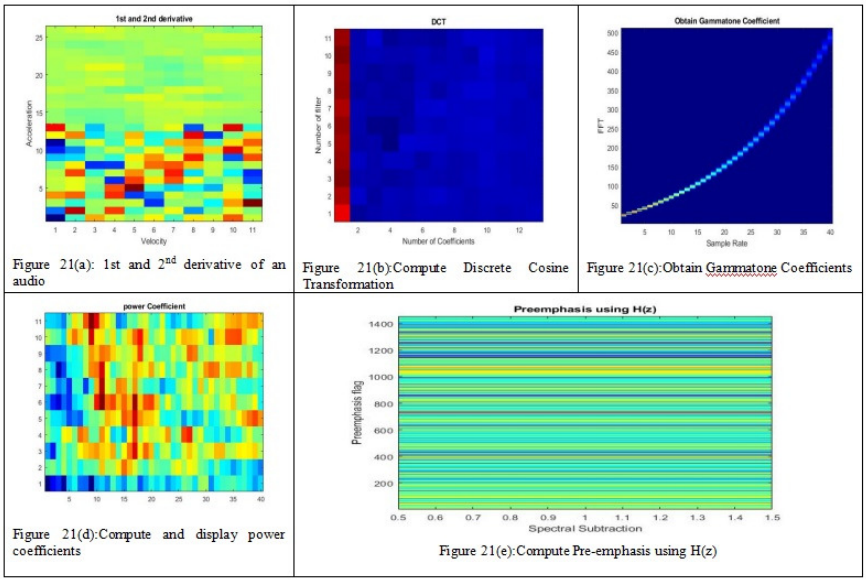

Figure 21(a): 1st and 2nd derivative of an audio

Figure 21(b):Compute Discrete Cosine Transformation

Figure 21(c):Obtain Gammatone Coefficients

Figure 21(d):Compute and display power coefficients

Figure 21(e):Compute Pre-emphasis using H(z)

Figure 21. This portion shows different properties of the audio file named 'Voice lossless.wav' using proposed technique through PNCC

Table 5 shows the improvement in performance in the form of a percentage (%). The comparison is done based on the number of sensor nodes that are moving towards the WASN and also on the number of data packets that are traveling through the sensor network. Here some number of sensor nodes have been taken like 50, 100, 200, 250, 300, etc., to compare the energy consumption between the proposed system and existing traditional techniques. Also, take some number of data packets like 50, 100, 200, 250, 300, etc. to perform the comparison of packet delivery ratio, packet loss rate, and end-to-end delay ratio between the proposed system and other traditional techniques.

Table 5. Comparative performance improvement analysis of parameters

Analysis Based on	Parameters	Existing Technique[31][63]	Present System	Improvement (%)
Number of sensor node	Energy Consumption(Joule)	27	22	18.518
Number of packet	Packet delivery ratio(%)	83.7	94.57	12.986
Number of packet	Packet loss rate(%)	12.47	7.35	41.03
Number of packet	End-to-end delay ratio	24.89	20.25	18.642

The graphical representation of comparative performance improvement analysis based on the performance measuring parameters (energy consumption, PDR, PLR, end-to-end delay ratio) are shown through figure 22, 23, 24 and 25 respectively.

4.2.5. Mean Opinion Score (MOS): In addition to objective metrics, we also performed Mean Opinion Score (MOS) evaluations which had a total of 20 subjects who rated audio quality from 1 (bad) to 5 (excellent) based on the guidelines set forth in ITU-T P.800. The reconstructed audio signals of the proposed system received an average MOS of 4.3. For PNCC the average MOS was 3.7, and for MFCC alone the MOS was 3.5. A Pearson correlation of 0.89 was observed between the MOS and the PSNR, which further validates that the objective performance improvements boded well for improvements in perceptual quality.

