NTU Journal of Engineering and Technology (ICSDT2025) special issue
DOI: https://doi.org/10.56286/ntujet.icsdt2025

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Challenges and Opportunities in Assistive Audio Recognition Technologies for Deaf and Hard-of-Hearing (DHH)

Israa Mohammed Ibrahim[1], Zaid H. Alsawaff [1], Fadwa Al Azzo[2,3]
[1]Department of Medical Instrumentation Techniques Engineering, Technical Engineering Col-lege, Mosul, Northern Technical University, Mosul, Iraq
[2]Center of Technical Research, Northern Technical University, Mosul, Iraq
[3]Technical Engineering College for Computer and AI, Northern Technical University, Mosul, Iraq
esraamohammed89@ntu.edu.iq , zaidalsawaff@ntu.edu.iq , fadwaalezzo@ntu.edu.iq .

Article Informations

Received: 15-03- 2025,
Accepted: 28-04-2025,
Published online: 01-03-2026

Corresponding author:
Name: Israa Mohammed Ibrahim
Affiliation : Northern Technical University
Email:
esraamohammed89@ntu.edu

Key Words:
Deaf,
Hard-Hearing-Impairment,
assistive techniques ,
wearable devices,
speaker dentification.

ABSTRACT

The advancements in assistive audio recognition technologies have progressed rapidly and have improved the accessibility for the Deaf and Hard-of-Hearing (DHH) community. This paper provides an extensive review of the current methods, including traditional techniques such as GMMs and HMMs, and modern deep learning-based techniques as Convolutional Neural Networks (CNN) and EfficientNet. The strengths and limitations of these methods are compared with respect to accuracy, computational efficiency, and noise robustness. Although these deep learning models have higher recognition rates, they impose significant computational requirements thereby restricting their usage in real-time and low-power devices. The review also revealed research gaps with a strong emphasis on energy-aware neural networks and their potential applications in smart environments, adaption strategies, and integration of IoT technologies for practical implementation. As deep learning models grow in complexity and the required amount of labeled data increased to unsustainable level, future research will need to explore hybrid models that balance between performance gains and efficiency, to ensure that assistive audio recognition systems become increasingly reliable, usable, and generalizable in a broad range of acoustic environments.

1. Introduction

One fundamental physical phenomenon that has a big impact on our daily lives is sound. Periodic changes in the medium's thickness and pressure are caused by things vibrating and sending wave signals through a variety of media, including solids, liquids, and gases [1]. Moreover, sound is considered very important for communication and human interaction since it contributes directly to thoughts and feelings transferred by words and non-verbal sounds [2]. Each person has a unique voice that refers to the speaker's characteristics, including age, gender, and ethnicity. These features give the ability to recognize the person through their voice due to the physiological variation (like vocal tract shapes, and larynx sizes) and behavioral variation (like speaking manner, pronunciation style, and vocabulary selection) [3]. [4].

The auditory system is divided into two main sections: the peripheral system, which consists of external, middle, and inner ear, as shown in Figure 1, and the central auditory system, which includes the auditory nerve and brain regions responsible for sound processing. The peripheral system captured sound waves by the outer ear and traveled through the auditory canal to the middle ear, where the vibrations transmitted to the inner ear. The cochlea converts vibrations into electrical signals, that travel through the auditory nerve to the central auditory system for sound analysis and recognition [5].

A normal hearing threshold range between 0 and 25 decibels (dB). When the threshold is exceeded, it gives an indicator of an auditory disorder that is considered a hearing impairment when recognized sound intensity reaches above 41 dB [7]. Hearing impairment is considered a growing global health problem. According to the World Health Organization, 5% of the world's population suffers from deafness, which means 430 million people [8].

Hearing loss negatively affects several aspects of patient life, like communication, speech, and language development, it also affects cognitive abilities and education, leading to psychological, social, and behavioral problems [9].

Assistive devices played a role in improving the quality of hearing-impaired people's lives by enhancing social interactions and community engagement by facilitating communication, as well as making progress in education and independence for this group [10]. The design of these devices is supported by embedded systems that control their functions and guarantee their instant response considering their low cost, ease of programming, and customization by open software. Moreover, these platforms support many sensors due to their flexibility and their diversity [11].

Unique voice features and characteristics can be utilized to develop applications and assistive devices like speaker recognition systems by comparing the voice biometrics of the speech with models prestored [12]. These devices analyzing vocal characteristics such as tone, spectral magnitudes, and formant frequencies, simulating the human hearing system distinguish between sounds. enhancing speaker recognition systems that identify individuals based on voice patterns [13].

Speaker recognition efforts started in 1950, relying on human experts who analyze individual speech patterns. With efforts and advancements over time, automated speaker recognition systems have developed and become valuable in recent decades, as they are the only remotely testable biometric via the telephone network, which makes speaker recognition more popular in variable applications. Among them is the assistive system [14].

Additionally, the progress of artificial intelligence and machine learning (ML) has enhanced the efficiency of speaker recognition assistive systems [15].

These systems are composed of three units. Starting with the preprocessing unit that removes any noise and enhances clarity and efficiency. Following the feature extraction unit, where unique characteristics are extracted from the processed signal by specific techniques that recognize the human voice and store it as a feature vector in the database, like Mel-Frequency Cepstral Coefficients (MFCC) allows these systems to analyze speech patterns similarly to how the human ear processes sound frequencies, enabling accurate speaker identification [16]. Finally, the classification unit involved model evaluation using features of new data to be compared with the reference features and take the decision [17].

