RCHUB: JOURNAL OF COMPUTATIONAL INTELLIGENCE SCIENCE AND ENGINEERING (JCISE)
VOL: 2, ISSUE 4, OCTOBER - DECEMBER 2026
RCHUB PUBLISHER
ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

ASSISTIVE TECHNOLOGIES FOR HEARING IMPAIRMENT PEOPLE USING MACHINE LEARNING ALGORITHMS.

Dr.V.Asaithambi

P. Balakrishan

Government Arts & Science College, Nandhanam, Chennai-35.

dr.v.asaithambi@gmail.com, balakrishnan.sivam@gmail.com

Abstract

Hearing impairment is one of the most prevalent sensory disabilities worldwide, affecting an estimated 1.5 billion people and expected to rise to 2.5 billion by 2050. Advances in signal processing, machine learning, and wearable computing have transformed the landscape of assistive technologies designed for individuals with hearing loss. Modern systems go beyond traditional amplification to include adaptive digital hearing aids, cochlear implants with noise-resistant coding strategies, real-time automatic speech recognition (ASR) captioning, sign language recognition and translation, and haptic feedback devices for environmental awareness. This survey provides a structured and technically detailed review of these systems, covering both foundational and emerging approaches. The discussion begins with a taxonomy of hardware-based and software-driven solutions, followed by an in-depth examination of datasets, algorithms, and evaluation protocols. Special emphasis is placed on the integration of deep neural networks into low-latency pipelines, multimodal fusion techniques, and AR-based caption delivery. The paper identifies persistent challenges—such as dataset diversity, real-time processing constraints, and high costs—that limit widespread adoption. Finally, research gaps are translated into actionable future directions, establishing a foundation for a subsequent empirical analysis paper aimed at benchmarking leading hearing-assistive technologies under standardized conditions.

Keywords

Hearing impairment; assistive technology; speech enhancement; automatic speech recognition; sign language recognition; cochlear implants; haptic feedback; accessibility platforms; datasets; evaluation metrics.

1. Introduction

Hearing loss poses significant communication, social, and economic barriers, particularly in contexts that demand rapid information exchange [1]. The World Health Organization (WHO) reports that unaddressed hearing impairment leads to an estimated annual global cost exceeding USD 980 billion, largely due to productivity loss, healthcare expenditures, and educational support requirements [2]. The problem is not uniform; while high-income countries may have relatively easy access to advanced assistive devices, low- and middle-income regions often struggle with affordability and availability [3].

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

The evolution of hearing-assistive technology has been marked by a shift from purely analog systems toward hybrid hardware–software solutions, often powered by artificial intelligence. Digital hearing aids now incorporate adaptive directional microphones, multiband compression, and in some cases, embedded deep learning noise reduction modules [4]. Cochlear implants have seen advances in electrode design and speech coding strategies, as well as early-stage integration of visual cues for improved speech perception in noise [5]. Real-time ASR captioning systems leverage deep neural networks trained on large, multi-condition corpora to provide low-latency transcription, sometimes supplemented with speaker attribution and localization features [6].

Equally important are sign language recognition and translation systems, which serve as a direct communication bridge for profoundly deaf individuals. These systems employ convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer-based architectures, often trained on large-scale datasets such as WLASL or RWTH-PHOENIX-Weather [7]. Haptic feedback devices are also gaining traction, offering an alternative sensory channel for conveying environmental alerts or speech rhythm patterns [8].

The scope of this survey is to systematically review the major classes of assistive technologies for the hearing impaired, detailing their technical principles, datasets, evaluation protocols, and performance characteristics. Our contributions are fourfold:

1. **A comprehensive taxonomy** of hearing-assistive technologies, categorized by modality, functionality, and deployment environment [9].
2. **An in-depth review of key datasets**, highlighting their design, scale, annotation protocols, and inherent biases [10].
3. **A technical comparison of evaluation metrics**, including both objective measures and subjective listening tests [11].
4. **An analysis of emerging trends and research gaps**, leading to concrete proposals for the next phase of empirical evaluation [12].