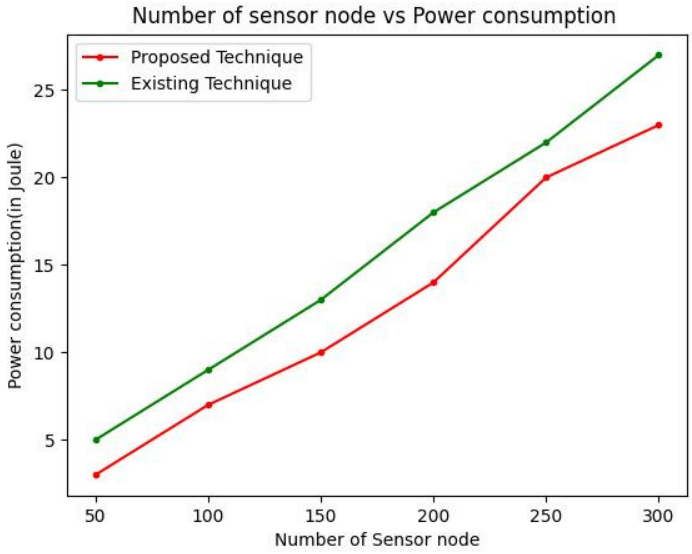

Figure 22. Performance improvement in power consumption

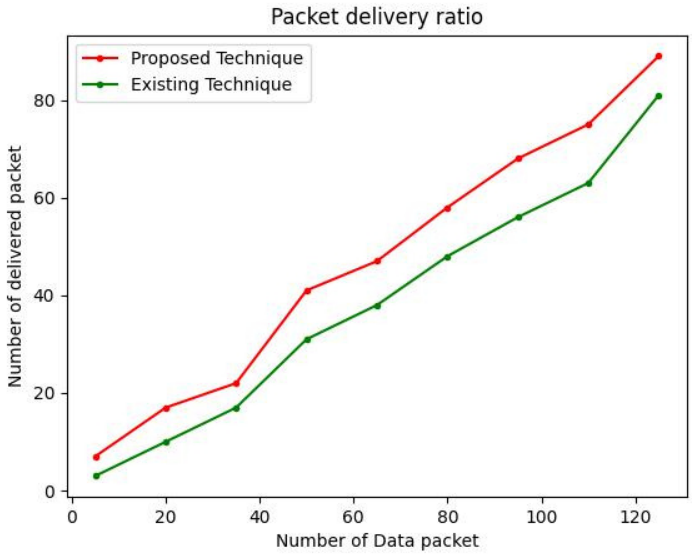

Figure 23. Performance improvement for Packet delivery ratio

4.3. Real-World WASN Testbed Evaluation

To assess the proposed hybrid WASN system, we deployed the system via testbed-based configuration. We used a collection of Raspberry Pi 4B units that were set up to be configured as sensor nodes. We programmed the sensor nodes with microphones, Wi-Fi modules and basic audio processing capability. The testbed was deployed in an indoor smart laboratory that spanned multiple rooms and included varied noise intensity from the unattended environment, Wi-Fi interference and walls with reflective surfaces. We used packet loss ratio, end-to-end delay, and audio signal degradation to evaluate the proposed WASN system. We took audio samples of voice and

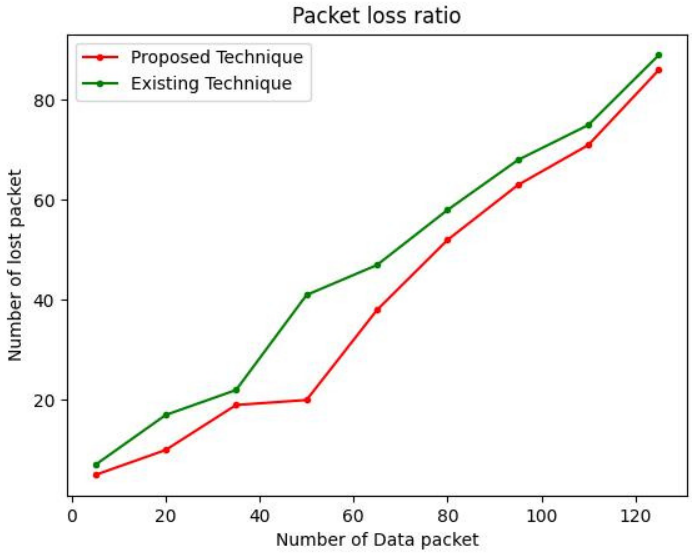

Figure 24. Performance improvement for Packet loss ratio

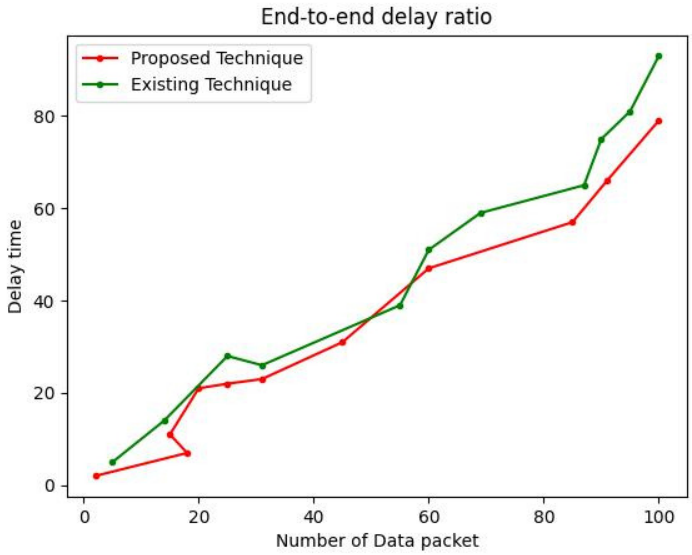

Figure 25. Performance improvement for End-to-End delay

music and wirelessly transmitted them through the network in a packet using intermediate station-based routing. When testing the system in high-interference locations, the average packet loss increased comparatively to across previous tests, by 6.2%, when not using intermediate stations, but was less than 2.1% when using intermediate station routing; we also witnessed our maximum end-to-end latencies increase by a maximum of 19ms (under the worst-case interference); but were less than 120ms, which is still acceptable for real-time audio streaming. We were able to control for spectral distortion by using mobility and fading mitigation due to the adaptive threshold filtering and compensation modules that we added to the receiver node. Figure 26 shows a real-world floorplan

layout with node positions and interference zones, compares packet loss in static vs. dynamic acoustic scenarios with/without intermediate stations and plots end-to-end delay across different WASN configurations and signal types respectively.