This stage may employ varying classification techniques to distinguish speakers. Among these techniques is the Gaussian Mixture Model (GMM), which is based on statistical analysis and effectively recognizes voices with limited sources.

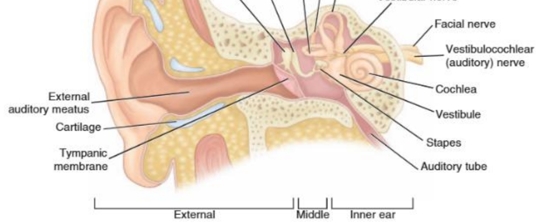

Fig. 1. Anatomical structure of the auditory system, illustrating the external, mid-dle, and inner ear [6]

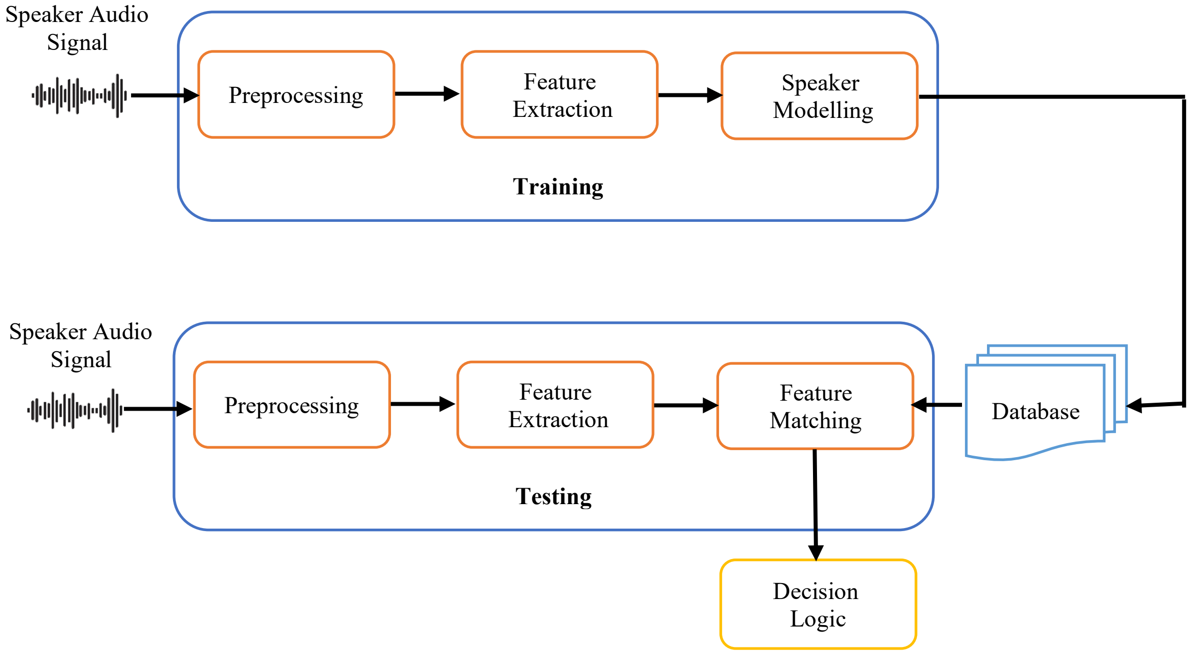

Fig. 1. Block diagram of speaker recognition system

Another technique is the Hidden Markov Model (HMM), which relies on temporal sequences, which can handle longer audio recordings and distinguish between overlapping voices. Also, a support vector machine (SVM) identifies the ideal border between speaker classes, which is effectively precise for limited data. Another technique that combines generative and discriminative approaches is hybrid models, which give more accurate results in complex audio environments [18].

For handling large audio data with high complexity, Deep Neural Networks (DNN) provide high-accuracy identification for speakers, although in a noisy environment [19]. These units must pass through two main stages, training and testing [20] as depicted in figure 2.

In training stage, the audio signals are collected and stored as reference models after feature extraction.as for testing stage, the new audio signals are captured and their features are compared with the saved models for recognition decision [21].

In addition, Streaming database systems have been proposed as a method to improve communication for the DHH community. These systems enable real-time data processing and facilitate interactions, ensuring seamless communication through instant data retrieval and translation between different modalities, such as text and sign language [22]. However, effective communication for the DHH community is not solely dependent on data processing; seamless integration between smart devices is also crucial. This is where the Internet of Things (IoT) plays a transformative role. The advancements in IoT technology provide a connected environment that enables assistive devices to exchange and analyze information in real time, which supports the intelligent environment to enhance the independence and safety of DHH and improve their lives [23].

2. Review Process

This review examines assistive devices designed for DHH, concentrating on sound and speaker identification techniques. To ensure alignment with recent advancements, studies were selected based on research published within the last ten years, sourced from IEEE Xplore, ScienceDirect, and SpringerLink databases, due to their inclusion of high impact scientific journals and conferences in engineering, artificial intelligence, and assistive technologies.

Twenty-five studies were reviewed, exploring various assistive devices, including wearable technologies, along with fixed and portable solutions. Additionally, these studies covered different sound recognition methods, ranging from basic acoustic features extraction, to advanced statistical methods, ML algorithm, and DL techniques.

This review is a crucial step in developing a new assistive device to enhance speaker recognition for DHH. By analyzing of current technologies, this study highlights the latest development in this field, covering acoustic feature extraction techniques, classification methods, and technologies utilized for system designed. Additionally, to addressing the limitation and challenges that face these systems, providing comprehensive overview that could support feature improvement and research.

