technologies

MDPI

Review

Non-Prosthetic Assistive Technologies for Persons with Hearing Losses: A Survey

Reemas Alsubaiei [1], Farah AlHayek [1], Mariam Alsahhaf [1], Ghadah Alajmi [1], Aliah Almutairi [1], Karim Youssef [1,*], Ghina El Mir [2], Sherif Said [1], Taha Beyrouthy [1] and Samer Al Kork [1]

[1] College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait
[2] College of Business Administration, American University of the Middle East, Egaila 54200, Kuwait
* Correspondence: clement.joseph@aum.edu.kw

Abstract

Millions of persons worldwide experience varying degrees of hearing loss, traditionally addressed through prosthetic solutions such as hearing aids and cochlear implants. However, a significant proportion of individuals cannot benefit from these technologies, cannot access them, or choose not to use them. In this context, non-prosthetic assistive technologies have emerged as a complementary paradigm, leveraging advances in sensing, artificial intelligence, and wearable computing to transform acoustic information into alternative perceptual representations rather than restoring auditory function. This survey provides a review of such systems, focusing on technologies that enhance environmental awareness, communication, and social interaction. Existing approaches are categorized along two main dimensions: the tasks they perform and the platforms on which they operate. Task-oriented analysis includes sound recognition (speech and non-speech), sound source localization, emotion recognition, sign language recognition, and related emerging functionalities. Platform-based analysis emphasizes wearable devices and mobile solutions enabling real-time and context-aware assistance. The survey further highlights key research trends, including real-time auditory scene analysis, portable processing, and artificial intelligence. It shows that recent studies increasingly demonstrate that combining auditory, visual, and haptic modalities improves robustness and usability in real-world conditions, particularly in noisy and dynamic environments. Finally, open challenges such as energy efficiency, latency, evaluation methodologies, and user acceptance are discussed. By synthesizing existing work and identifying open research directions, this survey aims to provide a structured foundation for future developments in intelligent, non-prosthetic assistive systems that redefine how auditory information is accessed and interpreted.

Keywords: sound recognition; sound classification; assistive technology; hearing-impaired; hard of hearing; hearing losses; sign language; real-time audio detection; sound source localization

Check for updates

Academic Editor: R. Simon Sherratt

Received: 22 March 2026
Revised: 24 April 2026
Accepted: 6 May 2026
Published: 13 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Hearing impairment is a widespread and growing global health concern that significantly affects the safety, independence, and quality of life of millions of individuals. According to the World Health Organization (WHO), over 430 million people currently live with disabling hearing loss, a number projected to rise to more than 700 million by 2050 [1]. Individuals who are deaf and hard-of-hearing (DHH) can repeatedly question about things not properly heard, inappropriately answer misheard questions, speak loudly,

and in unilateral hearing impairment, they turn the healthy ear to the sound source [2]. Additionally, they often face serious challenges in perceiving critical environmental sounds such as alarms, vehicle horns, emergency sirens, and spoken warnings, which can increase their risk of accidents and unsafe situations. Additionally, prejudices about deafness can lead to limiting the access of affected individuals to social interaction, education and employment for example [3].

Hearing loss can be conductive, resulting from a disruption of the transmission of sound waves to the cochlea, or sensorineural, resulting from problematic transmission of stimuli at the cochlea or after it [4]. Various reasons exist for hearing impairment, such as genetic, prenatal, perinatal, postnatal, head injuries and otitis media [4]. Additionally, the chance of developing hearing loss increases with age [5].

Traditionally, hearing aids and cochlear implants have been among the most commonly used methods to improve sound perceptions and well-being of individuals with hearing losses [6,7]. Other related technologies like middle ear implants [8] and auditory brainstem implants [9] have also been used. A hearing aid consists of three main parts: a microphone, an amplifier and a speaker and is worn in or behind the ear to amplify sounds [10]. Additionally, cochlear implants are devices designed to restore sensorineural hearing loss, and improve the hearing of individuals with severe hearing loss [11]. The field of cochlear implantation is continuously evolving, with advances in electrodes, signal processing and connectivity [12]. However, such technologies can be associated to negative effects on the individuals using them. Aside device manufacturing and malfunction issues, medical adverse effects like ear infection and skin concerns have been witnessed in hearing aids [13], and hearing aids can amplify undesired sounds [14] and cause discomfort [15]. Also, implant surgery can result in complications affecting results [16], and implants can cause physical discomfort and functioning problems, and complications requiring reimplantation [17], aside factors like reimplantation rejection, loss of audio processors and cognitive difficulties [18]. Additionally, hearing aids and cochlear implants do not provide meaningful contextual interpretation of auditory events or clearly indicate the nature and source of sounds. Indeed, many users face difficulties in complex acoustic environments, multi-party communication, safety awareness and accessibility to spoken and auditive non-spoken information. These limitations highlight a significant gap in these solutions and demonstrate the need for intelligent assistive technologies capable of detecting, interpreting, and conveying environmental sound information in real time to improve situational awareness and personal safety. Video remote interpreting technologies have been used to improve the communication between healthcare professionals and patients [19], benefiting from advances in telecommunication technologies but subjected to issues in connection and maneuvering flexibility [20]. Also, rapid advances in sensing, computing, wireless communication and human-machine interfaces have led to the emergence of a diverse set of assistive technologies that operate outside the traditional paradigms. Indeed, technologies such as assistive listening, visual and multimodal alerting [21], speech recognition [22,23], haptic and tactile devices [24–26] have been proposed and were shown to bring improvements to the quality of life of hearing impaired individuals. Unlike prostheses, such solutions often provide task-related assistance rather than auditory restoration, and are implemented on heterogeneous platforms ranging from wearable devices and smarthphones to smart environments and robotic systems.

Despite their growing impact, existing reviews of non-prosthetic hearing impaired-assistive devices tend to examine them in isolation or within narrow application domains, without systematically comparing the tasks they support or the platforms on which they are deployed. This fragmentation limits the ability to identify trends and gaps in functionality, as well as opportunities for cross-platform integration. In particular, there is a

lack of a unified framework that jointly analyzes these systems from both functional and implementation perspectives, making it difficult to understand how different approaches relate to each other and to identify emerging design principles. Furthermore, existing studies rarely provide a comprehensive comparison between semantic-level processing (e.g., sound recognition and speech transcription) and spatial-level processing (e.g., sound source localization), despite their complementary roles in enhancing situational awareness. In addition, limited attention has been given to the integration of multimodal feedback mechanisms (visual, auditory substitution, and haptic), and how such combinations impact usability and user experience in real-world scenarios. Another notable gap lies in the insufficient discussion of system-level constraints, such as real-time performance, energy efficiency, and scalability across different hardware platforms, which are critical for practical deployment. This shows the need for a structured review that categorizes these technologies according to the tasks they address and the platforms they use. These two dimensions are complementary, as task performance is inherently influenced by platform capabilities and constraints. Such a review can provide a unifying framework that can inform the design of next-generation multimodal systems, and support customizable and personalized solutions for hearing impaired users. Thus, in this literature review, assistive technologies for hearing impaired individuals, which are not within the hearing aid or implant fields are reviewed. This review focuses on the tasks performed by the different solutions and the platforms used. Figure 1 illustrates the proposed classification framework used in this survey, which is organized along two complementary dimensions: *tasks* and *platforms*. The *tasks* dimension refers to the type of information extracted from the acoustic environment, such as sound recognition, sound source localization, emotion recognition, and sign language recognition. These tasks represent different levels of interpretation, ranging from identifying sound events to understanding contextual and social cues. In contrast, the *platforms* dimension categorizes how these functionalities are implemented in practice, distinguishing between wearable devices and mobile or non-wearable systems. This dual classification enables a structured analysis of existing solutions by linking system functionality to implementation constraints. Among the listed tasks, emotion recognition plays a complementary role to speech recognition. While speech-to-text systems provide the semantic content of spoken language, they often fail to capture paralinguistic information such as tone, intent, or emotional state. Emotion recognition addresses this limitation by enabling the interpretation of affective cues, such as urgency, stress, or friendliness, which are essential for effective communication and social interaction. For hearing-impaired users, this additional layer of information enhances contextual understanding and reduces ambiguity in everyday interactions, justifying its inclusion as a distinct task in the proposed framework. It is important to note that the distinction between wearable devices and mobile platforms in this survey is based on the *primary interaction and deployment modality*, rather than the underlying computational architecture. In practice, many assistive systems adopt hybrid designs, where wearable devices are used for sensing and user interaction, while mobile devices or cloud services provide additional processing power. However, for the purpose of classification, a system is categorized as *wearable* if the main interaction with the user occurs through a body-worn device, even if external computation is involved. Conversely, a system is categorized as *mobile* if the primary interface and processing are handled by external platforms such as smartphones or robots, without requiring continuous body-worn interaction. This distinction allows for a clearer analysis of system usability, responsiveness, and real-world deployment, while acknowledging the existence of hybrid architectures that combine both paradigms. The review does not aim to be exhaustive, but to offer a structured and critical overview of representative works, with particular emphasis on recent relevant contributions, current trends, and future research directions.

Thus, this review will help to see the gaps in existing approaches, highlight underexplored combinations of tasks and platforms, and guide future research toward more integrated, efficient, and user-centered assistive solutions.

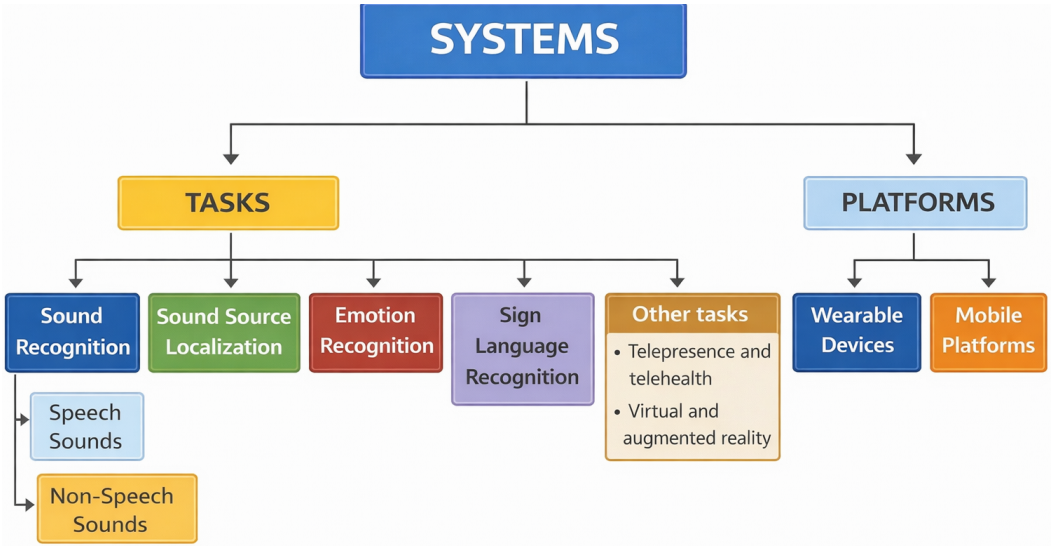


Figure 1. Aspects of hearing impaired assistive systems covered in the literature review.

The rest of the paper is organized as follows. Section 2 shows a review of the different tasks performed by non-prosthetic hearing impaired assistance devices. Section 3 reviews the different types of platforms used in these devices. Section 4 discusses the outcomes of this review and Section 5 concludes the paper.

2. Tasks

To implement an assistive wearable device, the first step is knowing how the system works and the functions it performs. This notably includes sound recognition, which covers both speech and non-speech sounds, as well as sound source localization and emotion recognition. Different reviewed work in this context are summarized in Table 1 and their percentages among the reviewed work are shown in Figure 2. While Figure 2 provides a statistical overview of the distribution of tasks across the reviewed literature, it does not fully reflect how these components can be integrated within a complete assistive system. In practice, multi-tasking modern assistive systems should follow a structured processing pipeline in which multiple tasks operate sequentially and complement each other.

Typically, the system begins with an audio acquisition stage, where environmental sounds are captured using microphones or microphone arrays. The captured signal is then preprocessed (e.g., filtering, framing, and feature extraction such as MFCC or spectrograms) to prepare it for analysis. The first functional stage is usually sound recognition, which identifies the type of sound event (e.g., speech, alarm, vehicle). This information provides the semantic context of the environment.

In parallel or subsequently, sound source localization estimates the direction of the sound source, providing spatial awareness. This is particularly important for safety-critical scenarios, such as detecting approaching vehicles or alarms. For speech signals, additional processing such as speech-to-text conversion extracts linguistic content, while emotion recognition can further analyze tone and prosody to provide contextual and affective information.

Table 1. Representative example works categorized by task and implementation approach.

Category	Work	Task	Method
Speech	Pavlidou and Lo [27]	Speech → vibration	Phoneme encoding; RNN; haptic
	Abi Sen et al. [28]	Keyword alerts	Google STT; vibration
	Yamamoto et al. [29]	Captioning	ASR
	Yağanoğlu [30]	Speech recognition	MFCC + DTW
	Tharwat et al. [31]	Speech + speaker ID	Whisper; RF
	Gupta and Vishwakarma [32]	Speech	Google STT
	Peddi et al. [33]	Speech + classification	MFCC; CNN
Non-speech	Adnan Habib et al. [34]	Multi-class classification	CNN/RNN
	Jain et al. [35]	Few-shot recognition	ProtoNet
	Chin et al. [36]	Siren detection	CNN
	Buhat et al. [37]	Environmental sounds	YAMNet
	Salem et al. [38]	Road sounds	CNN; RF ; KNN; DT
	Munasinghe and Dulanjani [39]	Emergency sounds	CNN; SVM ; KNN; RF
	Gurrala et al. [40]	Sudden sounds	Threshold-based
Localization	Choe et al. [41]	Localization + classification	GCC-PHAT; BSS
	Tang and Zhang [42]	Tracking + localization	ODAS; Kalman
	Matsuo et al. [43]	Direction estimation	ITD/ILD
	Lim et al. [44]	Localization + detection	MFCC; NN; ILD
Emotion	Ridha and Shehieb [45]	Speech emotion	CNN
	Fukui et al. [46]	Crowd emotion	VGGish
	R et al. [47]	Multimodal emotion	Transformers
	Gupta and Vishwakarma [32]	Emotion analysis	Wav2Vec2
SLR	Fu et al. [48]	Gesture recognition	Sensor glove
	Prasath and Panaiyappan [49]	Speech → sign	CNN + RNN
	Talaat et al. [50]	Translation	YOLOv8
	Tyagi et al. [51]	Real-time gestures	MediaPipe
Extended	Santomauro et al. [52]	Telehealth	Remote healthcare
	Liu et al. [53]	Telehealth usability	Evaluation
	Serafin et al. [54]	VR rehab	Virtual reality
	Mehra et al. [55]	AR audio	Augmented audio

Each of these tasks contributes complementary information: sound recognition answers *what* the sound is, localization determines *where* it originates, and emotion recognition provides insight into *how* it is expressed. The outputs of these stages are then fused and delivered to the user through appropriate feedback modalities, such as visual displays, haptic alerts, or augmented reality interfaces. Therefore, rather than functioning independently, these technologies are optimally combined into a unified system that transforms raw acoustic signals into meaningful, multimodal information, enhancing situational awareness and communication for hearing-impaired users. Figure 2 should therefore be interpreted as a distribution of functionalities, while an actual multitasking system implementation relies on the integration of these tasks within a unified processing pipeline.