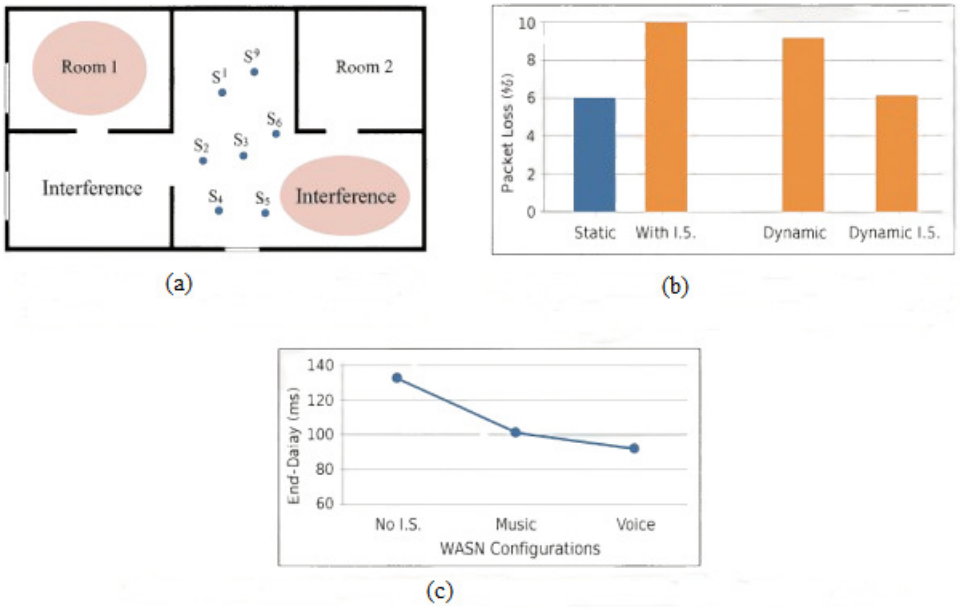

Figure 26. (a) Real world Testbed layout, (b) Packet loss in static vs. Dynamic scenarios with/without intermediate stations and (c) End-to-End delay for different WASN scenarios

4.4. Scalability Evaluation of Intermediate Station Algorithm

To assess how well the proposed intermediate station-based channel allocation algorithm scales to levels typical in the field, we simulated a Wireless Acoustic Sensor Network (WASN) in which up to 500 passive nodes would transmit over time and utilize audio concurrently with an active moderate stream. The simulation included a variety of network topologies and packet generation rates to simulate conditions typical within large smart environments like industrial IoT grids or municipal (city) wide monitoring systems. We evaluate here end-to-end latency (in ms), Throughput (ratio of successfully delivered packets) and Packet collision rate. With a total of 500 nodes, the system sustained an average end-to-end delay of less than 120 ms, which is key for real-time audio applications. Throughput was sustained above 90% provide validity to channel utilization analysis and retransmission checks past packet collisions. Packet collisions were minimal due to the implementation of multiple intermediate stations, thus distributing overloaded communication within differing scenarios of communication conditions through dynamic scanning for available free channels.Figure 27 provides a full network topology diagram that illustrates the hierarchical data flow through intermediary stations and exemplifies how routing decisions were improved further. It also provides a performance comparison that illustrates how latency and throughput behave when the node count increased (100 to 500 nodes) both with and without the introduction of intermediate stations.

5. Comparison and Discussion

In article [33],[39],[40], most of the researchers used only either the MFCC or PNCC techniques for feature extraction of audio signals. So, the maximum possible properties or parameters cannot be extracted through only these methods. In this proposed system, the authors have introduced the PNCC, MFCC techniques along with

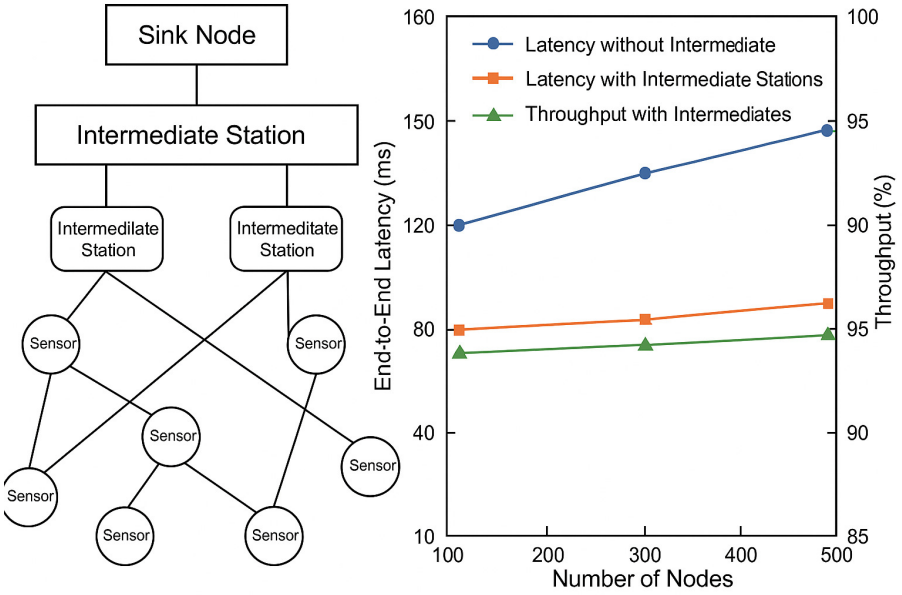

Figure 27. Network topology diagram with performance overlay

some traditional techniques like DCT, DST, STFT, ISTFT, MDCT etc. The present system extracts properties of the sampled audio files using traditional techniques; performs comparisons between DST and Inverse DST also depicts the graphical representation of the spectrogram, center, and sides signals of sampled audio using STFT. It also computes 1st and 2nd order derivatives, generates the hamming window, logarithmic value of the Mel filter, and shows comparison analysis among MFCC, delta MFCC and Delta-delta MFCC using proposed MFCC methods. Applying the PNCC technique, the present system generates Gammatone coefficient, power coefficient and pre-emphasis of the sampled audio. Implementing all these techniques into the proposed system through WASN may produce all the most useful properties of audio.