3. Related Studies

Hearing-impaired assistive devices have witnessed significant advancements, as numerous studies confirmed that discussed inventions with diverse techniques in this field.

Fernandes (2015) presented a speaker identification system that facilitates the fa-cilitation of real-time group conversation transcription, especially for DHH. The system depends on two methods for recognizing the speakers. First, it used MFCCs for voice feature extraction, which is more popular in voice recognition systems, and analyzed the unique audio characteristics of each speaker by ML. Second, the system used Sound Source Localization (SSL) to specify the speaker location, especially in loud environments. The prototype designed used a KL25Z board that analyzed the captured sound by the microphones, while the user interface (UI) displays the speaker identification on a connected computer. Additionally, to SSL by lighting LEDs in the exact direction. The system achieved 80% accuracy in speaker recognition [24].

Kugler et al. (2016) focused on real-time environmental sound recognition with field learning abilities in a system of DHH-aided devices. The system was developed on a Field Programmable Gate Arrays (FPGA) platform that can discover new sounds and add them to the database. The picked-up sound via MEMS microphone was pre-processed through band-pass filters, hair-cell compression, and a spike generator. Furthermore, the extracted sound features were classified by the K-Nearest Neighbor (KNN) algorithm. Although the system scored a high accuracy level of 95%, it reduced to 82.2% in a loud environment of 6 dB[25].

Ando et al. (2017) also proposed a sound recognition system for DHH people based on ML. The system consists of two programs: a learning and analysis program for the Raspberry Pi and a notification program for smartphones. The sound waves were captured using an external USB microphone and saved in WAV form, and the system applies the Fourier transform and correlation coefficient for feature extraction to be analyzed by the Python libraries (Numpy and Scipy) for sound matching, while the notifications are sent to the smartphone. The system accomplished an accuracy of 100% for the interphone sound, 90% for the clapping sound, and 80% for the door sound [26].

Saifan et al. (2018) likewise introduced another way to identify sounds and words in the alert system for DHH. The system contains two main engines, one of which recognizes the cautionary words by using the Soundex algorithm, which achieved 85% accuracy and improved to 100% by repeating words, while three classification algorithms were tested for the sound recognition engine: KNN, Decision Tree, and Neural Network (NN), where the last was more specific by achieving the highest accuracy of 91.7%. NN utilized triple-layer neural networks to classify the sound features that were extracted by using FFT and Praat software. The system sends sensory alerts like images and vibrations to the mobile user [27].

Yağanoğlu and Köse (2018) analyzed real-time sound recognition using vibrational wearable devices for DHH. The device is accomplished in Raspberry Pi real-time sound analysis and sending vibrational alerts with different patterns while successfully distinguishing important sounds. Despite the high achieved accuracy of 98% on the computer, it dropped to 94% in noisy environments. The audio fingerprinting method that is utilized in this system involves analyzing the sound temporal frequencies that were converted into a spectrogram, picking the peak points, and creating a unique sound fingerprint [28].

In the same year, Fanzeres et al. introduced an environmental sound recognition system as a mobile app for DHH. The system features a customizable sound database, making it practical and flexible. Starting with sound analysis relied on MFCC and Linear Predictive Coding (LPC) features for spectral and temporal property analysis, additionally for spectral features that enhance the classification accuracy. While classification sounds based on several ML algorithms, the Nearest Neighbor algorithm achieves the highest accuracy among them at 92.7% [29].

Anegundi et al. (2019) proposed two selections in a computer speech recognition application that enables people with hearing disabilities to recognize speech and speakers through near and far scenarios. To detect the speaker in the near scenario, the GMM was used to analyze the voice and extract the MFCC after fragmentation and convert it into FD, while in the far scenario, each speaker has a specific IP that can be recognized with it. The system accuracy reached 92%; nevertheless, the error rate increased as the number of speakers increased due to the similarity and interference between the voices [30].

Dhakal et al. (2019) studied spontaneous real-time pipelined speaker recognition. Gabor filter and CNN were utilized for the feature's extraction. Among several classifiers tested (SVM, DNN), the Random Forest (RF) enhances the system performance and increases its accuracy in speaker identifying where it was the most specific, sensi-tive, and accurate by achieving 94.87% accuracy [31].

In the context of Active speaker recognition development, Tapu et al. (2019) pre-sented an innovative system called DEEP-HEAR, that enhance DHH experience through tracking translation location dynamically with the active speaker. the system based on analyzing the audio spectrogram that extract via several techniques, where the short-time Fourier transform (STFT) and the residual-network (ResNet-34) were utilized for analyzing and classifying the audio signals, which assist in speaker recognition despite background noise or active speaker is not visible in scene. This study relied on in merge the audio signals with visual and textual data, which improve the recognition accuracy that reached rate of over 90%. However, the system faces challenges in handling real world noise while using only the audio stream for speaker recognition drops accuracy in to 63% [32].

Jain et al. (2020) moreover introduced two CNN-based deep learning models of sound identification for deaf and HI awareness systems [33][34]. The home awareness system for DHH operates on displays and smartwatches [33]. The trained Visual Geometry Group (VGG16) classification model accuracy reached 85%. while the intelligent smart watch application's average accuracy didn't exceed 81% and generally outperformed for certain voices [34]. The system running on either a watch, phone, or cloud-depending (VGG-Lite) as a classification model had the best performance among the other four models. Both systems provide the user with visual notifi-cations, including voice identity, loudness, and timestamps, along with vibration.