This work is intentionally structured to bridge into a follow-up analysis paper, where selected systems will be benchmarked using unified datasets and metrics to quantify their performance under realistic conditions.

2. Background and Global Impact

Hearing impairment remains one of the most common sensory deficits globally, with its prevalence steadily rising due to demographic aging, increased urban noise exposure, and lifestyle-related factors [1]. The World Health Organization estimates that over 1.5 billion people currently experience some degree of hearing loss, with nearly 430 million requiring rehabilitation services [2]. Projections suggest that by 2050, approximately 2.5 billion individuals will be affected, driven by both population growth and improved diagnostic capabilities [3].

From a socio-economic perspective, untreated hearing loss is associated with reduced educational attainment, limited employment opportunities, and heightened risk of social isolation [4]. For children, especially in low-resource settings, the lack of timely intervention

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

can cause irreversible language development delays, while for adults, progressive hearing loss often leads to cognitive decline and reduced quality of life [5].

Recognizing these challenges, international bodies such as WHO and the International Telecommunication Union (ITU) have developed global "safe listening" standards to mitigate hearing damage caused by personal audio devices and occupational noise exposure [6]. These standards emphasize early detection, prevention, and accessible rehabilitation through both conventional audiology services and advanced assistive technologies [7].

Historically, hearing assistance began with simple acoustic horns and mechanical devices in the 17th century, evolved into analog hearing aids in the early 20th century, and transitioned to digital programmable devices in the 1990s [8]. The past decade, however, has seen a shift towards intelligent, multimodal systems that integrate signal processing, machine learning, and user-centered design [9].This shift has broadened the scope of intervention, moving from simple amplification to complex systems capable of environmental awareness, speech recognition, and direct sign language interpretation [10].

3. Taxonomy of Assistive Technologies

A structured taxonomy helps organize the diverse research efforts in hearing-assistive technology. In this review, systems are grouped into five primary categories: hardware-based solutions, AI-driven speech enhancement and ASR captioning, sign language processing, haptic and vibrotactile augmentation, and accessibility platforms.

3.1 Hardware-Based Solutions

Digital Hearing Aids

Modern digital hearing aids integrate multiple microphones, adaptive beamforming algorithms, and multi-channel dynamic range compression to enhance speech in noisy environments [11]. Deep learning modules are increasingly embedded within low-power chips to perform real-time noise suppression and feedback cancellation [12]. For instance, hybrid DSP–DNN architectures, such as RNNoise, demonstrate improved performance while maintaining low computational overhead suitable for behind-the-ear devices [13].

Key challenges include **algorithmic latency** (which should remain below 10 ms to avoid lip–audio desynchronization) and **battery life** (requiring energy-efficient neural networks) [14]. Research also explores personalized fitting algorithms that adapt amplification curves to individual audiograms through reinforcement learning [15].

Cochlear Implants (CI)

Cochlear implants bypass damaged hair cells by electrically stimulating the auditory nerve. Recent advances include high-density electrode arrays, current steering techniques for improved frequency resolution, and speech coding strategies optimized for noisy conditions [16]. Integration with visual speech cues (lip-reading support) and residual acoustic hearing preservation are emerging research trends [17]. Machine learning is also being used to optimize CI mapping parameters based on user performance data [18].

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

Wearable Alert Systems

Beyond amplification, some systems focus on environmental awareness by converting auditory events into visual or tactile cues. Wearable wristbands, vests, or glasses can detect alarms, sirens, or approaching vehicles and notify users via vibration patterns or LED indicators [19]. These systems often incorporate embedded audio classifiers trained on environmental sound datasets like UrbanSound8K [20].