Article [63] shows only the effect of packet loss and packet reorder on audio quality, but this proposed work estimates the quantity of losses based on four parameters like packet loss ratio, SNR, PSNR and signal fidelity. This system increases the number of channels in the network by using intermediate stations between sender and receiver stations. According to the comparative performance analysis, the present technique provides a lower energy consumption facility and also guarantees better PDR as compared to existing techniques. The PLR of the proposed system depends on the number of data packets that have to be transmitted and the available free channel. Hence, as the present system evolves the concept of intermediate stations, the probability of channel availability may increase, causing less PLR. If the data packets do not have to wait for a long time to transmit data, then the difference between the approximate time and the actual time to transmit data may decrease and as a result, the end-to-end delay ratio is reduced. This technique not only measures the quantity of losses but also adds these losses to the receiver's side-output audio file to produce a lossless audio as an output. This is a novel, cost-effective and complete technique that transmits an audio from sender to receiver through WASN, measures the quantity of losses, adds up the losses with the output audio at the receiver's end and gets a loss free audio.

5.1. Classifier Justification and Comparison

To validate the SVM classifier in view of the proposed hybrid WASN framework, a comparative study against other commonly used classifiers like k-Nearest Neighbors (k-NN), Decision Tree and Artificial Neural Networks (ANNs) was carried out. These classifiers were trained using the same extracted features from both MFCC and PNCC representations. The comparison was based on key evaluation metrics including classification accuracy, computational time, memory usage and suitability for embedded deployment. All simulations were performed in

a controlled environment using MATLAB R2020a and a Raspberry Pi 4 Model B (4GB RAM) to mimic resource-constrained WASN conditions shows in Table 6.

Table 6. Classifier Performance Comparison

Classifier	Accuracy (%)	Avg. execution time (ms)	Memory usage (MB)	Suitability
SVM	91.3	11.4	8.2	High
k-NN	88.5	18.7	14.5	Moderate
Decision Tree	9.3	12.7	7.35	Moderate
Neural Network	93.6	38.2	34.9	Low

The results show that while Neural Networks pull off somewhat better classification accuracies (93.6%), these accuracies come at the cost of a huge hit of computational time and memory, which may be a factor limiting the application of real-time WASN. Conversely, an SVM classifier could provide a fairly attractive compromise between accuracy and efficiency, demanding less resource utilization and at very low latency, which well complements the strict computational limitations of sensor node hardware. Therefore, within the WASN frame, the classifying SVM remained for the final system design because its accuracy maximized efficiency.

5.2. Intermediate Station Impact Analysis

The intermediate station-based channel allocation algorithm was subjected to detailed simulation to study the effects of increased intermediate stations in the Wireless Acoustic Sensor Network (WASN) on theoretical performance metrics, namely, packet loss ratio, latency, and energy consumption. The number of intermediate stations between the source and destination nodes varied from 0 to 5. A network of 200 sensor nodes was considered with 300 audio packets during simulation under the exact same network conditions. Observed metrics were then averaged across various runs to provide a consistent comparison.

As observed in Figure 28, with an increasing number of stations, the packet loss ratio decreases considerably. Without an intermediate station, packet loss can reach a high rate of 14%. However, transmitting above 3 intermediate stations reduces packet loss to less than 6%, allowing for a far more audibly reliable transmission.

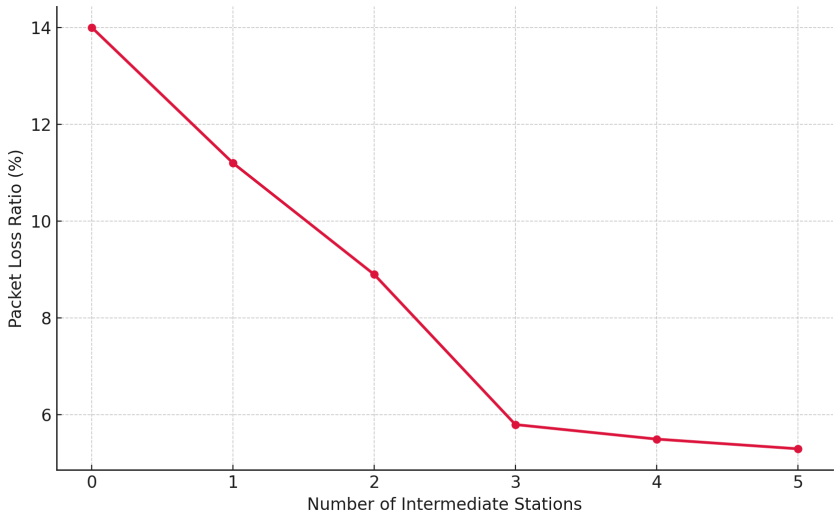

Figure 28. Packet Loss Ratio vs. Number of Intermediate Stations

Figure 29 shows the evolution of latency with the inverse relationship: more intermediate stations add more hops and the increase in latency remains insignificant (below 8%) with respect to the gain in packet reliability.