During the same year, Rosa Tavares and Victória Barbosa aimed to develop an IOT-based Apollo Sign Sound intelligent system for DHH. The system employed ESP8266 to process the collected sounds by using FFT. The system utilized the server for training the received data by a multi-layer perceptron neural network (MLPNN) with a sigmoid activation function to reduce the error. The LIBRAS-based application pushes customizable notifications on the user's mobile that can be via visualization, illumination, vibration, SMS, and email. The system's accuracy in detecting sounds achieved an F-score of 0.73 [35].

Furthermore, Young et al. (2020) developed assistive smart glasses for DHH using deep learning techniques. The system used MFCC characteristics as an input for the DNN and the Inception-v4 model for sound classification. Real-time speech and sound recognition were made possible by the IOT base system integrated into the ESP-8266 platform. Accuracy in identifying sounds, even in noisy environments, was 92% in the lab and 82% in the real world. The results were displayed on a tiny screen that was mounted on the glasses [36].

In similar efforts implemented to improve the sound awareness of DHH via smart glasses, Rahman et al. (2020) presented system focusing on real time localization of sounds and notifying users using visual and tactile manner. The system utilizes Dynamic Binaural Cues technique for precise sound source direction, integration Raspberry Pi 4B for high performance with low cost. Net of five microphones were used for analyzing the audio signals, supported with Active Noise Cancellation (ANC) technique, classifying important sounds like names, greetings (e.g., "Hello"), and warning sounds (e.g., car horns, dog barks) using template matching. The sound type and source direction displayed on embedded screen with vibration alert. The system is based on IOT for settings and sound update, and its work efficiently within the range of 6 meters. Although there are challenges with low-frequency or weak sounds, overlapping noises, and complex sound recognition.[37]

Goodman et al. (2021) discussed personal voice recognition system development for DHH, aimed at raising their awareness of environmental sounds through customizing ML models. The system relies on the recording users for voices with personal importance, such as home alerts or pets sounds, by using smartphone application like Rev Recorder. Interactive Machine Learning (IML) models are then trained, as Google Teachable Machine, in order to classify these sounds according to user recording. Studies revealed the effectiveness of the system in enabling the users customizing their audio experience, despite the challenges in recording unexpectable sounds or evaluation its quality by inexpert users. Moreover, dependence on the user for data collection, limits the diversity and accuracy of the trained models [38].

In the same year, Albishi et al. presented an intelligent emergency application that facilitates deaf and DHH connections with emergency services. The application is practical and easy to use by adopting the TeleTYpe technology (TTY) with Norman's interaction smart model. It enables direct texting within telephone networks. Moreover, it identifies the patient's location by Global Positioning System GPS and sends it to the emergency service [39].

Shimoyama (2021) proposed a DHH assistive device that tracks the sound sources by multi-directional vibrations. The system designed on an FPGA module that analyzed the frequency of the signal detected by the ear microphone relies on measuring the difference in sound arrival time between the ears to identify the source direction, while the distance evaluation is done by determining the sound intensity. as well It identifies the direction by sending vibration in the detected source direction with different intensity according to the source distance [40].

Mohamed et al. (2022) studied the audio signal processing for speaker recognition. MFCCs are considered the most important features that improve the clarity and quality of audio signals affected by noise. Speaker recognition techniques include Hidden Markov Models (HMM), Support Vector Machines (SVM), (DNN), and Gaussian Mixture Models (GMM), where the most common model in speaker recognition systems combines a GMM with a universal background model (UBM) while using it with SVM enhances the speaker identification accuracy at a rate of 70–75% between ten suspects [41].

During the same year, Jain et al. proposed a ProtoSound system that can distinguish the sound samples relying on prototypical networks, which are based on few shot learning, enabling the model to be trained using only five records for each sound category. The inputs of this system employed Log-Mel Spectrograms (LMS) features. For classification, Nearest Neighbor identified the sound based on Euclidean distance, while unfamiliar sounds were handled by Open-Set Classification by evaluating relative distances. The system achieves 88.9% accuracy as a real-time training and operation mobile app, making it the perfect awareness application for DHH [42].

Aiyengar et al. (2023) also introduced utilizing MFCC features that simulate the ear frequency response with GMM-HMM for sound classification in the DHH assistive system. The device that was implemented using Raspberry Pi 4 gives tactile notification represented by a specific pattern of vibrations. The system achieved 79% accuracy where the model performance ranged between good in some sounds like clapping and knocking and faced challenges for others like barking and laughing [43].

Buhat et al. (2023) designed a DNN-based assistive device for DHH in sound identification and classification. The wristband consists of a headband implemented on a Raspberry Pi 3B with electret microphones to analyze the collected audio signal. Also, the wristwatch used a NodeMCU ESP 8266 microprocessor that notified the user by specifying a vibration pattern, additionally to distinct color through an SMD RGB LED. The connection between the systems depends on the web service, where the classification is accomplished by yet another Audio Mobilenet Network (YAMNet) model. The obtained system accuracy was 70% for internal tests, while it dropped in external tests to 63.33% [44].

Within the same year, Chin et al. focused on developing wearable awareness devices that assist DHH in recognizing emergent road sounds while walking or driving. The system uses an Arduino Nano 33 BLE Sense, the system adopted to use short-time Fourier transform (STFT) to convert detected sounds into spectrograms that are analyzed through a pre-trained EfficientNet-based ensemble model with CREMA-D audio data. The device recognizes the emergent sounds and rapidly alerts the user through vibrations and OLED screen messages. The system shows significant performance, where the accuracy reached 97.1% in offline computing and 95.2% in edge computing, as it turns out this model surpassed the accuracy of IOS devices with 26.23% [45].