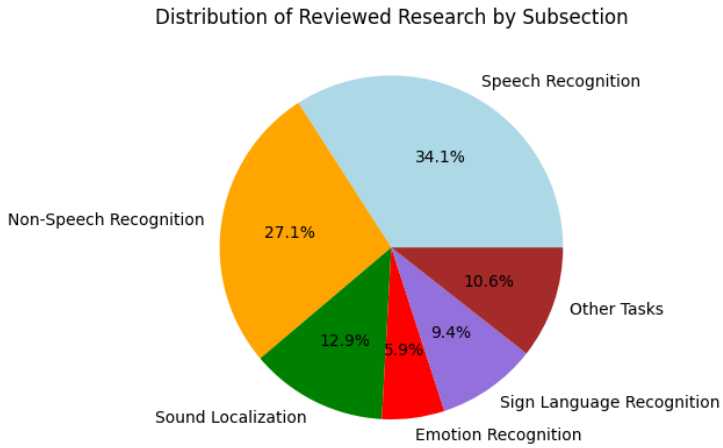


Figure 2. Distribution of reviewed citations across assistive-technology tasks for deaf and hard-of-hearing users. The chart shows the proportion of studies focusing on sound recognition, non-speech sound analysis, sound source localization, emotion recognition, and other assistive functions, highlighting the strong research emphasis on sound recognition and environmental sound understanding compared with localization and emotion-aware systems.

To improve clarity and readability, Table 1 presents a set of representative works grouped by task category, rather than listing all reviewed studies. This organization aligns with the task-based classification introduced in Figure 2 and allows for a clearer understanding of how different approaches relate to specific assistive functionalities. In addition to core auditory processing tasks, Table 1 includes selected extended assistive systems such as telepresence, telehealth, and AR/VR-based solutions. Although these systems do not always perform direct sound recognition or localization, they contribute to hearing assistance by enhancing communication, accessibility, and remote interaction. Their inclusion reflects the broader ecosystem of assistive technologies that support hearing-impaired users beyond isolated signal processing tasks.

2.1. Sound Recognition

Sound recognition involves detecting, analyzing, and classifying audio signals so that a system can understand and interpret different acoustic events in the environment. This function forms the core of assistive hearing technologies, enabling devices to differentiate between speech sounds, which carry linguistic meaning, and non-speech sounds, which include environmental cues such as alarms, traffic noise, and daily activity sounds.

2.1.1. Speech Sounds

Most of the newest wearable devices for deaf and hearing-impaired people focus on converting human speech into text and vibration instead of amplifying the sound. In 2021, Pavlidou and Lo [27] presented a portable phoneme-based speech-to-tactile translation system that enabled users to perceive speech through vibration while maintaining eye contact. The recognition process employed phoneme-level analysis using Fourier Transformation to extract frequency patterns and utilized machine learning (ML) techniques, including Recurrent Neural Networks (RNNs), to model the sequential nature of speech. The 44 English phonemes were encoded into binary patterns to drive vibration motors, with three coin-type motors representing seven letters derived from consonant–vowel combinations. Tested on three participants, the system achieved real-time processing of approximately 10–15 phonemes per second, and users required about 30 min to learn and identify the seven tactile letters. Abi Sen et al. [28] proposed a system, relying on an application that users can use to set up a group of words they want the system to

detect. After the application captures a sound, it converts it into text using Google library API STT (Available: https://docs.cloud.google.com/speech-to-text/docs/reference/rest, accessed on 27 January 2026.) (Application Programming Interface speech-to-text). Then, it compares the text to the stored words. If the text matches one of the stored words, the application will send an alert as vibrations. However, the number of words this system can identify and categorize is limited. Ridha and Shehieb [45] proposed a system using a smart glasses device with a microphone array added to it. The developed device became Android Augmented Reality (AR) glasses. This design integrates the real world with digital elements, translating speech into text to display the translation. The augmented reality appears in only one lens. Speech to text conversion was performed using Google Cloud Speech-to-text API[1], and to assist users in distinguishing what each person is saying, a color was assigned to each text item for differentiation purposes. The system used a convolutional Neural Network (CNN) trained on the UrbanSound dataset (Salamon et al. [56]) and they also created a dataset of voices they had collected. The model achieved a validation accuracy of approximately 52.5%. Also, Yamamoto et al. [29] developed a system in which automatic Speech Recognition (ASR) was employed to convert spoken language into visual text for deaf and hard-of-hearing users. The system utilized a directional microphone to capture the speaker's voice and a real-time speech-to-text engine displayed on a transparent screen, allowing both the speaker and the user to view the captions simultaneously. This design enabled immediate confirmation and correction of recognition errors, improving communication accuracy. Furthermore, the system presented by Yağanoğlu [30] recognized ten common sentence patterns (e.g., "How are you?", "Call me tomorrow") and converted them into distinct vibrations. The wearable platform uses a Raspberry Pi, microphone, and vibration motor, operating standalone without internet. Audio is sampled at 16kHz, filtered, framed (20 ms with 10 ms overlap), and processed with a Hamming window. MFCC features (39-dimensional vectors) are extracted, and DTW handles acoustic modeling, aligning input speech with reference patterns despite timing variations. The system achieves 95% accuracy overall, maintaining high performance even at 5dB Signal-to-Noise Ratio (SNR). While most wearable devices focus on converting speech into text or tactile vibration, Fu et al. [48] focused on using speech-to-gesture animation to bridge the communication gap. The paper proposed an intelligent Human–Computer Interaction (HCI) system based on the LD3320A voice module, which integrates speech recognition and voice output in a single chip. When speech is recognized, a hexadecimal command is sent to an Arduino Mega2560 microcontroller and then transmitted to Unity 3D to generate sign language animation for hearing-impaired users. The system was tested on six specific words (e.g., "study", "you", and numbers 1–4) and achieved an overall recognition accuracy of 90.4%.

In 2022, the system presented by Anwaar et al. [57] integrated audio recognition within a multimodal platform to detect and communicate external hazards. For speech processing, TinyML techniques were applied using Mel-Frequency Cepstral Coefficients (MFCCs) (also used by Youssef et al. [58,59]) to extract acoustic features through Mel-scale mapping and frequency-domain analysis. The preprocessing pipeline included analog-to-digital conversion, framing, and spectral transformation to enable lightweight keyword classification. A neural network (NN) model identified predefined warning words such as "Stop" and "Watch out," triggering directional haptic alerts. This integration of audio classification within the safety system enhanced situational awareness for hearing-impaired users. In addition, Senthil et al. [60] proposed a real-time Speech-to-Text (STT) system to help deaf individuals understand spoken language without requiring the speaker to know sign language. The system uses a microphone and Raspberry Pi 3B+ to capture and process speech, and perform speech recognition. Real-time testing showed accuracy between 86.66% and 92.3%, with the microprocessor enabling faster, multitask processing

compared to microcontroller-based systems. Finally, Prasath and Panaiyappan [49] presented a multi-stage real-time speech recognition system for deaf users. Voice Activity Detection (VAD) uses long-term differential entropy with an improved Mel-DCT filter, while Maximum a Posteriori (MAP) selection refines voiced/unvoiced detection and GrabCut segmentation isolates speech from audio–video frames. Segments are converted into 2D representations via a sliding window and classified using an ensemble Region-based Convolutional Neural Network (R-CNN), where CNNs extract spatial features and RNNs model temporal sequences, optimized with Stochastic Gradient Descent (SGD). Noise robustness was achieved through training with vehicle noise at multiple SNRs, forming a comprehensive assistive speech recognition framework.

In 2023, research has produced newer solutions in human speech recognition. Staš et al. [61] presented a mobile assistive application for deaf and hard-of-hearing users, named Hear IT. The application utilizes an open-source Kaldi engine (Povey et al. [62]) for automatic speech recognition (ASR). Based on ASR, the system converts speech into text in both online and offline modes. The ASR framework integrates voice activity detection to identify speech segments and processes audio streams in chunks. Intermediate recognition results are generated to display outputs with minimal response time. The final output is generated after speech completion, detected by VAD. The system then applies re-scoring using a Recurrent Neural Network Language Model (RNN-LM) to enhance accuracy and produces N-best hypotheses from the lattice for alternative recognition results. Bao et al. [63] used keyword spotting techniques through the LD3320 chip. The LD3320 chip is based on Automatic Speech Recognition (ASR) technology (Errattahi et al. [64]) which detects only predefined keywords. The chip performs sound signal processing, including amplification and noise reduction, before conducting semantic recognition to output the corresponding text. Rathna et al. [65] presented an assistive system with LED alerts, and vibrations. The system doesn't translate "speech to text" literally; it only recognizes speech cues or human warning calls. It uses lightweight ML algorithms, which were very accurate: K-Nearest Neighbors (KNN) = 93% and Support Vector Machine (SVM) = 94%.

Recent research studies in 2024 further support assistive systems for hearing-impaired and deaf (HID) individuals, focusing on human speech recognition. J et al [66] presented a method where an offline ASR system converted speech into text on a small screen, supporting long sentences and conversations, achieving 83.3% accuracy, 16.67% word error rate, and <0.03 s response time for short indoor distances. A related study by Tharwat et al. [31] proposed a real-time speech-to-text system with speaker identification. The system applied automatic speech recognition, combined with preprocessing steps such as sampling, normalization, and noise reduction. MFCC features were extracted and fed into a Random Forest (RF) classifier to improve accuracy and reduce recognition errors by 50% and identified speakers with 93.37% accuracy. Another study by Peddi et al. [33], a dual-model system combines Wake Word Detection and Speech Recognition. The user-defined wake word (e.g., their name) was trained on 100 audio samples using a two-layer CNN, achieving 95.2% accuracy. Upon detection, the wake word model triggers Google's Web Speech API for speech-to-text conversion and temporarily halts itself. If no speech is detected within 10 s, speech recognition stops and wake word detection resumes, optimizing resource use and system responsiveness. Also, G et al. [67] focused on facilitating short, direct conversations with rapid responses. Speech was converted to text using the Google Speech-to-Text API, while an Artificial Neural Network (ANN) model handled speech recognition and a Hidden Markov Model (HMM) supported text-to-speech conversion. However, R et al. [47] used Facebook AI Wav2Vec 2.0 (Coban et al. [68]), which includes cutting- edge pre-training and fine-tuning procedures for high accuracy in noisy environments. It also used Conditional Generative Adversarial Networks (cGANs) to generate

data to improve the accuracy. Building on a different approach, Talaat et al. [50] introduced a real-time Arabic avatar system translating spoken Arabic into dynamic Arabic Sign Language (ArSL) (Belmadoui [69]). Noise robustness is achieved using babble, white, factory, and Volvo noise, along with time, pitch, and speed augmentations. The four-phase ArSLGen algorithm first annotates Arabic speech with transcripts for ground truth. Next, speech or text input is processed with YOLOv8 to generate avatar gestures. The bidirectional system also converts recognized ArSL back to Arabic speech or text, achieving up to 99.5% Mean Average Precision (mAP).

Building on previous work, research in 2025 introduced further advances in assistive systems for HID individuals. Gupta and Vishwakarma [32] proposed a software-based system provided real-time speech-to-text transcription for hearing-impaired users using Google's Speech-to-Text API. Audio was preprocessed for noise reduction and converted into spectrograms to identify phonemes. Deep learning and language models enabled support for multiple languages and accents with low latency (250 ms). CNN-based speaker identification (99.3% accuracy). Similarly, Ubur [70] developed a speech recogntion-based captioning system using a lightweight Automatic Speech Recogntion (ASR) engine for real-time transcription. The recognized speech was displayed through a Unity-based AR application, allowing captions to be embedded directly within the user's visual field. This approach reduced cognitive load and improved comprehension for Deaf and Hard-of-Hearing (DHH) students. Senaha et al. [71] developed AR smart glasses to display speech as text while identifying the speaker. Evaluated on 90-s Japanese Language Proficiency Test audio clips, Voice Activity Detection (VAD) increased utterance segments by 29.6%, reduced characters per display by 21% for readability, and lowered word error rate (WER) by 33%. Speaker recognition on 100 participants achieved 99.7% (studio) and 87.7% (re-recorded) for 6-s clips, and 41% (studio) and 22.7% (re-recorded) for 1-s clips. Limiting to five speakers improved 2-s re-recorded clip accuracy to 93.3%. More recently, Tang and Zhang [42] proposed a robotic system capturing 5-s audio segments, analyzed via YAMNet to detect human speech. Recognized speech was converted to text using Google's Speech Recognition API and displayed via a Tkinter interface. Another work proposed by M et al. [72] focused on children with hearing loss, particularly those with congenital impairments, whose speech is often difficult to understand. The system uses a CNN model to classify children's speech signals and applies natural language processing (NLP) for post-processing. Initial accuracy was 23%, which improved to 67% after applying spectrogram augmentation. Also, the system proposed by J et al. [73] aimed to improve communication between HI (Hearing-Impaired) and mute individuals with the wider community, as not everyone knows how to use or understand sign language. The system used microphones and cameras to read lip and gesture movements, convert speech to text, alert users via vibration motors, and translate sign language into speech using CNNs and LSTMs. Gawli et al. [74] presented smart glasses for only speech to text. First, an electret microphone amplifier module captures the speech since it has low noise performance and sends it to a computer. Then Speech Recognition API processes the audio and Google's Web Speech API used to convert the audio to text and display on glasses. These processes take 0.2–0.4 s for microphone capture and pre-processing and 0.6–1.3 s for speech-to-text conversion via Google API with accuracy > 90%. And more recently, speech recognition was investigated by Hughes et al. [75] in its clinical applications for patients with hearing loss. This study supported the potential of speech recognition in real-time to improve communication for deaf and hard-of-hearing (DHH) persons in clinical care.