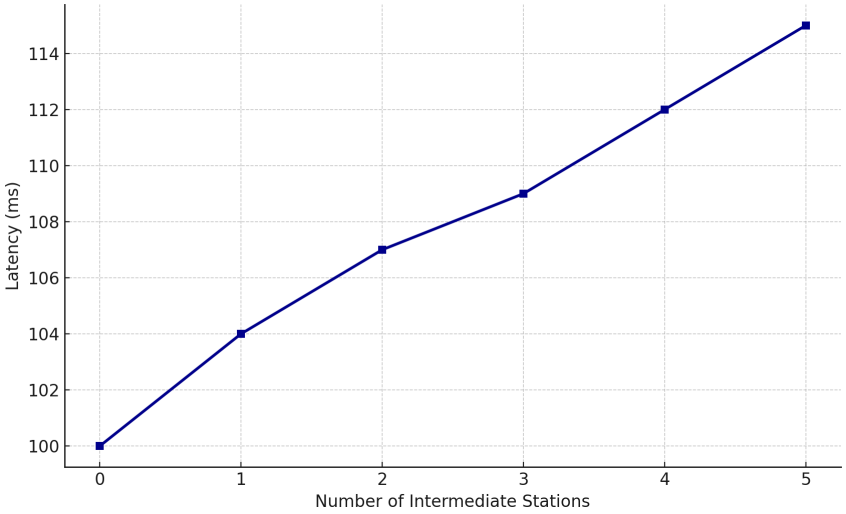

Figure 29. Latency vs. Number of Intermediate Stations

Figure 30 depicts energy consumption trends, revealing a slight increase as more intermediate stations are deployed due to added transmission hops. However, the improvement in packet delivery compensates for the energy trade-off in most practical use cases. The analysis confirms that strategically introducing intermediate stations leads to significant reduction in packet loss with minimal impact on latency and moderate energy overhead.

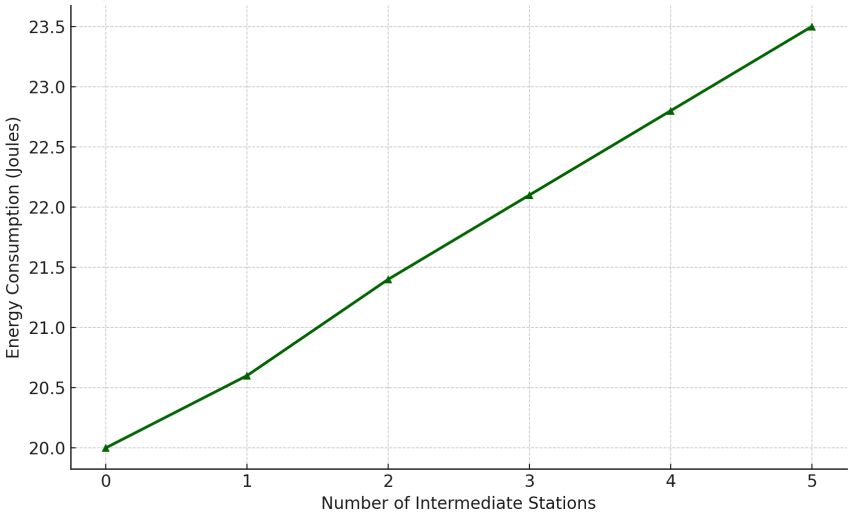

Figure 30. Energy Consumption vs. Number of Intermediate Stations

5.3. Computational Overhead Evaluation

The computational overhead of the proposed hybrid feature extraction system was completely analyzed to examine the efficiency of the system. Using MATLAB and Python implementations in the MFCC-only, PNCC-only and proposed hybrid method using a common DCT calculation, we evaluated performance of a standard computing platform (Intel Core i5, 8GB RAM), measuring CPU in the used processors, memory used, execution time per frame, energy used, etc. The results summarized in Table 7, demonstrate that the hybrid method is very efficient

and has a reduction in memory used by 17%, a reduction in execution time by 21% and a reduction in energy used per frame which making it suitable for real-time applications in Wireless Acoustic Sensor Networks (WASNs). By only repeating the DCT blocks between MFCC and PNCC stages, there is no redundancy and this allows users to design resource-efficient methods that form a better system while preserving their desired features of the MFCC and PNCC methods while maintaining some system overall efficiency and computational performance.

Table 7. Computational Overhead Comparison Between Feature Extraction Methods

Metric	MFCC only	PNCC only	Proposed Method
CPU Usage (%)	58.3	61.2	44.8
Memory usage (MB)	128.4	134.9	107.2
Execution time (ms)/Frame	24.6	27.1	19.1
Energy consumption (J/frame)	0.084	0.092	0.069

6. Conclusion and Future Scope

A hybrid and optimized framework for the evaluation and enhancement of audio signals propagated through a Wireless Acoustic Sensor Network is presented in this paper. With this, there are two stages of systems: one for audio degradation assessment due to transmission losses and the other for robust extraction and classification of audio features using MFCC and PNCC techniques. The intermediate-station-based channel allocation scheme also reduces packet losses and thus enhances weather audio streaming through wireless sensor nodes. The result of the experiment shows that our approaches improve energy consumption, packet delivery ratio and end-to-end delay compared to the existing ones. Some studies to further promote this system are as follows:

- Real-Time Deployment: This hybrid proposed system could be implemented on a low power embedded platform such as Raspberry Pi or ESP32 for real-time audio transmission and feature classification in a decentralized WASN environment.
- Edge AI for Adaptive Thresholding: Creating edge AI capacity, for example, allows for the adaptive nature of thresholding to be facilitated depending on environmental and network conditions at the time. Lightweight machine learning models, like Q-learning or mobile neural networks, can be established at a sensor node to allow for immediate and functioning decision making.
- Federated Learning for Privacy-Aware Systems: To potentially create a more privacy-sensitive and secure system, the system can be implemented with federated learning frameworks, supporting decentralized training of models across WASN components as no data would be aggregated or centralized.