Goyal and Basavarajappa (2023) likewise presented a wearable assistive device based on IOT that enables DHH to communicate with their family using the Switch Module and Vibration Module implemented on the Bluno Beetle microcontroller. Each family member has a switch module that activates the vibration module in the wearable device in a specific pattern for each switch via Bluetooth Low Energy (BLE). The device reflects its simplicity of use, energy efficiency, and low cost [46].

Tharwat et al. (2024) also produced wearable assistive devices that enhance communication with DHH through sound and speech recognition. The device was designed on Raspberry Pi 4 using MFCC to extract sound features to be analyzed and classified by an RF algorithm that recognizes the speaker and displays their identity on an OLED screen. The system achieved an accuracy of up to 93.37% in speaker recognition [47].

Abhiram et al. (2024) presented an assistive system that improves the safety and independence of DHH drivers by alerting them about emergency vehicles. The system is composed of a transmitter and receiver; the transmitter is represented by a USB microphone and a Raspberry Pi 4, where the sounds are analyzed. The picked-up sounds are processed by MFCC for feature extraction and CNN recognizes the sound and sends the result via local WiFi to the receiver, which is represented by the driver glove, where ESP32 sends a visual notification on the OLED screen and activates the vibrator motors on the fingertips. Even though the system shows promising results, it has a finite number of data points [48].

Table 1. Comparative overview of sound recognition approaches.

Model	Accuracy	Processing Speed	Noise Robustness	Computational Complexity
GMM-UBM + SVM	70–75%	Medium	Low	Low
HMM	80–85%	High	Medium	Medium
KNN / Decision Tree	85–90%	Medium	Low	Low
CNN	90–97%	Low	Very High	Very High
Efficient Net	95–97.1%	Low	Very High	Very High

4. Results and Discussion

Wearable assistive devices of the DHH have developed and improved simultaneously with both sound recognition and notification systems: all those technologies address various challenges in communication, alerting, and education; among them, it is possible to point at sound recognition as one of the most influential.

4.1. Comparative analysis of exisiting technologies

according to earlier research, sound recognition systems performance, affected by used model, the quality of training data, and adaptability to noisy environments. These systems are generally categorized into two models:

Traditional Models: Methods such as GMM, HMM, and KNN rely on statistical analysis and probabilistic modeling. These models are computationally efficient but highly sensitive to noise.

Deep Learning Models: Including CNNs and EfficientNet, these models offer high accuracy and robustness in noisy environments but require substantial computational resources. Table 1 provides a comparative overview of these approaches.

4.2. Limitations of traditional models

Traditional models as GMM and HMM relies on statistical characteristics, their performance is limited in complicated noisy situations. While techniques like KNN and Decision Tree perform well with clean data, their accuracy significantly drops under real-world noise. These models also lack the ability to extract hierarchical or complex features from audio inputs, making them less effective in dynamic settings.

4.3. Strengths and weaknesses of deep learning models

The CNN and EfficientNet models have proven to be very resilient to noisy environments, with accuracy rates of up to 95.2% in real-world scenarios and 97% in controlled settings. Nevertheless, these models are computationally and might not be appropriate for wearables with low power consumption.

In an effort to improve recognition, hybrid systems like the DEEP-HEAR architecture integrated textual, visual, and auditory data. Still, recognition accuracy declined to 63% when relying solely on audio in noisy environments.

4.4. Research gaps and future directions

Research Gaps and Future Directions Based on the analysis, several research directions are recommended:

- Improve the energy efficiency of deep learning models to make them viable for wearable applications.
- Develop hybrid approaches (e.g., CNN + HMM) to leverage the advantages of both traditional and modern techniques.
- Encourage self-learning systems that can adjust to changing conditions without requiring continual retraining.
- Enhance performance in outdoor settings with unpredictable noise, such as streets and industrial zones.

Table 1. The Comparisons between state-of art studies.

Ref. No.	Year	Method	Technology	Result	Notes
[24]	2015	MFCCs, SSL	KL527 board, mic, RGB LED	Accuracy 80%	Indicates sound direction
[25]	2016	Band-pass filters, hair-cell compression, spike generator, K-NN	FPGA, MEMS mic, LED	Accuracy reduced from 95% to 82.2%	Affected by environment
[26]	2017	Fourier transform, correlation coefficient, Python libraries (Numpy and Scipy)	Raspberry Pi, USB Mic, smartphone	100% accuracy for interphone sound, 90% for clapping sound, 80% for door sound	Affected by the noise and quality.
[27]	2018	FFT, KNN, Decision Tree, NN	Mobile app, MATLAB and Praat software, smartphone mic	91.7% accuracy in sound recognition (NN performed best)	Sends alerts via images & vibrations
[28]	2018	Audio fingerprinting	Raspberry Pi, USB mic, vibration motor	98% accuracy on computer, 94% in real-time	Affected by noise, Response time affected by distance.
[29]	2018	MFCC & LPC, Nearest Neighbor	Android devices, built-in mic	Accuracy 92.7%	Personalized databases