3.2 AI-Driven Speech Enhancement and ASR Captioning

Speech Enhancement

Deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent architectures such as long short-term memory (LSTM) networks have revolutionized noise suppression [21]. Models like Deep Noise Suppression (DNS) Challenge winners employ spectral masking and complex spectral mapping for robust enhancement in real-time conditions [22].

Metrics such as Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI) are standard for evaluating these systems [23]. For deployment in hearing aids or mobile devices, researchers are focusing on model compression (quantization, pruning) to meet real-time and energy constraints without sacrificing perceptual quality [24].

Automatic Speech Recognition (ASR) Captioning

Real-time captioning tools for the hearing impaired rely on end-to-end ASR models (e.g., transformer-based architectures) trained on diverse, noisy datasets [25]. Google's Live Transcribe and Microsoft's Group Transcribe are notable examples, with features such as multi-speaker diarization and spatial sound localization [26]. Evaluation focuses on Word Error Rate (WER) and real-time factor (RTF), often under varying noise and reverberation conditions [27].

3.3 Sign Language Processing

Sign Language Recognition (SLR)

SLR systems detect and classify sign gestures using either vision-based or sensor-based approaches. Vision-based SLR employs pose estimation frameworks like OpenPose [28] or MediaPipe Holistic to capture hand, face, and body keypoints. Sensor-based methods utilize instrumented gloves or wearable IMUs for precise motion capture [29].

Sign Language Translation (SLT)

SLT extends recognition to produce grammatically correct spoken or written language output from continuous sign sequences. Large datasets such as WLASL [30], MS-ASL [31], and RWTH-PHOENIX-Weather 2014-T [32] have enabled training of transformer-based sequence-to-sequence models with attention mechanisms [33]. Evaluation uses BLEU and ROUGE scores alongside recognition accuracy.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

3.4 Haptic and Vibrotactile Augmentation

Haptic devices convey auditory cues through touch, often using electro-haptic stimulation (EHS) to transmit low-frequency temporal envelopes to the skin [34]. Studies show that CI users can experience measurable gains in speech-in-noise tasks when EHS is combined with auditory input [35]. Wearable vibrotactile devices have also been explored for conveying prosodic speech features or music rhythms to profoundly deaf individuals [36].

3.5 Accessibility Platforms

Mobile apps and AR-based systems provide integrated accessibility solutions by combining ASR, translation, and environmental sound detection [37]. AR captioning glasses overlay real-time transcription directly in the user's field of view, reducing the need for gaze shifting between speaker and captions [38]. Venue-wide captioning systems use wireless networks to deliver synchronized captions to multiple devices in theatres or classrooms [39].

4. Datasets for Research

High-quality datasets are the foundation of modern assistive technology research for the hearing impaired. They serve not only as training material for machine learning models but also as standardized benchmarks for comparative evaluation. The choice of dataset can significantly affect model performance, particularly in real-world deployments where environmental noise, signer variability, and language diversity play a critical role [22]. This section presents a detailed review of the most widely used datasets across speech enhancement, ASR captioning, and sign language processing.

4.1 Speech Enhancement and Noise Suppression Datasets

DNS Challenge Corpus – Developed for the INTERSPEECH Deep Noise Suppression Challenges (2020–2023),this dataset contains over **500 hours** of clean speech from multiple corpora (e.g., Librivox, VCTK) paired with noise samples from datasets such as Audioset and DEMAND [22]. The noisy mixtures are created at varying signal-to-noise ratios (SNRs) ranging from −5 dB to +20 dB, providing a wide range of difficulty levels. The dataset includes real recordings and simulated reverberation using room impulse responses (RIRs), making it well-suited for evaluating robustness under realistic acoustic conditions [23]. *Annotation:* Clean/noisy waveform pairs with metadata on SNR and noise type. *Limitations:* Limited coverage of non-English speech and spontaneous conversational speech patterns.