These implementations will offer the potential to make this system more correctly deployable to different smart cities, industrial IoT or remote environmental monitoring applications, while also maintaining efficiency, scalability and respect for user privacy.

REFERENCES

1. eMusic, *Online music store and streaming service*, Available: https://www.emusic.com/, accessed: Oct. 5, 2024.
2. iHeart, *Radio, podcasts, and music streaming platform*, Available: https://www.iheart.com/, accessed: Aug. 10, 2024.
3. Y. Arifin, T. G. Sastria, and E. Barlian, *User experience metric for augmented reality application: a review*, Procedia Computer Science, vol. 135, pp. 648–656, 2018. doi: 10.1016/j.procs.2018.08.221.
4. A. A. Laghari, H. He, S. Karim, H. A. Shah, and N. K. Karn, *Quality of experience assessment of video quality in social clouds*, Wireless Communications and Mobile Computing, 2017.
5. L. Zhang, H. Dong, and A. El Saddik, *Towards a QoE Model to Evaluate Holographic Augmented Reality Devices: A HoloLens Case Study*, IEEE MultiMedia, 2018.
6. A. A. Laghari, H. He, M. Shafiq, and A. Khan, *Assessment of quality of experience (QoE) of image compression in social cloud computing*, Multiagent and Grid Systems, vol. 14, no. 2, pp. 125–143, 2018.