[30]	2019	MFCC, GMM	Google Cloud Speech API, client-server model	Accuracy 92% in near scenarios.	Error increases with similarity
[31]	2019	GF, CNN, RF, SVM, DNN.	Language Processing (NLP) integration	RF accuracy 94.87%	Affected by the noise
[32]	2019	Deep CNN (VGG16, ResNet), Multimodal Fusion	Software-based system, DEEP-HEAR Framework	Accuracy 90% by text, audio, and video, 63% by audio only.	challenges with real world noise
[33]	2020	VGG16 for home awareness system.	visual displays (Surface Pro 3), smartwatch	Accuracy 85.9%.	Provides visual notifications with vibration.
[34]	2020	Smartwatch app, VGG-Lite.	smartwatch, smartphone, or cloud	Accuracy 81.2%	Provides visual and vibrational alerts.
[35]	2020	MLPNN for sound classification.	ESP8266, MAX4466 sound sensor, LIBRAS app, mobile device	Detected environmental sounds, F-score of 0.73	Customizable notifications via visualization, illumination, vibration, SMS, and email.
[36]	2020	MFCC, Inception-v4	ESP-8266, smart glasses, small screen	Accuracy: 92% in labs, 82% in real-time	Affected by real environments
[37]	2020	Dynamic Binaural Cues, Template Matching, ANC	Raspberry Pi 4B, 5 mics, vibration motors, display	moderate accuracy recognized names, greetings, warning sounds	Struggled with low-frequency, overlapping noises, complex sound.
[38]	2021	Pre-trained Neural Networks (VGG, ResNet), IML	Smartphone, Rev Recorder & Teachable Machine	Customizable audio experience	Accuracy affected by lack of data diversity & unexpected sounds
[39]	2021	TTY-based intelligent app, Norman's model	Mobile phone, GPS	Supports texting & GPS alerts	Enhances emergency communication
[40]	2021	Tracks sound sources based on distance/direction estimation	FPGA modules, ear Mic, multi-directional vibration.	Localizing sound source.	vibrational alerts for direction and distance
[41]	2022	MFCC, GMM, SVM, UBM, DNN	Software-based system	Recognition accuracy rate of 70-75%	Hybrid models improve performance in noisy environments
[42]	2022	Using Prototypical Networks, LMS features, and Nearest Neighbor for classification	Mobile app.	Accuracy 88.9%	challenges in recording accuracy, and handling similar sound classes.
[43]	2023	MFCC for feature extraction, GMM-HMM for sound classification.	Raspberry Pi 4, Vibration Motor as a tactile notification device	The system accuracy in classifying common sounds reached 79%	struggles with background noise, lack of data, and sound overlap.
[44]	2023	YAMNet .	Raspberry Pi 3B, NodeMCU ESP8266, Electret Mic, SMD RGB LED, vibration	indoors accuracy 70.48%, 63.33% outdoors.	better performance indoors
[45]	2023	STFT, EfficientNet, fuzzy rank- ensemble.	Arduino Nano 33 BLE Sense, Efficient Net, fuzzy ensemble, CREMA-D data, OLED screen, vibrations	offline accuracy 97.1%, 95.2% edge.	Outperforming iOS by 26.23%.
[46]	2023	IoT-based Wearable device using Switch Module and Vibration Module	Bluno Beetle Microcontroller, BLE, vibration module.	Helping DHH family communication by wearable device	simple, low-cost, and energy-efficient, required multi-connection support.
[47]	2024	MFCC, Random Forest.	Raspberry Pi 4, OLED screen	Accuracy of 93.37 %	Displays speaker ID
[48]	2024	MFCC, CNN	Raspberry Pi 4, ESP32, OLED, Vibration motors	Effective emergency detection	Required more data, affected noisy environments

5. Conclusion

This review has discussed the progress and issues of assistive audio recognition technologies developed for the DHH community while centering around classical statistical models and contemporary neural networks. Traditional methods such as GMM and HMMs have been important for speaker and sound recognition but are limited in noise robustness and adaptation to different acoustic conditions. In contrast, deep learning methods outperformed traditional ones, especially on CNN applied and EfficientNet, where its recognition accuracy reached 97% in controlled conditions. Despite that, their usage is still limited from high computational cost and energy consumption perspective, thus making them unfeasible for wearable and real-time applications.

Whilst much progress has been made in this area, there are still many challenges that remain. Key challenges include enhancing the energy efficiency of the deep learning models involved, which is crucial for their deployment on low-power wearables. A different key challenge includes creating adaptive learning mechanisms with in which recognition systems can generalize well to a new yet different real-world setting without needing a retrain of their model. The convergence of artificial intelligence and IoT technologies can also provide an avenue for improving the ability of assistive devices to respond to users instantaneously and in context.

Future work can investigate the possibilities of hybrid models where such statistical strategies can be implemented with the deep learning models for speed versus accuracy. Developments such as multi-modal identification systems or noise reduction techniques might lead to further improvements in resilience. Addressing these gaps will be critical to enhancing the quality of life, safety, and communication of the DHH community by increasing the reliability, effectiveness, and accessibility of assistive audio recognition systems.

References

[1] W. A. Yost, Fundamentals of Hearing: An Introduction. San Diego: Academic Press (Elsevier)., 2021.

[2] B. A. Stach, "Introduction to Sound Acoustics for the Hearing and Speech Sciences Fourth Edition Editor-in-Chief for Audiology."

[3] R. Mohd Hanifa, K. Isa, and S. Mohamad, "A review on speaker recognition: Technology and challenges," Comput. Electr. Eng., vol. 90, no. January, p. 107005, 2021, doi: 10.1016/j.compeleceng.2021.107005.

[4] S. A. Imam, P. Bansal, and V. Singh, "Speaker recognition using automated systems," AGU Int. J. Eng. Technol., vol. 5, no. 5, pp. 31–39, 2017.

[5] Frank E. Musiek and J. A. Baran, The Auditory System Anatomy, Physiology, and Clinical Correlates, Second Edi. San Diego, CA, USA: Plural Publishing, Inc., 2020.