Comparative Analysis:

A comparison of the reviewed speech-based assistive systems reveals several key distinctions in terms of input processing, output modality, computational complexity, and application scope. From an input perspective, most systems rely on microphone-based audio capture, with some incorporating directional microphones or microphone arrays to enhance signal quality and noise robustness. In terms of processing techniques, approaches range from classical methods such as MFCC combined with DTW [30], to deep learning-based models including CNNs, RNNs, and transformer-based ASR systems such as Whisper [31].

Regarding output modalities, the majority of systems convert speech into visual text (e.g., AR displays, mobile screens), while others utilize haptic feedback through vibration patterns [27,30]. A smaller subset of works explores alternative modalities such as sign language animation [48]. In terms of functionality, some systems focus on continuous speech transcription, whereas others are limited to keyword spotting or predefined phrases, which reduces flexibility but improves computational efficiency.

Furthermore, system deployment varies between cloud-based solutions (e.g., Google Speech APIs) and fully embedded systems operating offline (e.g., Raspberry Pi-based implementations), highlighting a trade-off between accuracy and latency versus privacy and independence. Finally, robustness to noise and real-time performance remain key differentiating factors, with recent works increasingly incorporating noise augmentation and lightweight models to improve usability in real-world environments.

2.1.2. Non-Speech Sounds

Non-speech sound recognition has become a core component in modern assistive technology research, especially for systems designed to help deaf and hard-of hearing individuals interpret critical environmental audio events. In 2021, Ridha and Shehieb [45] used a MEMS microphone array and CNN models to classify a wide range of non speech environmental events such as glass breaking, gunshots, and car horns, projecting the recognized sound types and directions directly in AR glasses. During the same year, Yu et al. [76] detected household non speech sounds including fire alarms, boiling water, and door knocks using a ReSpeaker mic array and YAMNet, communicating results through LEDs, screen messages, or robot movement. In addition, Goodman et al. [77] focused on specialized environmental sounds and examined non-speech sound recognition with 14 DHH participants using a three-stage approach. First, participants trained Google's Teachable Machine [78] on three personalized sounds ensuring high-quality training samples. Second, they recorded daily non-speech sounds for one week, collecting at least 15 unique sound classes. Finally, researchers analyzed 677 samples using the pydub library (Available: https://pydub.com) and applied reflexive thematic analysis to interviews and diaries to understand participants' challenges and sense-making strategies. This study is particularly significant because it highlights the importance of user-driven personalization, showing that assistive sound recognition systems can be more effective when adapted to the specific daily environments and priorities of Deaf and Hard-of-Hearing users (Goodman et al. [77]). To better understand the evolution of non-speech sound recognition systems, the reviewed works can be compared in terms of input configuration, classification methodology, output modality, and deployment constraints.

By 2022, recognition capabilities became more structured, and model driven. In Anwaar et al. [57], researchers applied TinyML to classify non-speech outdoor sounds like horns and movement noise, generating real-time haptic alerts. Unlike this work where the dataset limited to only five classes, the dataset used by Adnan Habib et al. [34] contained exactly 50 classes which each classes consisting of 1200 samples. Both CNN and RNN mod-

els were employed within the same experimental context and all achieved high accuracy, reaching 98.67% for CNN and 97.01% for RNN. Contrary to the common expectation that RNN models typically achieve lower accuracy compared to CNNs in audio classification tasks, the result demonstrate only a small performance difference between the two models. In a more advanced methodology, Jain et al. [35] proposed "ProtoSound", an on-device personalized sound recognition system using few-shot meta-learning. After comparing five algorithms, Prototypical Networks [79] was chosen for its 95.6% accuracy and minimal training time. Audio signals were converted to log-mel spectrograms with Cepstral Mean and Variance Normalization (CMVN) and processed via a lightweight MobileNetV2 CNN [80], enabling real-time mobile inference. Pre-trained on 35 environmental sound classes from six libraries, the system included three innovations: context generalization, open-set classification, and a library of 10 difficult-to-produce sounds (e.g., sirens, fire alarms). Using a 5-way, 5-shot setup, it achieved 88.9% accuracy on recordings from 21 real-world locations covering 22 sound classes. During the same year, Fukui et al. [46] proposed a system that enables deaf and hard-of-hearing (DHH) users to follow sports events by detecting non-speech crowd cheering sounds. Also, a pretrained deep neural network (Hershey et al. [81] is employed as a feature extractor for environmental audio collected from recorded matches. Then the classification model is trained using Create ML and deployed for real-time analysis via SNAudioStreamAnalyzer on both mobile and wearable devices. Also, An et al. [82] focused on recognizing sounds through image-based deep learning. The system records sounds that exceed 60 dB and then transmits them to the server (Firebase). The server converts the sound signal into a Mel-Spectrogram. After that, the YOLO learning model classifies the sound into four major daily-life sounds: baby crying, fire alarm, washing machine shutdown, and doorbell. The classification accuracy of the system in real-life conditions was 85%, and for mixed sounds it was 80%.

During 2023, the field showed both specialization and broader expansion. Chin et al. [36] focused specifically on detecting emergency vehicle sirens using a lightweight CNN optimized for edge devices, demonstrating a narrow but highly practical single-class solution. Meanwhile, in Choe et al. [41], the Audio Spectrogram Transformer (AST) achieved a high accuracy for environmental sound classification. This result highlights the growing effectiveness of transformer-based models and standardized datasets for robust non-speech sound recognition. To enhance real-time performance in white-noise environments, Fast Fourier Transform (FFT) combined with Otsu's method was further applied as a preprocessing stage for noise reduction. To enhance real-time performance in white-noise environments, Fast Fourier Transform (FFT) combined with Otsu's Method was applied as a noise reduction preprocessing technique. Through the accuracy comparison of sound classification models, the AST achieved the highest accuracy. Also, T et al. [83] presented a solution focusing only on how to convert the sound to vibration feedback. The solution is a watch that works for specific features like alarms, and the benefit is that it is small and easy to use. This technique uses a watch that has a small microphone capture the sound, then converts it to electrical signals. The next step is to analyze the signal: first filtering, and second feature extraction such as frequency or amplitude. The final step is encoding, which converts the signal into vibration. Also in that year, Buhat et al. [37] used a pretrained YAMNet model to recognize a wide range of indoor and outdoor non-speech sound events, achieving practical accuracy without requiring extensive custom datasets.

Later, recognition systems reached more advanced levels. In 2024, Rathna et al. [65] integrated a multi-microphone array with lightweight embedded ML models to classify horns, alarms, engine noise, and general ambient sounds, while also offering directional awareness. Also, Sun et al. [84] introduced early non-speech recognition by identifying vehicle horns using basic filtering and signal-processing techniques. This work demonstrated

that traffic-related non-speech cues can be transformed into haptic alerts for user safety. The system by S P et al. [85] also targeted siren-based recognition using MFCC and spectrogram data with a CNN, offering partial non-speech coverage. Finally, Peddi et al. [33] presented a dataset that contains 8000 audio file including drilling, sirens, car horns, dog barks, children playing, air conditioners, engine sounds, hammers,street music, and gunshots, and more sounds can be added. For pre-processing, "Librosa" is used for the analysis audio signals. MFCC feature extraction is used for sound recognition. The dataset was trained on a CNN with 4 layer. The model achieved 92% on accuracy. And starting from 2025, Karroum et al. [86] implemented a complete multi-class recognition wearable with a cloud-based CNN trained on UrbanSound8K [87], enabling detection of horns, alarms, glass breaking, and door knocks. another 2025 contribution by Tang and Zhang [42] combined YAMNet classification with ODAS localization to recognize and track diverse non-speech events such as fire alarms, boiling water, and knocking allowing robots to navigate toward sound sources. In addition, Salem et al. [38] classified multiple non-speech road sounds such as car horns, emergency sirens, and engine noise using ML models including CNN, RF, KNN, and Decision Tree (DT), establishing reliable multi-class performance. Munasinghe and Dulanjani [39] proposed a system targeting sound-based risk assessment. User audio is first classified by an emergency sound detection model (ambulance, police, firetruck, gunshots, others) and then by an environment classifier (outdoor, indoor, in transport). Based on both results, the system calculates the risk level. audio samples were standardized to 3 s, with longer clips truncated and shorter ones zero-padded. A balanced dataset with class weighting was used during training. Among RF, SVM, KNN, and CNN models, the CNN achieved the highest accuracy at 94%. Another study by Lim et al. [44] proposed a system that targets selective emergency sounds (dog bark, car honk, siren). A dataset of 456 audio samples from Freesound was preprocessed using MFCC features and trained with a neural network. Incoming sounds are recorded for 5 s via a USB microphone if the decibel level is 80 or more. The system classifies these sounds with 90% accuracy. Finally, Gurrala et al. [40] offered a simpler version of non-speech sound processing by detecting sudden noises, alarms, or knocking based on amplitude thresholds rather than ML classification.

Across recent research, the most advanced implementations in non-speech sound recognition for assistive technology combine multi-class detection, real-time embedded processing, and actionable alerts. The approaches represent leading developments in the field, demonstrating robust performance and practical applicability in real-world assistive settings.

Comparative Analysis:

A comparative examination of the reviewed non-speech sound recognition systems reveals several key differences in design and performance. In terms of input configuration, most systems rely on single or multi-microphone setups, with microphone arrays providing enhanced spatial awareness and noise robustness [45,65]. Regarding classification approaches, earlier works primarily utilized CNN and RNN models, while more recent approaches incorporate transformer-based architectures such as AST [41] and few-shot learning techniques [35] to improve adaptability and scalability.

From a functionality perspective, systems vary between single-class detection (e.g., siren detection [36]) and multi-class environmental recognition (e.g., UrbanSound-based models [86]), highlighting a trade-off between specialization and generalization. In terms of output modalities, most solutions provide visual or haptic feedback, while a few integrate robotic responses or navigation capabilities [42].

Furthermore, deployment strategies differ significantly, with some systems relying on cloud-based processing for higher accuracy, while others implement lightweight embedded or TinyML solutions for real-time, low-latency operation [57]. Finally, robustness to noise and real-world variability remains a key challenge, with recent works addressing this through data augmentation, noise-aware preprocessing, and balanced datasets.

2.2. Sound Source Localization

Deaf and hearing-impaired individuals need assistive solutions that not only recognize sounds or speech but also localize the sound source. Unlike hearing humans, who use binaural cues (ITD and ILD) [88] to estimate sound direction, severely impaired individuals may miss environmental context even with automatic translators. Hence, many assistive systems emphasize artificial sound localization to alert users to the direction of alarms, vehicles, speech, or sudden events. In 2021, Yu et al. [76] focused on detecting important indoor sounds and notifying users via multiple interfaces. Using a "ReSpeaker", a microphone array for sound recognition, the system identified sound locations relative to the robot's position. This approach guided users successfully toward the sound source. To better analyze the evolution of sound source localization systems, the reviewed works can be compared in terms of sensing configuration, localization methodology, computational complexity, and achieved accuracy.

During 2022, new research proposed different methods for sound localization. For example, Anwaar et al. [57] proposed a multimodal vest for car drivers, combining a camera and LiDAR for sound and object localization. The camera captured detailed surroundings, while LiDAR provided a high-resolution 360° horizontal view for localizing the human speaking. Although each sensor worked well individually, fusing them proved impractical due to the microcontroller's limited image-processing capabilities.

In 2023, Choe et al. [41] used a ReSpeaker 4-mic array to localize sound sources via Time Difference of Arrival (TDoA). The system employed ODAS sound localization algorithms (Grondin and Michaud [89]) and applied Generalized Cross-Correlation with Phase Transform (GCC-PHAT) for real-time TDoA calculation. To handle mixed sound sources, Blind Source Separation (BSS) combining PCA (Jolliffe and Cadima [90]), NMF (Wang and Zhang [91]), and ICA (Grondin and Michaud [89]) were used to separate overlapping sounds and improve localization accuracy.

Later approaches in 2024, such as the approach presented by Tharwat et al. [31] used only microphones for localization. Their wearable headband employed four microphones to capture audio, processed by a Raspberry Pi, which sent directional cues to LEDs indicating sound origin from four sides. This setup provided 360° orientation, but accuracy did not exceed 60% due to the lack of noise filtering. In addition, Matsuo et al. [43] proposed a smart glasses-based system for sound source localization, using two outward-facing unidirectional microphones on the left and right sides. Using two microphones instead of four reduces response time. The system determines whether a sound originates in front of or behind the user based on the loudness difference, while the angle of the sound source is calculated from the time difference of arrival between the microphones. Also in 2024, Sun et al. [84] proposed a helmet-based system for basic sound localization. A microphone module collects audio, which is amplified, band-pass filtered, and compared to a threshold. When the amplitude exceeds the threshold, the microcontroller estimates whether the sound comes from the left or right, providing a simple directional alert rather than full-scene analysis. In contrast, Rathna et al. [65] proposed a real-time auditory assistive wearable. The system similarly uses a multi-microphone array to capture sounds from different directions, enabling basic directional awareness. The authors used sensitive

electret omnidirectional microphones, and then the processing pipeline estimates which side the sound is coming from. However, the localization capability was still limited.

By 2025, research indicated that relying solely on microphones or cameras for sound localization is often insufficient. Tang and Zhang [42] presented a different approach employs ODAS with a 4-channel microphone array and a Kalman filter adapted to the environment for real-time localization. Signal processing algorithms, including TDOA and GCC-PHAT, are used to estimate sound direction, achieving azimuth errors within ±5° at a 2-m range. A similar approach in Senaha et al. [71] also used a microphone array to estimate direction of arrival, comparing actual versus detected angles for right and left sources. The results showed substantial improvement in angular accuracy when assisted by sound localization. Some other studies, however, continue to rely on sound loudness alone. Also, Gurrala et al. [40] implemented a wristband system uses a small microphone module to measure sound intensity and generates distinct vibrations when a sound exceeds a set threshold. This provides a basic spatial cue indicating distance but does not enable precise localization. Finally, Lim et al. [44] used four MAX4466 microphones for sound localization. The system activates only for sounds $\geq$ 80dB. Sound direction is determined by intensity, calculated from recorded voltage values; the microphone with the highest voltage indicates the source. With microphones positioned left, right, front, and back, tests in various indoor and outdoor environments achieved an average real-time accuracy of 77%.

Comparative Analysis:

A comparative analysis of the reviewed sound localization approaches highlights several important distinctions. In terms of sensing configuration, systems range from single-microphone setups relying on amplitude differences [40,84], to multi-microphone arrays enabling more precise spatial estimation through TDoA and phase-based techniques [41,42]. Some works further integrate multimodal sensing using cameras and LiDAR Anwaar et al. [57], although such approaches introduce additional computational complexity and hardware constraints.