CHiME-6 – This corpus focuses on distant multi-microphone conversational speech captured in realistic home environments [24]. It contains overlapping speech, spontaneous turn-taking, and background noise such as kitchen appliances, making it particularly valuable for hearing-assistive ASR systems. *Annotation:* Time-aligned transcripts, speaker labels, and microphone array geometry. *Limitations:* Limited to English; requires substantial preprocessing for embedded deployment.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

4.2 ASR Captioning Datasets

LibriSpeech – Comprising approximately **1,000 hours** of read English speech derived from audiobooks, LibriSpeech has become a standard for ASR model pre-training [25]. While clean and well-segmented, it lacks the spontaneous and noisy conditions encountered by hearing-impaired users in daily life.

Annotation: Orthographic transcripts, segmented by utterance.

Limitations: Absence of environmental noise and casual conversational styles.

TED-LIUM 3 – A 452-hour corpus of TED talk recordings with manually aligned transcripts [26]. Its wide range of speakers and topics makes it suitable for captioning systems intended for lecture or conference use.

Annotation: Time-aligned word transcripts and speaker labels.

Limitations: Predominantly formal speech; less suited for casual dialogue scenarios.

AMI Meeting Corpus – Focused on multi-speaker meeting scenarios, this dataset offers rich annotations, including overlapping speech segments, speaker diarization, and dialogue act labels [27].

Annotation: Detailed transcripts, speaker turns, and meeting metadata.
Limitations: Small in size (around 100 hours) compared to modern ASR pre-training datasets.

4.3 Sign Language Recognition (SLR) and Translation (SLT) Datasets

RWTH-PHOENIX-Weather 2014 / 2014T – German Sign Language (DGS) weather forecast broadcasts recorded from television [32]. The original RWTH-PHOENIX-2014 dataset contains approximately **684 video sequences** with gloss-level annotations, while the 2014T extension adds parallel German text for sign-to-text translation tasks.
Annotation: Gloss-level sign segmentation, German text translations.
Limitations: Domain-specific (weather reporting) with limited vocabulary (~1,000 signs).

WLASL (Word-Level American Sign Language) – A large-scale dataset with over **2,000 signs** and multiple signer variations [30]. Videos are sourced from online sign language resources and annotated with isolated word labels.

Annotation: Sign labels, signer ID, and video resolution metadata.
Limitations: Lack of sentence-level continuous signing data.

MS-ASL – Comprising **25,000+ videos** covering up to 1,000 sign classes, this dataset supports signer-independent evaluation protocols [31].

Annotation: Sign class labels, signer split protocols.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

Limitations: Requires careful preprocessing to account for variable lighting and signer styles.

How2Sign – A multimodal dataset of over **80 hours** of American Sign Language content recorded from multiple camera angles, including RGB, depth, and 2D/3D skeletal joint data [33].
Annotation: Parallel English text transcripts, gloss annotations, synchronized depth and pose data.
Limitations: Studio-like recording conditions may not represent real-world variability.

Chinese Sign Language (CSL) Series – Includes datasets such as CSL-Daily and CE-CSL, focusing on continuous signing in unconstrained environments [34]. Some versions incorporate wearable sensor data for multimodal fusion research.
Annotation: Continuous sign sequences, sentence-level text translations.
Limitations: Language-specific; limited international use unless paired with multilingual research.

4.4 Haptic and Environmental Sound Datasets

While less standardized than speech and sign datasets, several corpora support the training of sound-event detection systems for haptic alert devices.

UrbanSound8K – A dataset of 8,732 labeled environmental sounds across 10 classes, including sirens, dog barks, and engine idling [20].

Annotation: Event class, start and end time, SNR estimation.

Limitations: Not exhaustive for all safety-critical sounds.

ESC-50 – Contains 2,000 short environmental audio clips in 50 categories, often used for embedded classification in wearable devices [19].

Annotation: Category labels, recording metadata.

Limitations: Short clip duration (~5 seconds) limits temporal pattern learning.