7. K. Tsioutas, G. Xylomenos, and I. Doumanis, *Aretousa: A Competitive Audio Streaming Software for Network Music Performance*, in Audio Engineering Society Convention 146, 2019.
8. A. A. Laghari, H. He, R. A. Laghari, A. Khan, and R. Yadav, *Cache Performance Optimization of QoC Framework*, EAI Endorsed Transactions on Scalable Information Systems, vol. 19, no. 20, 2019.
9. A. A. Laghari, H. He, A. Khan, N. Kumar, and R. Kharel, *Quality of experience framework for cloud computing (QoC)*, IEEE Access, vol. 6, pp. 64876–64890, 2018.
10. K. U. R. Laghari and K. Connelly, *Toward total quality of experience: A QoE model in a communication ecosystem*, IEEE Communications Magazine, vol. 50, no. 4, pp. 58–65, 2012.
11. D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, *An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments*, in Proc. 8th Int. Conf. on Quality of Multimedia Experience (QoMEX), pp. 1–6, 2016.
12. A. A. Laghari, H. He, M. Shafiq, and A. Khan, *Impact of storage of mobile on quality of experience (QoE) at user level accessing cloud*, in Proc. 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), pp. 1402–1409.
13. S. Dey, M. Sahidullah, and G. Saha, *An overview of Indian spoken language recognition from machine learning perspective*, ACM Transactions on Asian and Low-Resource Language Information Processing, 2022. doi: 10.1145/3523179.
14. S. Yadav, A. Kumar, A. Yaduvanshi, and P. Meena, *A review of feature extraction and classification techniques in speech recognition*, SN Computer Science, vol. 4, no. 6, 2023. doi: 10.1007/s42979-023-02158-5.
15. X. Wang, Y. Long, Y. Li, and H. Wei, *Multi-pass training and cross-information fusion for low-response end-to-end accented speech recognition*, in Proc. INTERSPEECH 2023, pp. 2923–2927, 2023. doi: 10.21437/Interspeech.2023-142.
16. A. Becerra, J. I. D. I. Rosa, E. d. J. Velasquez, and G. Zepeda, *Portable student attendance management module for university environment by using biometric mechanisms*, Multimedia Tools and Applications, 2023. doi: 10.1007/s11042-023-15482-y.
17. F. Toosy and M. S. Ehsan, *BAQ and QoE: Subjective Assessment of 3D Audio on Mobile Phones*, in Audio Engineering Society Convention 146, 2019.
18. A. Ragano, E. Benetos, and A. Hines, *Adapting the Quality of Experience Framework for Audio Archive Evaluation*, in 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–3. IEEE.
19. L. Bartel and A. Mosabbir, *Possible Mechanisms for the effects of sound vibration on human health*, Healthcare, vol. 9, no. 5, pp. 597, 2021.
20. C. H. Hansen and K. L. Hansen, *Noise control: from concept to application*, CRC Press, 2021.
21. A. Alfayly, I. H. Mkwawa, L. Sun, and E. Ifeachor, *QoE-driven LTE downlink scheduling for VoIP application*, in Proc. 12th IEEE Consumer Communications and Networking Conference (CCNC), pp. 603–604, 2015.
22. C. C. Wu, K. T. Chen, C. Y. Huang, and C. L. Lei, *An empirical evaluation of VoIP playout buffer dimensioning in Skype, Google Talk, and MSN Messenger*, in Proc. 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video, pp. 97–102, ACM, 2009.
23. S. Tasaka and H. Yoshimi, *Enhancement of QoE in audio-video IP transmission by utilizing tradeoff between spatial and temporal quality for video packet loss*, in Proc. IEEE GLOBECOM, pp. 1–6, 2008.
24. N. K. Karn, H. Zhang, F. Jiang, R. Yadav, and A. A. Laghari, *Measuring bandwidth and buffer occupancy to improve the QoE of HTTP adaptive streaming*, Signal, Image and Video Processing, pp. 1–9, 2019.
25. L. Zhu, L. Chen, D. Zhao, J. Zhou, and W. Zhang, *Emotion recognition from Chinese speech for smart affective services using a combination of SVM and DBN*, Sensors, vol. 17, no. 7, 2017.
26. J. E. Noriega-Linares and J. M. Navarro Ruiz, *On the application of the Raspberry Pi as an advanced acoustic sensor network for noise monitoring*, Electronics, vol. 5, no. 4, 2016.
27. A. Alasadi, T. H. H. Aldhyani, R. R. Deshmukh, and A. H. Alahmadi, *Efficient Feature Extraction Algorithms to Develop an Arabic Speech Recognition System*, Engineering, Technology and Applied Science Research, vol. 10, no. 2, pp. 5547–5553, 2020. doi: 10.48084/etasr.3465.
28. J. Lee, M. F. B. Abbas, C. K. Seow, Q. Ceo, K. P. Yar, S. L. Keoh, and I. Mcloughlin, *Non-Verbal Auditory Aspects of Human-Service Robot Interaction*, in Proc. 15th IEEE Int. Conf. on Service Operations and Logistics, and Informatics (SOLI), Singapore, 2021. doi: 10.1109/SOLI54607.2021.9672366.
29. A. Glowacz, *Diagnostics of rotor damages of three-phase induction motors using acoustic signals and SMOFS-20-EXPANDED*, Archives of Acoustics, vol. 41, no. 3, pp. 507–515, 2016.
30. A. Glowacz, *Fault diagnosis of single-phase induction motor based on acoustic signals*, Mechanical Systems and Signal Processing, vol. 117, pp. 65–80, 2019.
31. M. Kunicki and A. Cichon, *Application of a phase resolved partial discharge pattern analysis for acoustic emission method in high voltage insulation systems diagnostics*, Archives of Acoustics, vol. 43, no. 2, pp. 235–243, 2018. doi: 10.24425/122371.
32. D. Mika and J. Jozwik, *Advanced time-frequency representation in voice signal analysis*, Advances in Science and Technology Research Journal, vol. 12, no. 1, pp. 251–259, 2018.
33. L. Zou, Y. Guo, H. Liu, L. Zhang, and T. Zhao, *A method of abnormal states detection based on adaptive extraction of transformer vibroacoustic signals*, Energies, vol. 10, no. 12, 2017.
34. T. Padhi, M. Chandra, A. Kar, and M. N. S. Swamy, *Design and analysis of an improved hybrid active noise control system*, Applied Acoustics, vol. 127, pp. 260–269, 2017.
35. S. C. Lee, J. F. Wang, and M. H. Chen, *Threshold-based noise detection and reduction for automatic speech recognition system in human-robot interactions*, Sensors, vol. 18, no. 7, Article ID 2068, 2018.
36. J. Ni, F. Duan, and J. Cao, *Prescribed-time distributed observer based practical predefined-time leader-follower output consensus of second-order multiagent system with communication noises*, Information Sciences, vol. 634, pp. 271–289, 2023. doi: 10.1016/j.ins.2023.03.116.
37. J. W. Hung, J. S. Lin, and P. J. Wu, *Employing robust principal component analysis for noise-robust speech feature extraction in automatic speech recognition with the structure of a deep neural network*, Applied System Innovation, vol. 1, no. 3, Article ID 28, 2018.