Regarding localization methodologies, classical signal processing techniques such as GCC-PHAT and TDoA remain widely used due to their efficiency and real-time capability, while more advanced systems incorporate Blind Source Separation (BSS) and filtering techniques to improve performance in multi-source environments. In contrast, simpler approaches based on sound intensity or thresholding offer lower computational cost but provide limited directional resolution.

In terms of performance, there is a clear trade-off between system complexity and localization accuracy. High-precision systems using microphone arrays and adaptive filtering achieve low angular error, whereas lightweight wearable systems often exhibit reduced accuracy due to limited processing and lack of noise mitigation. Additionally, deployment considerations vary between embedded wearable platforms and robot-assisted systems, influencing both latency and usability.

2.3. Emotion Recognition

Emotion recognition is included as a distinct assistive task because understanding the emotional state of a speaker provides critical contextual information that cannot be captured through speech transcription alone.

One of the newest features used in recent research papers is speech emotion recognition, which helps deaf users understand the emotional state of the speaker and therefore interpret speech more effectively. In 2021, Ridha and Shehieb [45] proposed a CNN-based ML model trained on two open-source datasets to recognize different emotions from various languages and produce accurate emotion predictions. The detected emotion is then

displayed on the AR glasses as an emoji. The model achieved a validation accuracy of approximately 71.3%.

In 2022, Fukui et al. [46] proposed a system that detected crowd emotions by classifying cheering sounds. The system formulated environmental cheering recognition as a three-class classification task, labeling sounds as Heat-up, Normal, or Heat-down to represent excitement levels. A pretrained deep neural network (VGGish, Hershey et al. [81]) extracted audio features, while Create ML was employed to train the classification model. The system achieved 91% precision and 94% recall for the Heat-up class in soccer data, and 80% precision and 73% recall when tested on unseen martial arts audio, indicating reasonable cross-sport generalization.

In 2024, R et al. [47] used a strategy based on a Vision-and-Language Transformer, which focuses on speech and facial expressions at the same time to avoid missing information from either side and to make emotion identification easier and more stable. The system also used a privacy protection method called Federated Learning, which sent the model to the device instead of taking information and sending it to a server, so each device could learn using its own data. Additionally, the system learned from users' responses and could be customized based on experience using a method called Proximal Policy Optimization (PPO). Also in the testing phase, Deep Emotional Neural Networks (DENNs) were used for classification and performance evaluation.

By 2025, Gupta and Vishwakarma [32] used a pre-trained deep learning model based on the transformers library, specifically superb/wav2vec2-base-superb-er, to classify the speaker's emotional state based on recorded audio. This model is a specialized instance of the Wav2Vec2 architecture, chosen because it learns directly from raw waveforms using self-supervised learning. The system took the recorded sound, prepared it into a format suitable for the model, and then passed it through to the model to calculate the probability score for each emotion category and select the highest one. The model achieved reasonably good accuracy according to the results presented in the paper. Similarly, Ubur [70] incorporated an emotion recognition feature, which was used since speech-to-text alone was not sufficient to understand the context for Deaf and Hard-of-Hearing individuals. The system used a ML model to analyze both verbal and nonverbal cues, extract emotional metadata from tone, facial expressions, and gestures.

Comparative Analysis:

A comparison of emotion recognition systems reveals notable differences in sensing modalities and modeling strategies. Early approaches primarily rely on audio-only inputs, extracting features such as spectral or learned representations using CNN-based models [45]. In contrast, more recent systems adopt multimodal approaches that combine speech, facial expressions, and contextual cues to improve robustness and interpretability [47,70].

From a methodological perspective, traditional deep learning models such as CNNs and DNNs are increasingly complemented or replaced by transformer-based architectures (e.g., Wav2Vec2), which learn richer representations directly from raw audio. Additionally, the integration of adaptive learning techniques such as reinforcement learning and personalization mechanisms allows systems to evolve based on user-specific data, improving long-term performance.

In terms of deployment, some systems rely on centralized processing, while others adopt privacy-preserving approaches such as federated learning [47], highlighting a trade-off between data privacy and computational efficiency. Furthermore, performance varies depending on dataset diversity and modality integration, with multimodal systems generally achieving more stable and context-aware emotion recognition.

2.4. Sign Language Recognition

Sign language recognition (SLR) converts gestures into textual or spoken words to improve communication between hearing-impaired and hearing individuals. Over the years, various approaches have been proposed, ranging from sensor-based systems to deep learning-based vision models. Fu et al. [48] designed a data glove equipped with bending and attitude sensors, which was employed to capture detailed hand motion information. A Back Propagation (BP) NN was used for gesture recognition, achieving an overall accuracy of 95.18%. The system also converted recognized speech into 3D gesture animations using 22 human body angle data points in Unity 3D.

In 2022, alternative hardware-based approaches were explored. The work in Senthil et al. [60] replaced traditional flex sensors with an ADXL345 three-axis gyroscope capable of 360-degree rotation, improving accuracy and enabling recognition of over 30 unique inputs. The recognized gestures were mapped to pre-recorded MP3 outputs. Similarly, Prasath and Panaiyappan [49] developed a system that translated spoken words into visual sign language using an R-CNN classifier and a sign language dictionary, displaying the corresponding signs through a graphical interface.

More advanced vision-based and real-time systems emerged in 2024. Talaat et al. [50] introduced a real-time 3D Arabic avatar system designed for ArSL, combining Sign Language Translation (SLT) [92] and Recognition (SLR) [93]. The proposed ArSLGen algorithm translateed spoken or written Arabic into ArSL movements and also interpreted signs produced by deaf individuals. The system utilizes YOLOv8 for real-time gesture recognition and is trained on Arabic Alphabet Sign Language (AASL), Sign-language-detection Image (SLDI), and (ArSL), achieving a peak accuracy of 99.4% on AASL.

Recent works in [94–97] further integrated image processing and machine learning techniques to enhance recognition accuracy and execution speed, demonstrating the growing reliance on deep learning-based vision models. In addition, Tyagi et al. [51] proposed a real-time sign language recognition system to enhance safety for deaf and mute women. The system focused on 28 emergency-related gestures such as “Danger”, “Police”, and “Help”. It combined MediaPipe for multi-modal keypoint extraction, Dynamic Time Warping (DTW) for temporal alignment, and a voting algorithm for classification. The proposed approach achieved 98.45% accuracy, demonstrating the effectiveness of computer vision techniques in safety-critical SLR applications. The system is designed as a real-time solution intended for deployment in environments where marginalized communities can use it as a portable safety tool during emergencies.

Comparative Analysis:

Sensor-based systems (e.g., data gloves and inertial sensors) provide high accuracy and precise motion capture but require specialized hardware, limiting usability. In contrast, vision-based approaches using deep learning (e.g., YOLO, CNNs) offer more flexible and scalable solutions, enabling real-time recognition without wearable sensors. Recent methods increasingly combine temporal modeling and keypoint extraction to improve robustness and accuracy in dynamic gestures.

2.5. Other Tasks

Aside the previously mentioned tasks, work can be found in the recent literature on tasks like:

- Telepresence and telehealth: telepresence allows to move from video calls to immersive audio-visual communications, with recent advances in different aspects, notably robotics (Youssef et al. [98] and AlMutairi et al. [99]). Telehealth can be used to provide healthcare for patients in rural areas (Santomauro et al. [52]) for example. While not

designed for them, telehealth has been used with deaf and hard-of-hearing (DHH) patients and was shown in Liu et al. [53] to come with certain barriers in this context, showing the importance of improving its accessibility and support. Similar findings were reported in Moreland et al. [100] and Bhamjee et al. [101].

- Virtual and augmented reality: several work have been proposed for the incorporation of virtual and augmented reality technologies into the assistance of hearing impaired persons. Virtual reality has the potential of being integrated in training and rehabilitation for persons with hearing impairments Serafin et al. [54]. In Mehra et al. [55], it was suggested that augmented reality technologies can be used to augment auditory objects based on the user wants to attend. Also, mixed reality (MR) technologies can be used in avatar-based remote collaboration systems for persons with auditory disabilities, integrating automatic speech recognition (Waldow and Fuhrmann [102]).

3. Platforms

Hearing-assistive technologies have been implemented across multiple platforms, most notably wearable and mobile systems—each tailored to meet different user needs and environments. In this context, the platform refers to the hardware and computational environment used to implement and evaluate assistive systems.

These platforms influence key aspects such as processing capability, latency, and real-time performance, which are essential for sound perception tasks. The platforms discussed in the reviewed literature—including Raspberry Pi, Arduino, mobile processors, and embedded systems—demonstrate various implementations of sound perception algorithms. These prototype implementations, while not representing clinically approved systems, provide insights into the computational feasibility and trade-offs of different algorithmic approaches for real-time sound perception in resource-constrained environments.

Different reviewed work in this context are summarized in Table 2. And their percentages among the reviewed work are shown in Figure 3.

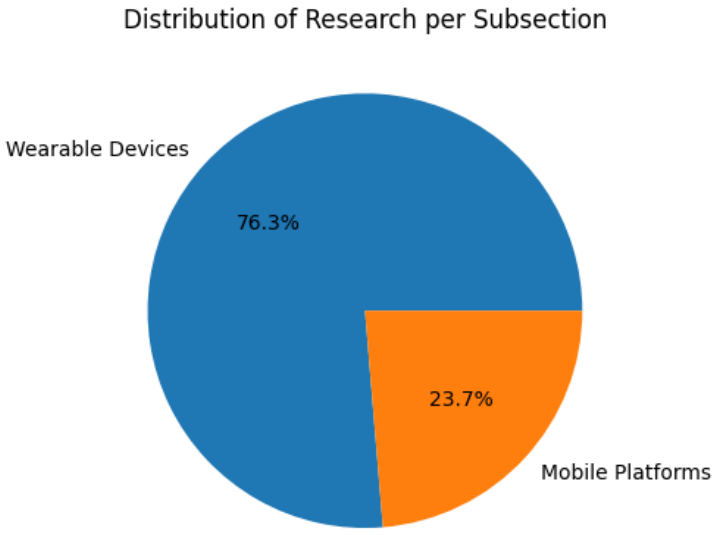


Figure 3. Distribution of reviewed citations across assistive-technology platforms for deaf and hard-of-hearing users. The chart illustrates the proportion of studies categorized by implementation platform, distinguishing between wearable devices and mobile systems. It highlights a predominant focus on wearable technologies, reflecting an emphasis on portability and real-time, unobtrusive user assistance, in contrast to the smaller subset of research addressing mobile applications and assistive robotic systems.

Table 2. Representative example platforms grouped into wearable and mobile assistive systems.

Category	Work	Platform Used	Output
Wearable	Pavlidou and Lo [27]	Arduino + computer	Vibration outputs (BLDC motors)
	Ridha and Shehieb [45]	AR glasses + mic array	Subtitles, sound events, emotions
	Abi Sen et al. [28]	Mobile app + Arduino bracelet	Vibration alerts
	Yağanoğlu [30]	Raspberry Pi wearable	Vibration feedback
	Fu et al. [48]	Data glove + Arduino	Gesture animation + voice
	Anwaar et al. [57]	Wearable vest + sensors	Directional vibration alerts
	Fukui et al. [46]	Smartwatch	Haptic notifications
	An et al. [82]	Bracelet + mobile app	Visual + vibration alerts
	Goyal and Basavarajappa [103]	BLE wearable device	Haptic communication signals
	T et al. [83]	Arduino watch	Controlled vibration
	Buhat et al. [37]	Headband + wristwatch	LCD, LEDs, vibration
	S et al. [104]	Wearable vest + sensors	Audio + GPS + alerts
	Choe et al. [41]	Mesh vest + mic array	Directional LED alerts
	Matsuo et al. [43]	Smart glasses	Localization angles
	J et al. [66]	Smart glasses + Raspberry Pi	Speech-to-text display
	Sun et al. [84]	Smart helmet	Visual + vibration alerts
	Rathna et al. [65]	Wearable mic array	Spatial awareness feedback
	S P et al. [85]	Wearable glove	OLED + vibration alerts
	Senaha et al. [71]	AR glasses	Direction + transcription
	de la Banda et al. [105]	Watch + belt	Hazard vibration alerts
Mobile	Tharwat et al. [31]	Raspberry Pi platform	Speech + speaker recognition
	Du et al. [106]	Standalone Raspberry Pi	Navigation + object detection
	Karroum et al. [86]	ESP32 + cloud CNN	Visual + haptic alerts
	Ubur [70]	AR mobile application	Visual annotations
	Gawli et al. [74]	Smart glasses + ESP32	Caption display
	Yu et al. [76]	Robot (BADA)	LED + movement alerts
	Jain et al. [35]	Smartphone	Sound recognition results
	Bao et al. [63]	Robot + camera	Motion + LED alerts
	Mahmud et al. [107]	Mobile platform	Integrated assistive services
	Talaat et al. [50]	Mobile app	3D avatar output
	Salem et al. [38]	Vehicle system (Raspberry Pi)	Dashboard display
	Tang and Zhang [42]	Robot + LiDAR	Mapping + navigation

3.1. Wearable Devices

Wearable assistive technologies have emerged as practical and innovative solutions to support people with deafness or hearing impairments in perceiving environmental and speech-related audio information.

In 2021, Pavlidou and Lo [27] explored tactile sound substitution by converting speech into vibration patterns for individuals with profound hearing loss. Using an Arduino Leonardo and a computer, speech phonemes were mapped to 3-bit vibration outputs across three Brushless DC (BLDC) motors. User evaluations showed the feasibility of a

tactile phoneme-based language. Although limited in vocabulary and adaptability, the study introduced a novel wearable approach that bypassed the auditory pathway and demonstrated the potential of haptic feedback for speech perception. During the same year, Ridha and Shehieb [45] presented smart AR glasses that combined real-time speech-to-text, environmental sound detection, and emotion recognition to support educational inclusion. The system integrated a Micro-Electro-Mechanical Systems (MEMS) microphone array, a Teensy microcontroller, and an Android AR platform. The device identified speech and directionality while displaying color-coded subtitles and sound events, achieving 71.3% accuracy in emotion recognition. It was one of the most feature-rich solutions, combining multiple audio-processing tasks into a single AR wearable. Similarly, Abi Sen et al. [28] addressed domestic communication challenges by delivering real-time haptic alerts triggered by predefined spoken keywords. The system utilized a mobile application integrated with a speech recognition module to wirelessly transmit notifications via Bluetooth to an Arduino-controlled bracelet, which generated tactile vibration feedback. In addition, Yağanoğlu [30] developed a wearable system designed to be mounted on the user's back, offering a low-cost, lightweight, and fully portable solution suitable for daily activities. It operated independently, without requiring a smartphone, internet connection, screen, or AR glasses. The hardware consists of a Raspberry Pi as the main processor, a Grove Shield for sensor connectivity, a microphone to capture environmental audio, and a vibration motor that delivers tactile feedback to the user. In contrast, Fu et al. [48] presented a complete translation system in which a deaf user wears a data glove to capture the posture of the hand and forearm. The glove integrates two types of sensors to ensure detailed motion tracking. The system also incorporates the LD3320A voice module, controlled by an Arduino Mega 2560, combining speech recognition and voice output in a single chip. This design reduced system size and allows it to be integrated into a compact wearable assembly.