[6] D. J. Seikel, J. Anthony; Drumright, David G.; Hudock, Anatomy & Physiology for Speech, Language, and Hearing, 6th Editio. San Diego: Plural Publishing, 2023.

[7] G. G. . Celesia, G. Hickok, and P. . Afra, The human auditory system : fundamental organization and clinical disorders. Elsevier, 2015.

[8] W. H. Organization, "Deafness and hearing loss," 2024.

[9] A. K. S. Alramamneh, O. A. O. Sabayleh, S. M. A. A. Hazim, and S. M. A. Drei, "Psychological and social problems of hearing-impaired students and the adopted coping strategies in deaf schools," J. Educ. Soc. Res., vol. 10, no. 2, pp. 205–215, 2020, doi: 10.36941/jesr-2020-0039.

[10] W. A. Zeleke, J. Leoste, E. Smolianina, and A. Valencia-Arias, "Benefits and development of assistive technologies for Deaf people's communication: A systematic review."

[11] J. Á. Ariza and J. M. Pearce, "Low-Cost Assistive Technologies for Disabled People Using Open-Source Hardware and Software: A Systematic Literature Review," IEEE Access, vol. 10, pp. 124894–124927, 2022, doi: 10.1109/ACCESS.2022.3221449.

[12] S. Bunrit, T. Inkian, N. Kerdprasop, and K. Kerdprasop, "Text-independent speaker identification using deep learning model of convolution neural network," Int. J. Mach. Learn. Comput., vol. 9, no. 2, pp. 143–148, Apr. 2019, doi: 10.18178/ijmlc.2019.9.2.778.

[13] Z. Saquib, N. Salam, R. P. Nair, N. Pandey, and A. Joshi, "A Survey on Automatic Speaker Recognition Systems," in International Conference on Multimedia, Computer Graphics, and Broadcasting, Springer Berlin Heidelberg, 2010.

[14] H. Beigi, "Speaker Recognition," Encycl. Cryptogr. Secur. Priv., pp. 1–17, 2021, doi: 10.1007/978-3-642-27739-9_747-2.

[15] A. Irum and A. Salman, "Speaker verification using deep neural networks: A review," Int. J. Mach. Learn. Comput., vol. 9, no. 1, pp. 20–25, Feb. 2019, doi: 10.18178/ijmlc.2019.9.1.760.

[16] C. S. Kumar and P. M. Rao, "Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm," Int. J. Comput. Sci. Eng., vol. 3, no. 8, pp. 2942–2954, 2011.

[17] K. J. A. J. Sumit Srivastava, "Analysis of Human Voice for Speaker Recognition: Concepts and Advancement," J. Electr. Syst., vol. 20, no. 1s, pp. 582–599, 2024, doi: 10.52783/jes.806.

[18] L. Li et al., "CN-Celeb: multi-genre speaker recognition ⋆," Speech Commun., vol. 137, pp. 77–91, 2022.

[19] et al. Kudashev, Oleg, "Usage of DNN in speaker recognition: advantages and problems," in Advances in Neural Networks – ISNN 2016, Part I, Springer International Publishing, 2016,

pp. 82–91. doi: 10.1007/978-3-319-40663-3_10.

[20] M. M. Monowar, "Deep Speaker Recognition : Process , Progress , and Challenges," IEEE Access, vol. 9, pp. 89619–89643, 2021, doi: 10.1109/ACCESS.2021.3090109.

[21] S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal, and R. Wang, "Speaker identification features extraction methods: A systematic review," Expert Syst. Appl., vol. 90, pp. 250–271, 2017, doi: 10.1016/j.eswa.2017.08.015.

[22] A. T. Y. T. Al Abd Alazeez, "Streaming Database System for Deaf People," NTU J. Eng. Technol., vol. 1, no. 3, pp. 41–48, 2022, doi: 10.56286/ntujet.v1i3.154.

[23] B. Alsharif and M. Ilyas, "technologies in healthcare for people with hearing impairments," IoT Big Data Technol. Heal. Care, pp. 299–308, 2022.

[24] A. De and L. Fernandes, "CALTRANSCENSE: A REAL-TIME SPEAKER IDENTIFICATION SYSTEM," 2015. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-55.html

[25] M. Kugler, T. Tossavainen, M. Nakatsu, S. Kuroyanagi, and A. Iwata, "Real-time hardware implementation of a sound recognition system with in-field learning," IEICE Trans. Inf. Syst., vol. E99D, no. 7, pp. 1885–1894, Jul. 2016, doi: 10.1587/transinf.2015EDP7432.

[26] K. Ando, S. Serikawa, and Y. Kitazono, "Voice Recognition and Information Transmission System for Hearing Impaired People," in 2017 5th Intl Conf on Applied Computing and Information Technology/4th Intl Conf on Computational Science/Intelligence and Applied Informatics/2nd Intl Conf on Big Data, Cloud Computing, Data Science (ACIT-CSII-BCD), IEEE, Jul. 2017, pp. 53–58. doi: 10.1109/ACIT-CSII-BCD.2017.66.

[27] R. R. Saifan, W. Dweik, and M. Abdel-Majeed, "A machine learning based deaf assistance digital system," Comput. Appl. Eng. Educ., vol. 26, no. 4, pp. 1008–1019, Jul. 2018, doi: 10.1002/cae.21952.

[28] M. Yağanoğlu and C. Köse, "Real-time detection of important sounds with a wearable vibration based device for hearing-impaired people," Electron., vol. 7, no. 4, Apr. 2018, doi: 10.3390/electronics7040050.