38. K. Feroze and A. R. Maud, *Sound event detection in real life audio using perceptual linear predictive feature with neural network*, in Proc. 15th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, pp. 377–382, 2018. doi: 10.1109/IBCAST.2018.8312252.
39. M. Ehsan and E. I. Abbas, *Isolated word recognition based on PNCC with different classifiers in a noisy environment*, Applied Acoustics, vol. 195, no. 7, pp. 108848, 2022. doi: 10.1016/j.apacoust.2022.108848.
40. A. Kaur and A. Singh, *Power-Normalized Cepstral Coefficients (PNCC) for Punjabi automatic speech recognition using phone based modelling in HTK*, in Proc. 2nd Int. Conf. on Applied and Theoretical Computing and Communication Technology, Bangalore, India, 2016.
41. H. A. Silva, *Comparison of techniques with speech enhancement and nonlinear rectification for robust speaker identification*, in Proc. 3rd Int. Conf. on Electronics, Communications and Information Technology (CECIT), China, pp. 170–175, 2022. doi: 10.1109/CECIT58139.2022.00038.
42. A. Alasadi, R. R. Deshmukh, and S. D. Waghmare, *Review of ModGDF & PNCC technique for feature extraction in speech recognition*, in Proc. IEEE Int. Conf. on Electrical, Computer and Communication Technologies (ICECCT), Tamil Nadu, India, 2019.
43. P. Borah and D. Gupta, *Review: Support Vector Machines in Pattern Recognition*, International Journal of Engineering and Technology, vol. 9, no. 3S, pp. 43–48, 2017. doi: 10.21817/ijet/2017/v9i3/170903S08.
44. U. Ghosh and U. K. Mondal, *Improved wireless acoustic sensor network for analysing audio properties*, International Journal of Information Technology, vol. 15, no. 5, pp. 3679–3687, 2023. doi: 10.1007/s41870-023-01411-7.
45. F. Bu, *An exploration of calculating the packet loss rate by using the block rate*, Advances in Computer Science Research, in Proc. 3rd Int. Conf. on Communications, Information Management and Network Security (CIMNS 2018), vol. 65, pp. 147–149.
46. G. Rajan and V. Thiagarajan, *Signal-to-noise ratio estimation techniques for wireless communication systems: A survey*, Wireless Personal Communications, vol. 71, no. 3, pp. 1987–2011, 2013. doi: 10.1007/s11277-012-0931-4.
47. Q. Huynh-Thu and M. Ghanbari, *Scope of validity of PSNR in image/video quality assessment*, Electronics Letters, vol. 44, no. 13, pp. 800–801, 2008. doi: 10.1049/el:20080522.
48. J. Shen, Y. Tang, and W. Pan, *A new fidelity metric for evaluating signal reconstruction performance in communication systems*, IEEE Communications Letters, vol. 20, no. 5, pp. 930–933, 2016. doi: 10.1109/LCOMM.2016.2526074.
49. Y. Ruan, Y. Xiao, Z. Hao, and B. Liu, *A convex model for support vector distance metric learning*, IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 8, pp. 3533–3546, 2022. doi: 10.1109/TNNLS.2021.3053266.
50. G. Nathan, F. Britten, and J. Burnett, *Sound Intensity Levels of Volume Settings on Cardiovascular Entertainment Systems in a University Wellness Center*, Recreational Sports Journal, vol. 41, pp. 20–26, 2017.
51. C. Zhang and M. Dong, *An improved speech endpoint detection based on adaptive sub-band selection spectral variance*, in Proc. 35th Chinese Control Conference (CCC), pp. 5033–5037, Chengdu, China, 2016.
52. V. Rosao and A. Aguileira, *Method to calculate LAFmax noise map from LAeq noise maps, for roads and railways*, in Proc. INTER-NOISE and NOISE-CON Congress and Conference Proceedings, vol. 22, pp. 5337–5345, 2019.
53. M. Huang, L. Chen, and Y. Zhang, *A spatio-temporal noise map completion method based on crowd-sensing*, Environmental Pollution, vol. 274, 115703, 2021. doi: 10.1016/j.envpol.2020.115703.
54. C. Zhang and M. Dong, *An improved speech endpoint detection based on adaptive sub-band selection spectral variance*, in Proc. 35th Chinese Control Conference (CCC), pp. 5033–5037, Chengdu, China, 2016.
55. Y. Zhang, K. Wang, and B. Yan, *Speech endpoint detection algorithm with low signal-to-noise based on improved conventional spectral entropy*, in Proc. 12th World Congress on Intelligent Control and Automation (WCICA), pp. 3307–3311, Guilin, China, 2016.
56. K. Sayood, *Introduction to Data Compression*, 3rd ed., Katey Bircher: Cambridge, MA, USA, pp. 47–53, 2017.
57. M. Magno, F. Vultier, and B. Szebedy, *Long-term monitoring of small-sized birds using a miniaturized Bluetooth-low-energy sensor node*, in Proc. 2017 IEEE SENSORS, pp. 1–3, Glasgow, UK, 2017.
58. S. Aggarwal, V. Gurusamy, S. Sethuramalingam, B. Pant, K. Kaur, A. Verma, and G. N. Binegde, *Audio Segmentation Techniques and Applications Based on Deep Learning*, Scientific Programming, vol. 6, pp. 1–9, 2022. doi: 10.1155/2022/7994191.
59. S. Iwata and R. Kitani, *Phase-resolved partial discharge analysis of different types of electrode systems using machine learning classification*, Electrical Engineering, vol. 103, pp. 3189–3199, 2021. doi: 10.1007/s00202-021-01306-5.
60. Y. Yang, Z. Peng, W. Zhang, and G. Meng, *Parameterised time-frequency analysis methods and their engineering applications: A review of recent advances*, Mechanical Systems and Signal Processing, vol. 119, pp. 182–221, 2019. doi: 10.1016/j.ymssp.2018.07.039.
61. A. Schilling, R. Gerum, C. Metzner, A. Maier, and P. Krauss, *Intrinsic noise improves speech recognition in a computational model of the auditory pathway*, Frontiers in Neuroscience, vol. 16, 908330, 2022. doi: 10.3389/fnins.2022.908330.
62. M. Sigmund, *Speaker Discrimination Using Long-Term Spectrum of Speech*, Information Technology and Control, vol. 48, no. 3, pp. 446–453, 2019. doi: 10.5755/j01.itc.48.3.21248.
63. A. A. Laghari, R. A. Laghari, A. A. Wagan, and A. I. Umrani, *Effect of Packet Loss and Reorder on Quality of Audio Streaming*, EAI Endorsed Transactions on Scalable Information Systems, Online First, 2019. doi: 10.4108/eai.13-7-2018.160390.