In 2022, new wearable assistive technologies were introduced. In Anwaar et al. [57], the authors developed a system for hearing-impaired users that detected hazards using audio, Light Detection and Ranging (LiDAR), and a camera. In this system, TinyML with MFCC classified critical sounds like "Stop," while YOLOv4 handled object detection. Directional vibrations indicated hazard locations. Despite its effectiveness, the device was relatively heavy (2.3 kg) and was still under testing, highlighting challenges in portable, safety-critical wearable design. Furthermore, Fukui et al. [46] integrated a smartwatch interface to provide real-time haptic feedback for deaf and hard-of-hearing users during sports events. When the system detected high excitement, the smartwatch delivered vibration-based notifications, allowing users to perceive important moments without relying on auditory cues. This wearable interaction enhances awareness through tactile feedback that converts environmental audio into accessible physical cues. To conclude this year, An et al. [82] proposed an assistive technology for deaf and hard-of-hearing individuals incorporating a wearable vibration bracelet. The mobile application captured sounds over 60 dB and displayed the classification results, while the vibration bracelet, implemented using an Arduino Nano and a Bluetooth module, delivered tactile alerts through a vibration motor. This wearable component enabled intuitive real-time notification beyond visual feedback.

In 2023, Goyal and Basavarajappa [103] developed a wearable system for close-contact, home environments, enabling non-verbal communication via BLE microcontrollers and vibration motors. Caregivers could send customized haptic signals through button inputs. While it lacked sound recognition or AI, the device was low-power, affordable, and easy to use, making it a practical solution for personalized home-care communication. Also, T et al. [83] introduced a simple device that was suitable for daily use. It was a watch with an Arduino and a breadboard, low-voltage audio power amplifier (LM386 IC), a capacitor, a resistor, a power supply unit, an electret microphone capsule, an electric

motor, and connecting wires. The vibration motor used was a Linear Resonant Actuator (LRA) to ensure comfort, as it is small, lightweight, and capable of producing precise and controlled vibrations. In addition, Buhat et al. [37] proposed a wearable system consisting of a headband and a wristwatch designed for sound classification and multi-modal alerts. Using Raspberry Pi, NodeMCU ESP8266, Liquid Crystal Displays (LCDs), Light Emitting Diodes (LEDs), and vibration motors, the system provided flexible output. Its modular design was a strength, but limited outdoor accuracy, short battery life, and reliance on electret microphones constrained real-world use. Also, Chin et al. [36] aimed at road safety for deaf users by recognizing emergency sirens instantly using edge computing. The system employed a CNN trained on spectrograms of sirens and achieved 95.22% accuracy on-device. It performed classification locally, without cloud dependency. Though highly specialized, it demonstrated how deep learning could function on embedded platforms. It offered limited sound-type coverage and lacked flexibility and wearable miniaturization. Moreover, S et al. [104] developed an AI-based wearable vest for mobility and safety of visual and hearing-impaired users. It combined ultrasonic sensors with bone-conduction audio for real-time alerts, Global Positioning System (GPS) navigation, Global System for Mobile Communications (GSM) SOS, and live tracking. AI-based image and face recognition enhanced environmental awareness, while the hands-free, multi-functional design demonstrated the potential of unifying several assistive tasks in a single wearable platform. Similarly, in Choe et al. [41], EchoVest was introduced as a wearable mesh vest for environmental sound awareness, integrating a Raspberry Pi 3 and a ReSpeaker 4-mic array. The initial design used Transcutaneous Electrical Nerve Stimulation (TENS), which was later replaced with LED-based feedback for user safety. The system conveyed sound information through LED brightness and arrangement, representing distance and direction. Experimental evaluation demonstrated accurate and responsive alerts in real-world conditions.

In 2024, Matsuo et al. [43] optimized a wearable device, which consisted of smart glasses equipped with two microphones. Data processing was performed on a computer. The smart glasses lens was used to display the angle of the sound source's location in degrees, with a positive or negative sign indicating the direction. Similarly, J et al. [66] focused on offline real-time speech-to-text transcription integrated into a lightweight smart-glass platform. Using a Raspberry Pi and the Vosk ASR engine, the system converted nearby spoken language to an Organic Light Emitting Diode (OLED) display, delivering high accuracy and ultra-low latency. The offline processing ability and integration into a lightweight wearable design made it highly effective for reliable indoor communication. In Sun et al. [84], a smart wearable helmet was proposed to detect vehicle horn sounds and alert users via vibrations and visual signals. The system employed an STM32 microcontroller, an MPU9250 motion sensor, and an ensemble algorithm for horn classification, with solar power enhancing sustainability. While effective for the targeted outdoor scenario, the prototype was limited in flexibility for broader auditory scene understanding, highlighting challenges in designing versatile wearable assistive devices. Similarly, Rathna et al. [65] proposed a wearable device that provided continuous environmental sound awareness and spatial feedback for hearing-impaired users. Using a four-microphone array and embedded classifiers (KNN, SVM), the system detected speech, alarms, and other essential sounds with up to 98% accuracy in close range. While effective in combining classification and directional awareness, its real-world usability was limited by form factor and robustness. Furthermore, S P et al. [85] discussed a wearable glove-based system that detected approaching emergency sirens for 14 hearing-impaired drivers. It used an ESP32 to process MFCC and spectrogram data with a CNN, displaying results on an OLED display screen and providing vibration feedback. The device operated offline and was tailored for vehicle use,

efficiently detecting a specific class of sounds. In addition, Tharwat et al. [31] introduced a smart-worn device that enabled speech-to-text and speaker recognition using OpenAI Whisper for ASR and RF classifiers for speaker identification. The outputs were displayed on an OLED via Raspberry Pi. Despite high accuracy, dependency on Whisper and power demands could limit portability. Finally, Du et al. [106] introduced a voice-controlled wearable platform supporting navigation, audio feedback, and object detection. Using deep learning and image processing, it enhanced user awareness of surroundings and operated independently on a Raspberry Pi, highlighting early steps toward intelligent, voice-driven wearable assistive systems.

Recently in 2025, Karroum et al. [86] aimed to improve user safety and awareness by identifying common environmental sounds, providing instant visual and vibration alerts. Using an ESP32 and an INMP441 microphone, the system enhanced ambient audio detection and transmitted it to a cloud-hosted CNN model trained on UrbanSound8K and real data. Sound labels were shown on a screen while haptic alerts were triggered. It was useful for environmental awareness, but its cloud-based architecture reliance and lack of on-device inference may limit use in offline settings or latency-critical applications. In Ubur [70], the wearable device was a Unity-powered AR application using the Mixed Reality Toolkit 3 (MRTK3), with a camera and microphone to capture real-time input for both verbal and non-verbal signals. The output was displayed as tone tags, visual highlights, and gesture-based annotations, which help the user focus on the speaker without being visually scattered. Also, the user can adjust settings such as font size, contrast, and placement. The device received successful feedback from participants in terms of usability and user focus. Furthermore, in Senaha et al. [71] an AR solution integrated speaker identity, transcription, and sound direction, using microphone arrays and VAD to enhance real-time speech capture. It used Raspberry Pi, Whisper ASR, and x-vector features to show directional arcs. Sound source direction was visualized using an Augmented Reality Caption Overlay (arc overlay). The WER dropped by 33% with VAD, and speaker recognition accuracy reached 87.7% using 6-s clips. It offered an advanced visual display model for real-time auditory scene representation, best suited indoors. Moreover, de la Banda et al. [105] introduced a smart wearable device for deaf people aimed at hazard awareness by detecting environmental sounds. The wearable system used a CNN-based sound classification approach to detect hazardous non-speech sounds, including car horns, fire alarms, and sirens. The device delivers real-time vibration alerts to warn users of perceived dangers without relying on auditory or visual cues. It was implemented as a wearable watch and belt that supports continuous safety monitoring and timely hazard awareness. Also, Garg et al. [108] proposed a glove that works as an assistive device to enhance communication for hearing- and speech-impaired individuals. The system enabled bidirectional communication by translating sign language into speech and converting spoken words into text using flex sensors and an MPU6050 accelerometer to capture hand movements, which are processed by a machine learning algorithm to recognize gestures accurately. This glove facilitates portable, inclusive, and real-time communication in daily environments. Similarly, A et al. [109] reviewed smart glove technologies for speech- and hearing-impaired users. The paper reviewed multiple approaches for sign language translation and gesture recognition, including microcontrollers, sensors, and machine learning-driven methods. The findings highlight the importance of balancing affordability, accuracy, and reliability in wearable communication solutions. Also, the assistive device proposed by Gawli et al. [74] contains an ESP32 Dev Board as the main processor, a microphone module, an OLED display, a mirror, a magnifying glass lens, and a 3D glasses frame. The device was designed as smart glasses, where a mirror is positioned in front of the user's eye within the frame, and a magnifying glass lens is placed between

them to provide easier visibility. Finally, Lim et al. [44] implemented a wearable device that consisted of three parts: headband, armband, and wristband. The headband has four MAX4466 microphones, and the armband has a USB microphone and a Raspberry Pi. The wristband has the TTGO T-display and vibration motor. Every time the TTGO-T display receives data, it sends a vibration signal through the vibration motor.

Comparative Analysis:

Wearable assistive systems offer several advantages, including portability, real-time interaction, and continuous user support. Devices such as smart glasses, vests, and wristbands enable direct and immediate feedback through visual or haptic channels, making them highly suitable for dynamic environments and daily activities. However, these systems often face limitations related to battery life, computational constraints, and hardware complexity. In some cases, bulky designs or reliance on external processing (e.g., cloud-based inference) may reduce usability and independence.

In terms of real-world applicability, lightweight and standalone wearable systems demonstrate strong potential for everyday use, particularly in indoor or controlled environments. Conversely, multi-sensor or multimodal wearable platforms, while offering enhanced functionality, may be less practical due to increased weight, cost, and energy consumption.

Collectively, these reviewed wearable systems demonstrate progressive advancement toward translating acoustic information into accessible formats through tactile, visual, and intelligent processing technologies.

3.2. Mobile Platforms

Instead of relying solely on wearable technologies, several studies have explored non-wearable assistive systems that support deaf and hard-of-hearing users without requiring them to carry or wear additional devices. In research Yu et al. [76], the authors proposed BADA, a robotic assistant inspired by hearing dogs, designed to recognize environmental events and provide deaf-friendly notifications. Arduino was used as a low-level controller connected to the motors and micro switches, allowing the robot to move and change direction when it collides with an object. A Dynamixel Motor was used to control the angle of the camera. The system includes a Coral USB Accelerator for fast AI processing, and an Adafruit LED Dot Matrix to display direction and notification. A ReSpeaker Microphone Array is used to capture input sound for event recognition. The main processor is UP² Board, and an RPLiDAR-A2 sensor is used for discovering and mapping the environment using 2D SLAM (Simultaneous Localization and Mapping) (G-Mapping). All these components work together to ensure the successful operation of the BADA robot, as demonstrated by the experimental results.

Other studies adopted mobile-based assistive solutions instead of standalone robotic platforms. In 2022, Fukui et al. [46] implemented their assistive application on a smartphone-based system that continuously captures environmental audio through the built-in microphone and performs real-time sound analysis. The mobile interface displays the detected class label and confidence level, supported by color-coded visual feedback that reflects different excitement levels to convey the atmosphere of the stadium. In addition, Jain et al. [35] implemented a real-time sound recognition system on Android smartphones, utilizing the device's built-in microphone and processing capabilities for on-device audio analysis. The system captures 16 kHz audio and processes it in 1-s buffers, enabling continuous monitoring directly on the mobile hardware without reliance on external servers. This hardware-based implementation demonstrates the feasibility of using resource-constrained mobile devices for real-time environmental sound monitoring. An et al. [82] developed an assistive mobile application and a simple vibration bracelet for deaf and hard-of-hearing

individuals. The workflow of the system is that the mobile phone captures sounds over 60 dB, then processes them to display the results in the application and send an alert to the vibration bracelet connected to the app via Bluetooth.

Similarly, by 2023, Bao et al. [63] proposed a hearing-dog robot that attract users' attention visually. The robot utilizes face detection to ensure that the user has noticed the output. Hiwonder's TonyPi robot was used as the experiment robot. The robot performs three types of body movements depending on event urgency level: bowing for low urgency, twisting the waist for normal urgency, and moving forward to do left and right hooks, and moving back for high urgency. In addition, the robot displayed corresponding words or numbers, such as the police number on an LED screen. The robot also used a camera to detect obstacles along its path in order to safely avoid them. Moreover, Mahmud et al. [107] provided a broader perspective on mobile-based assistive system solutions to support communication, accessibility, and daily living for physically disabled adults. The paper introduces the concept of a centralized mobile platform to integrate and improve assistive services for users. It emphasizes existing gaps and the need for accessibility, and highlights opportunities for mobile assistive technology development. Addressing this need, Talaat et al. [50] developed a real-time Arabic Sign Language translation system was specifically designed for deployment as a mobile application, providing a smooth and efficient communication tool usable anywhere. The proposed ArSLGen algorithm translates spoken or written Arabic into dynamic 3D avatar movements displayed directly on a smartphone interface. The system also supports gesture recognition, and the Arabic Sign Language (ArSL) dataset was recorded using a mobile phone camera, enhancing its practicality and real-world applicability.