[29] L. A. Fanzeres, A. S. Vivacqua, and L. W. P. Biscainho, "Mobile Sound Recognition for the Deaf and Hard of Hearing," arXiv, pp. 1–25, 2018.

[30] "A Computer-Based Application for Speech Recognition in Multi-Speaker Environment to Assist Hearing Impaired People." [Online]. Available: https://ssrn.com/abstract=3509379

[31] P. Dhakal, P. Damacharla, A. Y. Javaid, and V. Devabhaktuni, "A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface," Mach. Learn. Knowl. Extr., vol. 1, no. 1, pp. 504–520, Dec. 2019, doi: 10.3390/make1010031.

[32] R. Tapu, B. Mocanu, and T. Zaharia, "DEEP-HEAR: A Multimodal Subtitle Positioning System Dedicated to Deaf and Hearing-Impaired People," IEEE Access, vol. 7, pp. 88150–88162, 2019, doi: 10.1109/ACCESS.2019.2925806.

[33] D. Jain et al., "HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users," in Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Apr. 2020. doi: 10.1145/3313831.3376758.

[34] D. Jain, H. Ngo, P. Patel, S. Goodman, L. Findlater, and J. Froehlich, "SoundWatch: Exploring Smartwatch-based Deep Learning Approaches to Support Sound Awareness for Deaf and Hard of Hearing Users," in ASSETS 2020 - 22nd International ACM SIGACCESS Conference on Computers and Accessibility, Association for Computing Machinery, Inc, Oct. 2020. doi: 10.1145/3373625.3416991.

[35] J. E. da Rosa Tavares and J. L. Victória Barbosa, "Apollo SignSound: an intelligent system applied to ubiquitous healthcare of deaf people," J. Reliab. Intell. Environ., vol. 7, no. 2, pp. 157–170, Jun. 2021, doi: 10.1007/s40860-020-00119-w.

[36] F. Young, L. Zhang, R. Jiang, H. Liu, and C. Wall, "A deep learning based wearable healthcare iot device for ai-enabled hearing assistance automation," Proc. - Int. Conf. Mach. Learn. Cybern., vol. 2020-Decem, pp. 235–240, 2020, doi: 10.1109/ICMLC51923.2020.9469537.

[37] M. L. Rahman and S. A. Fattah, "Smart Glass for Awareness of Important Sound to People with Hearing Disability," 2020 IEEE Reg. 10 Symp. TENSYMP 2020, no. June, pp. 1527–1530, 2020, doi: 10.1109/TENSYMP50017.2020.9230986.

[38] S. M. Goodman, "Toward User-Driven Sound Recognizer Personalization with People Who Are d / Deaf or Hard of Hearing," vol. 5, no. 2, 2021.

[39] T. E. Ahmed et al., "Smart Deaf Emergency Application Based on Human-Computer Interaction Principles," IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 21, no. 4, 2021, doi: 10.22937/IJCSNS.2021.21.4.35.

[40] R. Shimoyama, "Omnidirectional Haptic Guidance for the Hearing Impaired to Track Sound Sources," Signals, vol. 2, no. 3, pp. 490–507, Sep. 2021, doi: 10.3390/signals2030030.

[41] A. A. Mohamed, A. Eltokhy, and A. A. Zekry, "An Overview of the Development of Speaker Recognition Techniques for Various Applications," J. Commun., vol. 17, no. 8, pp. 632–642, Aug. 2022, doi: 10.12720/jcm.17.8.632-642.

[42] D. Jain et al., "ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users," Conf. Hum. Factors Comput. Syst. - Proc., pp. 1–24, 2022, doi: 10.1145/3491102.3502020.

[43] A. Aiyengar, T. H. Selvan, and V. Moganarengam, "WEARABLE VIBRATION-BASED DEVICE FOR HEARING-IMPAIRED PEOPLE USING ACOUSTIC SCENE CLASSIFICATION," Int. Res. J. Eng.

Technol., 2022, [Online]. Available: www.irjet.net
[44] R. G. A. Buhat et al., “AICanHear: Assistive Alerting Device for Deaf and Hard-of-Hearing People (DHH),” in 8th International Conference on Recent Advances and Innovations in Engineering: Empowering Computing, Analytics, and Engineering Through Digital Innovation, ICRAIE 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICRAIE59459.2023.10468094.
[45] C. L. Chin et al., “A Wearable Assistant Device for the Hearing Impaired to Recognize Emergency Vehicle Sirens with Edge Computing,” Sensors, vol. 23, no. 17, Sep. 2023, doi: 10.3390/s23177454.
[46] M. Goyal and G. Basavarajappa, “A Wearable IoT Based Assistive Device to Aid Communication With Hearing Impaired,” in 2023 IEEE Microwaves, Antennas, and Propagation Conference, MAPCON 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/MAPCON58678.2023.10464174.
[47] M. Tharwat, Y. Wardak, S. Balbaid, and E. Radin, “Wearable Device With Speech and Voice Recognition for Hearing-Impaired People,” in 21st International Learning and Technology Conference: Reality and Science Fiction in Education, L and T 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 221–226. doi: 10.1109/LT60077.2024.10469567.
[48] S. P. Abhiram, S. Umamaheswari, M. S. Muhammed Shybin, A. Saju Campbell, V. G. Sreena, and N. R. Raji, “SI-LERT (Assistive Device for Hearing Impaired Drivers),” in Proceedings of International Conference on Circuit Power and Computing Technologies, ICCPCT 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1516–1521. doi: 10.1109/ICCPCT61902.2024.10673416.