Other alternative designs were also proposed in 2025. One Such example was presented by Salem et al. [38]. The proposed system was specifically designed for deaf drivers, where four sound sensors are installed at the front and rear of the vehicle to capture and analyze environmental sounds. The system provides timely alerts to drivers through a 3.5-inch visual display on the dashboard, which shows the output and the type of detected sound as explained in the subsection on non-speech recognition. All components are connected to a Raspberry Pi 4, which enabled reliable real-time processing and fast system performance as demonstrated in their experiments. In a related study, Tang and Zhang [42] proposed a robot to help deaf people. The proposed design featured a robot equipped with an array of eight microphones, connected to two computers, one dedicated to sound source localization and the other to sound recognition, as explained separately in their section. To ensure the robot moves successfully, they incorporated features for map generation and navigation. A LiDAR sensor was employed to draw the map and determine safe paths for the robot to move without collision. The robot utilized Google's Cartographer algorithm, a state of-the-art SLAM (Simultaneous Localization and Mapping) framework, to generate a 2D map of its surroundings. The process works as follows: the LiDAR sends data to the system, which handles it in real time and uses the map generated for navigation. Their outdoor navigation experiments achieved high accuracy with no collisions in any test environment.

Comparative Analysis:

Mobile and non-wearable assistive systems offer advantages in terms of computational power, scalability, and flexibility. Smartphone-based and robotic platforms can support more complex processing tasks, such as real-time sound recognition, localization, and navigation, without strict energy or size constraints. These systems are also easier to update and integrate with existing infrastructures. However, their main limitation lies in reduced immediacy and user dependence on external devices. Unlike wearables, mobile platforms

may require the user to remain within a specific range or actively interact with the system, which can limit responsiveness in urgent situations. Additionally, robotic systems, while powerful, may introduce complexity in navigation, cost, and maintenance.

From a real-world applicability perspective, mobile platforms are well suited for structured environments such as homes, classrooms, or vehicles, where infrastructure and space allow effective deployment. However, they are generally less suitable for continuous, on-the-go assistance compared to wearable solutions.

3.3. Overall Comparison

A direct comparison between wearable and mobile assistive systems highlights a clear trade-off between immediacy and computational capability. Wearable devices excel in providing continuous, real-time, and user-centered assistance, making them ideal for dynamic and outdoor scenarios. In contrast, mobile platforms offer higher processing power and system flexibility, enabling more advanced functionalities but often at the cost of reduced portability and responsiveness. Therefore, the choice between wearable and mobile systems depends on the intended application context, with hybrid approaches combining both paradigms representing a promising direction for future assistive technologies.

3.4. Linking Tasks and Platforms

While the proposed classification distinguishes between *tasks* and *platforms*, these two dimensions are inherently interconnected. The performance, design, and feasibility of a given task are strongly influenced by the underlying platform on which it is deployed.

For instance, computationally intensive tasks such as speech recognition, emotion recognition, and multimodal fusion often achieve higher accuracy when supported by platforms with greater processing power and memory resources, such as mobile devices or external computing units. However, recent advances in embedded systems and edge AI have enabled some wearable devices to perform increasingly complex computations locally. In addition, Internet of Things (IoT) architectures allow wearable systems to offload processing to connected devices or cloud services, thereby extending their computational capabilities beyond on-device limitations. Nevertheless, wearable platforms still face constraints related to energy consumption, hardware size, and real-time responsiveness, which may necessitate the use of optimized or lightweight models in certain scenarios. Similarly, tasks such as sound source localization may exhibit reduced accuracy on compact wearable systems with limited sensor configurations, while more advanced platforms (e.g., robots or multi-sensor environments) can leverage richer spatial data and signal processing techniques to enhance performance.

Therefore, tasks and platforms should not be viewed as independent categories, but rather as complementary dimensions, where platform capabilities, connectivity (e.g., IoT), and system design choices collectively shape the achievable functionality, accuracy, and real-world applicability of assistive systems.

4. Discussion

The surveyed body of work reveals a clear transition toward intelligent, multimodal, and context-aware assistive architectures, benefiting from advances in different technological domains, while simultaneously exposing important technological and methodological challenges.

4.1. From Amplification to Information Transformation

Traditional assistive technologies for hearing loss primarily aim to restore auditory function through signal amplification or electrical stimulation. In contrast, non-prosthetic assistive systems adopt a different paradigm: they transform auditory information into alternative representations. Rather than compensating for physiological deficits, these

systems reinterpret environmental acoustic information into accessible forms such as visual representations and vibrotactile cues. This approach shifts the objective of assistance from signal restoration to situational awareness, aligning with multimodal human-machine interaction frameworks.

4.2. From Speech Recognition to Auditory Scene Understanding

Early non-prosthetic systems largely focused on speech-to-text transcription. While speech recognition remains central for communication support, recent research demonstrates a broader objective: auditory scene understanding. Modern systems increasingly integrate sound classification, sound source localization, multi-source detection, and event inference. The incorporation of deep learning has significantly improved robustness in noisy and reverberant environments. Consequently, the focus is evolving from addressing "What was said?" to understanding "What is happening in the environment?" This conceptual expansion reflects a transition from speech-centric assistance to contextual intelligence.

4.3. Wearable and Multimodal Architectures

Wearable computing is an important platform for non-prosthetic assistive systems. Smart glasses, worn devices, head-mounted displays, vibrotactile accessories, and portable processing units illustrate the importance of mobility and discretion in assistive design. Notably, augmented reality (AR) technologies and smart glasses have allowed a promising interface for non-prosthetic assistive systems targeting persons with hearing loss. These devices provide a visual layer that transforms auditory and contextual information into real-time overlays within the user's field of view, enabling continuous and hands-free access to relevant environmental and communicative cues. One of the primary applications of AR glasses is real-time speech-to-text transcription. Several studies have demonstrated the effectiveness of this approach in improving communication accessibility in daily interactions, classrooms, and professional settings. AR-based assistive devices increasingly incorporate computer vision and machine learning techniques to enhance environmental awareness. For example, systems have been proposed to detect and classify environmental sounds and display corresponding visual cues in real time (Asakura [110]).

Another notable field with advances benefiting assistive systems is wearable haptic technologies. Indeed, they use tactile stimuli to provide information to users [24] and it became possible to address the concept of "haptic hearing aids" (Fletcher et al. [111]) allowing to improve outcomes for persons using, and not using cochlear implants. The advances in haptic actuator technologies were reported in Fletcher et al. [111] to enable new audio-to-tactile conversion methods. Such devices are not only speech-related, where applications like emotional cue translation can be found [112], but can also be used in music experience [113,114]. The evolutions in Internet of Things (IoT) have also allowed the development and improvement of different technologies for persons with hearing losses [115,116], aside intelligent sensors [117].

Recent developments in wearable and multimodal architectures emphasize spatial sound visualization through augmented reality, tactile feedback, and low-latency processing. In this context, multimodal sensor fusion, combining audio, vision, and inertial sensing, can enhance reliability under acoustically challenging conditions. However, trade-offs persist between computational complexity, energy consumption, latency, and ergonomic constraints. Achieving a balance between these factors remains an engineering challenge. The diverse hardware platforms reviewed in Section 3—including Raspberry Pi, Arduino, mobile processors, and embedded systems—demonstrate the feasibility of deploying sound perception algorithms across varying computational constraints. These prototype implementations provide empirical evidence of the trade-offs between algorithm complexity,

real-time performance, and energy efficiency that must be addressed in wearable assistive device design. While clinical deployment requires certified embedded systems and medical-grade validation, the computational insights from these research prototypes inform the design requirements and performance baselines for next-generation wearable assistive technologies.

4.4. Artificial Intelligence and Open Challenges

Artificial intelligence is highly involved in non-prosthetic systems. Deep learning models enable speech recognition, environmental sound classification, and scene interpretation. Advances in deep learning have enabled high accuracies in several of the tasks reported in Section 2. Indeed, advances in deep learning allow benefits in different areas of digital healthcare [115], in addition to other areas like sign language interpretation, speech recognition and text-to-speech synthesis [118]. And aside robustness in speech recognition for non-prosthetic assistive systems, deep learning has been used in sound classification and detection of important and alarming sounds [82,115,119]. Deep learning has also been relied on in hearing aids [120–122], notably for signal enhancement. Modern computational platforms and connectivity increasingly allow real-time deployment in wearable platforms. Despite substantial progress, several challenges remain. These include energy limitations in portable devices, strict real-time constraints, dataset bias affecting generalization, and privacy concerns arising from continuous environmental audio monitoring. Future research must therefore prioritize explainability, privacy-preserving computation, and adaptive personalization mechanisms to ensure sustainable and ethical deployment.

4.5. Evaluation and User Acceptance

Non-prosthetic systems may alleviate barriers associated with traditional prosthetic devices, including stigma, discomfort, cost, and maintenance complexity. However, many proposed solutions remain validated primarily in controlled laboratory settings. There is a clear need for standardized evaluation metrics, long-term usability studies, and ecologically valid real-world trials. Furthermore, personalization strategies must be strengthened to accommodate diverse user preferences, cultural contexts, and degrees of hearing loss. Addressing these gaps is essential for transitioning from experimental prototypes to reliable daily-life assistive technologies.

4.6. Future Research Directions

Future research should focus on developing more energy-efficient and low-latency models suitable for continuous real-time operation on wearable platforms. In addition, personalization mechanisms that adapt to individual user preferences, hearing conditions, and environmental contexts remain an important area for improvement. Another key direction is the integration of multimodal sensing and feedback, combining audio, visual, and haptic information to enhance situational awareness. Furthermore, there is a need for large-scale, real-world validation studies involving diverse user populations to ensure robustness, usability, and long-term acceptance. Finally, hybrid architectures that combine wearable and mobile systems may offer a promising pathway toward more comprehensive and scalable assistive solutions.

From a practical perspective, the design of assistive systems must balance accuracy, responsiveness, and user comfort. Lightweight and unobtrusive wearable devices are more likely to be adopted in daily life, while minimizing battery consumption and ensuring reliable performance in noisy environments. Privacy considerations are also critical, particularly for systems relying on cloud-based processing or multimodal data collection. Therefore, future implementations should prioritize on-device processing and secure data handling. Additionally, user-centered design and accessibility considerations must guide

system development to ensure that solutions are intuitive, customizable, and aligned with real-world user needs.

5. Conclusions

While initially aiming to provide persons with hearing loss with information extracted from their auditory environments, non-prosthetic assistive technologies are increasingly transitioning toward integrated, multimodal systems that combine sound signal processing tasks with contextual interpretation. Such a shift reflects a move away from isolated functionalities toward intelligent assistive solutions capable of enhancing environmental awareness and interaction. At the same time, system capabilities are strongly influenced by the underlying platforms, highlighting the importance of jointly considering both dimensions in system development. This is particularly important for ensuring that assistive systems remain both effective and practical, where functionality and performance must be balanced with portability, responsiveness, and user comfort. In addition, extensive real-world validation is important to address, in order to guarantee robustness and long-term usability beyond controlled settings.

Future advancements in this field, driven by advancements in fields like AI, wearable technologies, mobile computing and IoT, are expected to make assistive systems more adaptive, robust, and seamlessly integrated into complex everyday environments. These developments increase their usability and acceptability, thus promising to significantly enhance inclusiveness and improve quality of life by enabling greater independence and social participation for individuals with hearing loss.

Author Contributions: Conceptualization, K.Y., G.E.M., S.S., T.B. and S.A.K.; methodology, K.Y., G.E.M., S.S., T.B. and S.A.K.; software, R.A., F.A., M.A., G.A. and A.A.; validation, R.A., F.A., M.A., G.A., A.A., K.Y., G.E.M., S.S., T.B. and S.A.K.; formal analysis, R.A., F.A., M.A., G.A., A.A. and K.Y.; investigation, R.A., F.A., M.A., G.A., A.A. and K.Y.; resources, R.A., F.A., M.A., G.A., A.A. and K.Y.; data curation, R.A., F.A., M.A., G.A., A.A., K.Y., S.S. and S.A.K.; writing—original draft preparation, R.A., F.A., M.A., G.A., A.A., K.Y., S.S. and S.A.K.; writing—review and editing, R.A., F.A., M.A., G.A., A.A., K.Y., G.E.M., S.S., T.B. and S.A.K.; visualization, R.A., F.A., M.A., G.A., A.A. and K.Y.; supervision, K.Y., G.E.M., S.S., T.B. and S.A.K.; project administration, K.Y., G.E.M., S.S., T.B. and S.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Deafness and Hearing Loss. Available online: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed on 25 January 2026).
2. Zahnert, T. The Differential Diagnosis of Hearing Loss. *Dtsch. Ärzteblatt Int.* **2011**, *108*, 433–444 . [CrossRef]
3. Aldè, M.; Ambrosetti, U.; Aldè, S. The Ongoing Challenges of Hearing Loss: Stigma, Socio-Cultural Differences, and Accessibility Barriers. *Audiol. Res.* **2025**, *15*, 46 . [CrossRef]
4. Anastasiadou, S.; Al Khalili, Y. Hearing Loss. Available online: https://www.ncbi.nlm.nih.gov/books/NBK542323/ (accessed on 25 January 2026).
5. Hearing Loss: A Common Problem for Older Adults. Available online: https://www.nia.nih.gov/health/hearing-and-hearing-loss/hearing-loss-common-problem-older-adults (accessed on 25 January 2026).
6. Treatment for Hearing Loss—Hearing Health Foundation. Available online: https://hearinghealthfoundation.org/treating-hearing-loss (accessed on 25 January 2026).

7. Portelli, D.; Lombardo, C.; Loteta, S.; Galletti, C.; Azielli, C.; Ciodaro, F.; Mento, C.; Aguennouz, M.; Rosa, G.D.; Alibrandi, A.; et al. Exploring the Hearing Improvement and Parental Stress in Children with Hearing Loss Using Hearing Aids or Cochlear Implants. *J. Clin. Med.* **2025**, *14*, 2. [CrossRef] [PubMed]
8. Haynes, D.; Young, J.; Wanna, G.; Glasscock, M. Middle Ear Implantable Hearing Devices: An Overview. *Trends Amplif.* **2009**, *13*, 206–214. [CrossRef] [PubMed]
9. Deep, N.; Choudhury, B.; Roland, J. Auditory Brainstem Implantation: An Overview. *J. Neurol. Surg. Part B Skull Base* **2019**, *80*, 203–208. [CrossRef]
10. Hearing Aids—Styles/Types & How They Work | NIDCD. Available online: https://www.nidcd.nih.gov/health/hearing-aids (accessed on 25 January 2026).
11. Sutton, A.E.; Krogmann, R.J.; Al Khalili, Y. Cochlear Implants—StatPearls—NCBI Bookshelf. Available online: https://www.ncbi.nlm.nih.gov/books/NBK544280/ (accessed on 25 January 2026).
12. Shawkey, E.C.; Johns, J.D.; Kocharyan, A.; Corle, B.; Woolf, E.; Parks, A.; Briggs, S.E. Recent Advances in Cochlear Implantation. *J. Otorhinolaryngol. Hear. Balance Med.* **2025**, *6*, 9. [CrossRef]
13. Lu, Q.; Husein, M.; Jeyakumar, A. Review of Reported Adverse Events Associated with Prescription Hearing Aids in the Manufacturer and User Facility Device Experience (MAUDE) Database. *Cureus* **2025**, *17*, e76737. [CrossRef]
14. Hearing Aid Benefits and Limitations | FDA. Available online: https://www.fda.gov/medical-devices/hearing-aids/hearing-aid-benefits-and-limitations (accessed on 25 January 2026).
15. Marcos-Alonso, S.; Almeida-Ayerve, C.N.; Monopoli-Roca, C.; Coronel-Touma, G.S.; Pacheco-López, S.; Peña-Navarro, P.; Serradilla-López, J.M.; Sánchez-Gómez, H.; Pardal-Refoyo, J.L.; Batuecas-Caletrío, Á. Factors Impacting the Use or Rejection of Hearing Aids—A Systematic Review and Meta-Analysis. *J. Clin. Med.* **2023**, *12*, 4030. [CrossRef]
16. Neagoș, C.; Nenec, B.; Neagos, A.; Sin, A. Complications in cochlear implant surgery: A comprehensive review. *J. Med. Life* **2025**, *18*, 939–945. [CrossRef]
17. Huber, M. Cochlear implant-specific risks should be considered, when assessing the quality of life of children and adolescents with hearing loss and cochlear implants–not just cochlear implant-specific benefits–Perspective. *Front. Neurosci.* **2022**, *16*, 985230. [CrossRef]
18. Calvino, M.; Sánchez-Cuadrado, I.; Gavilán, J.; Lassaletta, L. Long-Term Non-Users of Transcutaneous Auditory Implants: Thirty Years of Experience at a Single Institution. *Int. J. Environ. Res. Public Health* **2023**, *20*, 6201. [CrossRef]
19. Kushalnagar, P.; Paludneviciene, R.; Kushalnagar, R. Video Remote Interpreting Technology in Health Care: Cross-Sectional Study of Deaf Patients' Experiences. *JMIR Rehabil. Assist. Technol.* **2019**, *6*, e13233. [CrossRef]
20. Yabe, M. Healthcare providers' and deaf patients' interpreting preferences for critical care and non-critical care: Video remote interpreting. *Disabil. Health J.* **2020**, *13*, 100870. [CrossRef]
21. Kim, J.; Kim, C. A Review of Assistive Listening Device and Digital Wireless Technology for Hearing Instruments. *Korean J. Audiol.* **2014**, *18*, 105–111. [CrossRef]
22. Fastelli, A.; Clignon, G.; Corasaniti, D.; Orzan, E. Speech-to-Text Captioning and Subtitling in Schools: The Results of a SWOT Analysis. *Audiol. Res.* **2025**, *15*, 105. [CrossRef]
23. Shezi, M.; Ade-Ibijola, A. Deaf Chat: A Speech-to-Text Communication Aid for Hearing Deficiency. *Adv. Sci. Technol. Eng. Syst. J.* **2020**, *5*, 826–833. [CrossRef]
24. Ramones, A.; del Rio, M. Recent Developments in Haptic Devices Designed for Hearing-Impaired People: A Literature Review. *Sensors* **2023**, *23*, 2968. [CrossRef]
25. Orzech, G.; Luo, Y.; Huang, G. Haptic Technology for Hearing Loss: A Systematic Review of Technical Feasibility, Usability, and User Experience. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2024**, *68*, 1537–1539. [CrossRef]
26. Snir, A.; Cieśla, K.; Vekslar, R.; Amedi, A. Highly compromised auditory spatial perception in aided congenitally hearing-impaired and rapid improvement with tactile technology. *iScience* **2024**, *27*, 110808. [CrossRef] [PubMed]
27. Pavlidou, A.; Lo, B. Artificial ear—A wearable device for the hearing impaired. In *Proceedings of the 2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4. [CrossRef]
28. Abi Sen, A.A.; S Aljohani, A.A.; Bahbouh, N.M.; Alhaboob, O. Designing a Smart Bracelet based on Arduino for Deaf Parents to Interact with their Children. In Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 17–19 March 2021; pp. 380–384.
29. Yamamoto, K.; Suzuki, I.; Shitara, A.; Ochiai, Y. See-Through Captions: Real-Time Captioning on Transparent Display for Deaf and Hard-of-Hearing People. In Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY, USA, 18–22 October 2021. [CrossRef]
30. Yağanoğlu, M. Real time wearable speech recognition system for deaf persons. *Comput. Electr. Eng.* **2021**, *91*, 107026. [CrossRef]
31. Tharwat, M.; Wardak, Y.; Balbaid, S.; Radin, E. Wearable Device With Speech and Voice Recognition for Hearing-Impaired People. In Proceedings of the 2024 21st Learning and Technology Conference (L&T), Jeddah, Saudi Arabia, 15 Jamuary 2024; pp. 221–226. [CrossRef]

32. Gupta, M.; Vishwakarma, V. Development of a Speech-to-Text Emotion and Speaker Recognition System for Individuals with Hearing Impairments. In Proceedings of the 2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC), Putrajaya, Malaysia, 29 October 2025; pp. 2012–2020. [CrossRef]
33. Peddi, M.; Vardhan, S.H.V.H.; Peddi, G.; Ponnam, S.; Rani, C.; Rajesh Kumar, M. Real-time Audio Recognition for Hearing Impaired. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT), Wuhan, China, 13–15 September 2024; pp. 1–6. [CrossRef]
34. Adnan Habib, M.; Raiyan Arefeen, Z.; Hussain, A.; Rownak Shahriyer, S.; Islam, T.; Rahman, R.; Zavid Parvez, M. Sound Classification using Deep Learning for Hard of Hearing and Deaf People. In Proceedings of the 2022 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 20–22 July 2022; pp. 184–190. [CrossRef]
35. Jain, D.; Huynh Anh Nguyen, K.; M. Goodman, S.; Grossman-Kahn, R.; Ngo, H.; Kusupati, A.; Du, R.; Olwal, A.; Findlater, L.; E. Froehlich, J. Protosound: A personalized and scalable sound recognition system for deaf and hard-of-hearing users. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 30 April–5 May 2022; pp. 1–16.
36. Chin, C.L.; Lin, C.C.; Wang, J.W.; Chin, W.C.; Chen, Y.H.; Chang, S.W.; Huang, P.C.; Zhu, X.; Hsu, Y.L.; Liu, S.H. A Wearable Assistant Device for the Hearing Impaired to Recognize Emergency Vehicle Sirens with Edge Computing. *Sensors* **2023**, *23*, 7454. [CrossRef]
37. Buhat, R.G.A.; Capuno, C.B.; De Guzman, R.A.D.; Pascua, M.D.; Tan, D.J.S.; Gallego, M.P.; Noriega, M.E.A. AICanHear: Assistive Alerting Device for Deaf and Hard-of-Hearing People (DHH). In *Proceedings of the 2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
38. Salem, O.; Mehaoua, A.; Boutaba, R. Vehicle Sound Recognition Assistance in IoT Systems for Hearing-Impaired Drivers. *IEEE Internet Things Mag.* **2025**, *8*, 14–21. [CrossRef]
39. Munasinghe, T.; Dulanjani, Y. AI Powered Context Aware Emergency Sound Recognition System for Enhancing Situational Awareness in Hearing Impaired Individuals. In Proceedings of the 2025 7th International Conference on Advancements in Computing (ICAC), Sri Jayawardenepura Kotte, Sri Lanka, 9–10 December 2025; pp. 1–6. [CrossRef]
40. Gurrala, V.K.; Talasila, S.; Kumar, C.K. Deaf Guard: Wristband Innovation For Hearing Impaired. In Proceedings of the 2025 6th International Conference for Emerging Technology (INCET), Belagavi, India, 22–24 May 2025; pp. 1–5. [CrossRef]
41. Choe, J.; Sood, S.; Park, R. EchoVest: Real-Time Sound Classification and Depth Perception Expressed through Transcutaneous Electrical Nerve Stimulation. *arXiv* **2023**, arXiv:2307.04604.
42. Tang, W.; Zhang, B. Sound Source Analysis for a Hearing Assistance Robot. In Proceedings of the 2025 9th International Conference on Robotics and Automation Sciences (ICRAS), Osaka, Japan, 27–29 June 2025; pp. 407–411. [CrossRef]
43. Matsuo, A.; Itami, T.; Yoneyama, J. 360° Sound Localization Support System for Deaf and Hard-of-Hearing People Using Smartglasses Equipped with Two Microphone. In Proceedings of the 2024 IEEE/SICE International Symposium on System Integration (SII), Ha Long, Vietnam, 8–11 January 2024; pp. 295–300. [CrossRef]
44. Lim, V.Y.; Chua, H.S.; Mohmad, S.; Lau, K.B; Tan, S.J. Emergency Sound Recognition and Direction Indication Using Machine Learning for Individuals with Hearing Loss. *J. Kejuruter.* **2025**, *37*, 3035–3043. [CrossRef]
45. Ridha, A.M.; Shehieb, W. Assistive Technology for Hearing-Impaired and Deaf Students Utilizing Augmented Reality. In *Proceedings of the 2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–5. [CrossRef]
46. Fukui, H.; Kim, J.E.; Bessho, M. Supporting Deaf and Hard-of-Hearing People to Watch Sports by Detecting Excitement using Mobile and Wearable Devices. In *Proceedings of the 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*; IEEE: Piscataway, NJ, USA, 2022; pp. 716–718. [CrossRef]
47. Alexander, R.; Sinduja, R. Deaf and Mute Communication Transformation: A Unified Framework for Real-Time Scribing and Emotion Recognition. In Proceedings of the 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 11–13 December 2024; pp. 1724–1729. [CrossRef]
48. Fu, Q.; Fu, J.; Zhang, S.; Li, X.; Guo, J.; Guo, S. Design of Intelligent Human-Computer Interaction System for Hard of Hearing and Non-Disabled People. *IEEE Sens. J.* **2021**, *21*, 23471–23479. [CrossRef]
49. Prasath, A.; Panaiyappan, A. Design of an integrated learning approach to assist real-time deaf application using voice recognition system. *Comput. Electr. Eng.* **2022**, *102*, 108145. [CrossRef]
50. Talaat, F.M.; El-Shafai, W.; Soliman, N.F.; Algarni, A.D.; Abd El-Samie, F.E.; Siam, A.I. Real-time Arabic avatar for deaf-mute communication enabled by deep learning sign language translation. *Comput. Electr. Eng.* **2024**, *119*, 109475. [CrossRef]
51. Tyagi, A.; Prasanth, V.S.; Lunawat, P.; Ashok, S. Real-Time Sign Language Gesture Recognition for Women's Safety Using Dynamic Time Warping and Voting Algorithms. *Frankl. Open* **2025**, *13*, 100428. [CrossRef]
52. Santomauro, C.; Best, D.; Wray, B.; Burgmann, F.; Pilotto, T.; Pearce, S.; McLanders, M. A new reality for telehealth: A simulation-based comparison of wearable mixed reality with videoconferencing for clinician-to-clinician telehealth. *Digit. Health* **2025**, *11*, 20552076251388404. [CrossRef]

53. Liu, A.W.; Yi, S.J.; Chari, D.A. Telehealth utilization and perceptions among deaf or hard of hearing adults: A cross-sectional analysis of the HINTS6 national dataset. *Am. J. Otolaryngol.* **2025**, *46*, 104716. [CrossRef]
54. Serafin, S.; Adjorlu, A.; Percy-Smith, L.M. A Review of Virtual Reality for Individuals with Hearing Impairments. *Multimodal Technol. Interact.* **2023**, *7*, 36. [CrossRef]
55. Mehra, R.; Brimijoin, O.; Robinson, P.; Lunner, T. Potential of Augmented Reality Platforms to Improve Individual Hearing Aids and to Support More Ecologically Valid Research. *Ear Hear.* **2020**, *41*, 140S–146S. [CrossRef]
56. Salamon, J.; Jacoby, C.; Bello, J.P. A Dataset and Taxonomy for Urban Sound Research. In *Proceedings of the 22nd ACM International Conference on Multimedia*; ACM: New York, NY, USA, 2014. [CrossRef]
57. Anwaar, T.; Arun Srinivasan, V.K.; Nandha Kizor, V.; Rajiv, N. Design of an Assistive Technology Wearable Vest for persons with Hearing Disability. In Proceedings of the 2022 10th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, 15–18 November 2022; pp. 347–352. [CrossRef]
58. Youssef, K.; Argentieri, S.; Zarader, J.L. Binaural speaker recognition for humanoid robots. In Proceedings of the 2010 11th International Conference on Control Automation Robotics & Vision, Singapore, 7–10 December 2010; pp. 2295–2300. [CrossRef]
59. Youssef, K.; Breteau, B.; Argentieri, S.; Zarader, J.L.; Wang, Z. Approaches for Automatic Speaker Recognition in a Binaural Humanoid Context. In Proceedings of the European Symposium on Artificial Neural Networks 2011, Bruges, Belgium, 27–29 April 2011.
60. Senthil, K.R.; Jayaram, B.; Srimathi, B.; Narmatha, T. A Novel System For Blind, Deaf And Dumb People Assistant Using Raspberry Pi. In *Proceedings of the 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
61. Staš, J.; Juhár, J.; Lojka, M.; Durkáč, M.; Hudzik, F. Hear IT—A Mobile Assistive Technology for Hearing Impaired People in Slovak. In Proceedings of the 2023 13th International Conference on Advanced Computer Information Technologies (ACIT), Wrocław, Poland, 21–23 September 2023; pp. 426–431. [CrossRef]
62. Povey, D.; Ghoshal, A.; Boulianne, G.; Burget, L.; Glembek, O.; Goel, N.; Hannemann, M.; Motlicek, P.; Qian, Y.; Schwarz, P.; et al. The Kaldi speech recognition toolkit. In *Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*; IEEE Signal Processing Society: Piscataway, NJ, USA, 2011.
63. Bao, M.; Li, S.; Jin, N.; Zhang, Q. Visual-Based Hearing-Dog Robots For Users With Hearing Impairment. In Proceedings of the 2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Suzhou, China, 2–4 November 2023; pp. 136–145. [CrossRef]
64. Errattahi, R.; El Hannani, A.; Ouahmane, H. Automatic Speech Recognition Errors Detection and Correction: A Review. *Procedia Comput. Sci.* **2018**, *128*, 32–37. . [CrossRef]
65. Rathna, R.; Maria Anu, V.; Mishra, S. Real-time Smart Auditory Assistive Wearable (RESAAW) for People with Different Degrees of Hearing Loss. *Rev. D'Intelligence Artif.* **2024**, *38*, 1319–1325. [CrossRef]
66. Xavier, A.; Mubarak, H.; Ashfaq, K.K.; Abdulla, S.M. Ear Assist for Hearing Impaired. In Proceedings of the 2024 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR), Kuttippuram, India, 25–26 September 2024; pp. 1–6. [CrossRef]
67. Yadava, T.; Nagaraja, B.G.; Reddy, S.; Rohan, K.; Mohamed, L.M. Advancements in Speech-to-Text Systems for the Hearing Impaired. In *Proceedings of the 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [CrossRef]
68. Coban, E.B.; Syed, A.R.; Pir, D.; Mandel, M.I. Towards large scale ecoacoustic monitoring with small amounts of labeled data. In *Proceedings of the 2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*; IEEE: Piscataway, NJ, USA, 2021; pp. 181–185.
69. Belmadoui, S. Arabic Sign Language (Unaugmented) Dataset. 2024. Available online: https://www.kaggle.com/datasets/sabribelmadoui/arabic-sign-language-unaugmented-dataset (accessed on 15 February 2024).
70. Ubur, S.D. Augmenting Captions with Emotional Cues: An AR Interface for Real-Time Accessible Communication. *arXiv* **2025**, arXiv:2504.17171. [CrossRef]
71. Senaha, T.; Midorikawa, R.; Baba, T.; Asaka, T. Bridging the Auditory Gap: AR Smart Glasses for Real-Time Speech-To-Text and Directional Audio Visualization for the Hearing-Impaired. *IEICE Commun. Express* **2025**, *14*, 279–282. [CrossRef]
72. Chandrashekar, H.M.; Anushree, T.R.; Chaitra, T.P.; Meghana, K.; Monisha, T.U. Automatic Recognition of Hearing-Impaired Children's Kannada Speech with Limited Vocabulary. In Proceedings of the 2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India, 21–22 March 2025; pp. 1–5. [CrossRef]
73. Rusmitha, J.; Srihari, G. Deep Learning and Computer Vision based Smart Intercom Device for the Deaf and Mute People. In Proceedings of the 2025 International Conference on Visual Analytics and Data Visualization (ICVADV), Tirunelveli, India, 4–6 March 2025; pp. 1227–1232. [CrossRef]

74. Gawli, P.; Patil, P.; Palse, V.; Parsewar, P.; Amilkanthwar, P. IOT Based Smart Glasses with Real-Time Speech to Text. In *Proceedings of the 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–4. [CrossRef]
75. Hughes, S.E.; Wu, L.Y.; Ma, L.J.; Jain, D.; McKee, M.M. Assessing the Role of Medical Caption Technology to Support Physician-Patient Communication for Patients With Hearing Loss: Mixed Methods Pilot Study. *JMIR Rehabil. Assist. Technol.* **2026**, *13*, e79073. [CrossRef] [PubMed]
76. Yu, H.; Lee, H.; Bae, J.; Kim, M.; Choi, S.; Hwang, J. The Development of a Social Robot Accessible to the Deaf. In Proceedings of the Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, New York, NY, USA, 8–11 March 2021; pp. 634–635. [CrossRef]
77. Goodman, S.M.; Liu, P.; Jain, D.; McDonnell, E.J.; Froehlich, J.E.; Findlater, L. Toward User-Driven Sound Recognizer Personalization with People Who Are d/Deaf or Hard of Hearing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 63. [CrossRef]
78. Carney, M.; Webster, B.; Alvarado, I.; Phillips, K.; Howell, N.; Griffith, J.; Jongejan, J.; Pitaru, A.; Chen, A. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 25–30 April 2020; pp. 1–8. [CrossRef]
79. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, *30, 4080–4090* .
80. Raghu, A.; Raghu, M.; Bengio, S.; Vinyals, O. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv* **2019**, arXiv:1909.09157.
81. Hershey, S.; Chaudhuri, S.; Ellis, D.P.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN architectures for large-scale audio classification. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2017; pp. 131–135.
82. An, J.H.; Koo, N.K.; Son, J.H.; Joo, H.M.; Jeong, S. Development on Deaf Support Application Based on Daily Sound Classification Using Image-based Deep Learning. *JOIV Int. J. Informatics Vis.* **2022**, *6*, 250–255. [CrossRef]
83. Thenmozhi, T.; Monisha, P.; Anusha, G.; Tharish, K.; Seshadhri, V.; Madeshwaren, G. Hearing Aiding System for Impaired. In Proceedings of the 2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET), Bangkok, Thailand, 21–22 January 2023; pp. 1–3. [CrossRef]
84. Sun, P.; Luo, J.; Wang, C.; Wu, D. An Improved Smart Helmet for Safe Travel of Deaf People Based on Embedded System. In *Proceedings of the 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*; IEEE: Piscataway, NJ, USA, 2024; pp. 183–188. [CrossRef]
85. Abhiram, S.P.; Umamaheswari, S.; Shybin, M.; Campbell, A.S.; Sreena, V.G.; Raji, N.R. SI-LERT (Assistive Device For Hearing Impaired Drivers). In Proceedings of the 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 8–9 August 2024; Volume 1, pp. 1516–1521. [CrossRef]
86. Karroum, R.B.; Safadi, S.; Boudargham, N.; Hamad, M. SoniWear: A Wearable Device For The Hearing-Impaired. In Proceedings of the 2025 Sixth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), Zouk Mosbeh, Lebanon, 24–26 September 2025; pp. 1–6. [CrossRef]
87. UrbanSound8K. Urban Sound Datasets. Available online: https://urbansounddataset.weebly.com/urbansound8k.html (accessed on 2 March 2026).
88. Garg, C. Generating Spatial Binaural Environmental Audio Alert on a Personal Audio Device. In Proceedings of the 2024 4th International Conference on Robotics, Automation and Artificial Intelligence (RAAI), Singapore, 19–21 December 2024; pp. 191–196. [CrossRef]
89. Grondin, F.; Michaud, F. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robot. Auton. Syst.* **2019**, *113*, 63–80. [CrossRef]
90. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef]
91. Wang, Y.X.; Zhang, Y.J. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 1336–1353. [CrossRef]
92. Núñez-Marcos, A.; Pérez-de Viñaspre, O.; Labaka, G. A Survey on Sign Language Machine Translation. *Expert Syst. Appl.* **2023**, *213*, 118993. [CrossRef]
93. Rastgoo, R.; Kiani, K.; Escalera, S. Sign Language Recognition: A Deep Survey. *Expert Syst. Appl.* **2021**, *164*, 113794. [CrossRef]
94. Alabdullah, B.; Alneil, A. Revolutionizing sign language recognition for hearing-impaired persons using ensemble of deep learning techniques with fine tuning model. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *38*, 41. [CrossRef]
95. Assiri, M.; Selim, M. Gesture recognition for hearing impaired people using an ensemble of deep learning models with improving beluga whale optimization-based hyperparameter tuning. *Sci. Rep.* **2025**, *15*, 21441. [CrossRef]

96. Arularasan, R.; Balaji, D.; Garugu, S.; Jallepalli, V.R.; Nithyanandh, S.; Singaram, G. Enhancing Sign Language Recognition for Hearing-Impaired Individuals Using Deep Learning. In Proceedings of the 2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 26–27 July 2024; pp. 1–6. [CrossRef]
97. Alotaibi, N.; Al-Dayil, R.; Aljehane, N.O.; Rizwanullah, M. Enhanced feature fusion with hand gesture recognition system for sign language accessibility to aid hearing and speech impaired individuals. *Sci. Rep.* **2026**, *16* . [CrossRef] [PubMed]
98. Youssef, K.; Said, S.; Al Kork, S.; Beyrouthy, T. Telepresence in the Recent Literature with a Focus on Robotic Platforms, Applications and Challenges. *Robotics* **2023**, *12*, 111. [CrossRef]
99. AlMutairi, S.; AlHajri, A.; Youssef, K.; Said, S.; Al Kork, S. An Immersive Telepresence System with User and Automatic Motion Control. In Proceedings of the 2025 6th International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 14–16 May 2025; pp. 1–4. [CrossRef]
100. Moreland, C.J.; Rao, S.R.; Jacobs, K.; Kushalnagar, P. Equitable Access to Telehealth and Other Services for Deaf People During the COVID-19 Pandemic. *Health Equity* **2023**, *7*, heq.2022.0115. [CrossRef] [PubMed]
101. Bhamjee, A.; le Roux, T.; Swanepoel, D.W.; Graham, M.A.; Schlemmer, K.; Mahomed-Asmail, F. Perceptions of Telehealth Services for Hearing Loss in South Africa's Public Healthcare System. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7780. [CrossRef] [PubMed]
102. Waldow, K.; Fuhrmann, A. Addressing Deaf or Hard-of-Hearing People in Avatar-Based Mixed Reality Collaboration Systems . In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Los Alamitos, CA, USA, 22–26 March 2020; pp. 594–595. [CrossRef]
103. Goyal, M.; Basavarajappa, G. A Wearable IoT Based Assistive Device to Aid Communication With Hearing Impaired. In *Proceedings of the 2023 IEEE Microwaves, Antennas, and Propagation Conference (MAPCON)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–4. [CrossRef]
104. Ezhilarasi, P.; Rajagopalan, V.G.; Ramakrishnan, H. Integrated AI Based Smart Wearable Assistive Device for Visually and Hearing-Impaired People. In Proceedings of the 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC), Online, 10–11 February 2023; pp. 1–6. [CrossRef]
105. de la Banda, S.S.Z.; Mamulang, C.J.T.; Garcia, E.R.G. Development of the SENSE - A Smart Wearable Device for Hazard Sound Detection and Vibration Feedback for Deaf Individuals. In Proceedings of the 2025 8th International Conference on Electronics Technology (ICET), Chengdu, China, 17–20 May 2025; pp. 766–773. [CrossRef]
106. Du, Y.C.; Yu, H.C.; Ciou, W.S.; Li, Y.L. A Wearable Assistive Listening Device with Immersive Function Using Sensors Fusion Method for the 3-D Space Perception. *IEEE Sens. J.* **2024**, *24*, 2108–2117. [CrossRef]
107. Mahmud, M.; Kolivand, H.; Al-Jumeily, D.; Khan, W. New Multipurpose Assistive Technology to Support Physically Disabled Adults. In Proceedings of the 2023 16th International Conference on Developments in eSystems Engineering (DeSE), Istanbul, Turkey, 18–20 December 2023; pp. 564–568. [CrossRef]
108. Garg, A.; Sharma, A.; Binjola, A.; Smriti.; Bhatia, A.; Mehto, A.; Bajaj, A. Enhancing Communication for the Hearing Impaired Through Innovative Smart Glove Technology: The GloveLingo. In Proceedings of the 2025 12th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 18–19 September 2025; pp. 1–6. [CrossRef]
109. Anitha, A.; Rajput, B.S.; Madarkhandi, A.; Naik, M.; Kumar, N.; Nishanth, V.K. Smart Gloves for Communication: A Survey on Technology for the Hearing and Speech Impaired. In Proceedings of the 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 18–20 February 2025; pp. 424–429. [CrossRef]
110. Asakura, T. Augmented-Reality Presentation of Household Sounds for Deaf and Hard-of-Hearing People. *Sensors* **2023**, *23*, 7616. [CrossRef] [PubMed]
111. Fletcher, M.; Akis, E.; Verschuur, C.; Perry, S. Improved tactile speech perception using audio-to-tactile sensory substitution with formant frequency focusing. *Sci. Rep.* **2024**, *14*, 4889. [CrossRef]
112. de Lacerda Pataca, C.a.; Hassan, S.; May, L.; Olson, M.M.; D'aurio, T.; Peiris, R.L.; Huenerfauth, M. Tactile Emotions: Multimodal Affective Captioning with Haptics Improves Narrative Engagement for d/Deaf and Hard-of-Hearing Viewers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 26 April–1 May 2025. [CrossRef]
113. Trivedi, U.; Alqasemi, R.; Dubey, R. Wearable musical haptic sleeves for people with hearing impairment. In Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments, New York, NY, USA, 5–7 June 2019; pp. 146–151. [CrossRef]
114. Cavdir, D. Development of embodied listening studies with multimodal and wearable haptic interfaces for hearing accessibility in music. *Front. Comput. Sci.* **2024**, *5*, 1162758. [CrossRef]
115. Young, F.; Zhang, L.; Jiang, R.; Liu, H.; Wall, C. A Deep Learning Based Wearable Healthcare Iot Device for AI-Enabled Hearing Assistance Automation. In Proceedings of the 2020 International Conference on Machine Learning and Cybernetics (ICMLC), Adelaide, Australia, 12–15 July 2020; pp. 235–240. [CrossRef]
116. Maashi, M.; Iskandar, H.; Rizwanullah, M. IoT-driven smart assistive communication system for the hearing impaired with hybrid deep learning models for sign language recognition. *Sci. Rep.* **2025**, *15*, 6192. [CrossRef]

117. Soares, C.; Silva Zendron, L.; Fernandes, A.; Villarrubia, G.; Leithardt, V.; Parreira, W. Intelligent sensors in assistive systems for deaf people: A comprehensive review. *PeerJ Comput. Sci.* **2024**, *10*, e2411. [CrossRef]
118. Zaineldin, H.; Adel Gamel, S.; M. Talaat, F.; Aljohani, M.; Baghdadi, N.A.; Malki, A.; Badawy, M.; Elhosseini, M. Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication. *Artif. Intell. Rev.* **2024**, *57*, 188. [CrossRef]
119. Ramirez, A.E.; Donati, E.; Chousidis, C. A siren identification system using deep learning to aid hearing-impaired people. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105000. [CrossRef]
120. Andersen, A.; Santurette, S.; Pedersen, M.; Alickovic, E.; Fiedler, L.; Jensen, J.; Behrens, T. Creating Clarity in Noisy Environments by Using Deep Learning in Hearing Aids. *Semin. Hear.* **2021**, *42*, 260–281. [CrossRef] [PubMed]
121. Ashkanichenarlogh, V.; Folkeard, P.; Scollie, S.; Kuehnel, V.; Parsa, V. Objective Evaluation of a Deep Learning-Based Noise Reduction Algorithm for Hearing Aids Under Diverse Fitting and Listening Conditions. *Trends Hear.* **2025**, *29*, 23312165251396644. [CrossRef] [PubMed]
122. Park, G.; Cho, W.; Kim, K.S.; Lee, S. Speech Enhancement for Hearing Aids with Deep Learning on Environmental Noises. *Appl. Sci.* **2020**, *10*, 6077. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.