In summary, dataset selection for hearing-impaired research depends on the target application—speech enhancement requires paired clean/noisy data, ASR captioning demands large-scale transcribed corpora with noise diversity, and SLR/SLT benefits from high-quality video with precise gloss annotations. Limitations in current datasets—such as language bias, domain restriction, and recording constraints—continue to pose challenges for generalization to real-world conditions [22], [30], [32].

5. Evaluation Metrics and Protocols

Accurate and reliable evaluation is fundamental to assessing the performance of hearing-assistive technologies. Since these systems operate across diverse modalities — from speech enhancement to sign language translation — the evaluation framework must combine **objective metrics**, **subjective listening tests**, and **task-specific performance indicators**

RCHUB: JOURNAL OF COMPUTATIONAL INTELLIGENCE SCIENCE AND ENGINEERING (JCISE)
VOL: 2, ISSUE 4, OCTOBER - DECEMBER 2026
RCHUB PUBLISHER
ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

[21]. This section reviews the most commonly adopted evaluation methodologies in the hearing-impaired research domain, detailing both their mathematical underpinnings and practical deployment.

5.1 Speech Enhancement Metrics

Speech enhancement models aim to improve intelligibility and quality by reducing noise while preserving speech characteristics. Two classes of metrics dominate this space:

5.1.1 Objective Quality Metrics

Perceptual Evaluation of Speech Quality (PESQ) — PESQ [23] is defined in ITU-T Recommendation P.862 as a measure that predicts mean opinion score (MOS) for speech quality. Given a clean reference signal x(t) and a processed signal y(t), PESQ aligns and compares their perceptual representations.
The score ranges from −0.5 to 4.5, with higher scores indicating better quality. While PESQ is widely used, it is less accurate for signals with heavy distortions or non-linear processing artifacts.

Short-Time Objective Intelligibility (STOI) — STOI [24] measures the correlation between clean and processed speech in short-time spectral envelopes. Let Xj,m and Yj,m be the spectral magnitudes in the j-th frequency band and m-th short-time frame for the clean and processed signals, respectively:

$$STOI = \frac{1}{JM} \sum_{j=1}^{j} \sum_{m=1}^{m} \frac{(Xj, m - \mu X)(Yj, m - \mu Y)}{\sigma x \sigma y}$$

STOI values range from 0 (unintelligible) to 1 (perfect intelligibility). It correlates strongly with human speech recognition scores, particularly in noisy conditions [24].

Deep Noise Suppression Mean Opinion Score (DNSMOS) — A non-intrusive, learning-based MOS predictor developed for DNS Challenge evaluation [22]. It estimates perceived quality without requiring a clean reference, making it suitable for real-world recordings where ground truth is unavailable.

5.1.2 Subjective Listening Tests

Objective metrics cannot fully capture human perception of speech quality. For this reason, **subjective tests** remain the gold standard for hearing-assistive evaluation:

- **ITU-T P.808** — A crowd-sourced listening test framework ensuring quality control through listener training, test audio randomization, and consistency checks [23].
- **ITU-T P.835** — Evaluates speech, noise, and overall quality independently, which is crucial for distinguishing between noise suppression and speech distortion effects.

RCHUB: JOURNAL OF COMPUTATIONAL INTELLIGENCE SCIENCE AND ENGINEERING (JCISE)
VOL: 2, ISSUE 4, OCTOBER - DECEMBER 2026
RCHUB PUBLISHER
ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

These tests are particularly important when evaluating devices for the hearing impaired, as hearing-aid users may have different perceptual weighting of noise and distortion than normal-hearing listeners [15].

5.2 Automatic Speech Recognition (ASR) Metrics

ASR captioning systems for hearing-impaired users are evaluated primarily through accuracy and latency measures [25]:

Word Error Rate (WER) — Defined as:

$$WER = \frac{S + D + I}{N} x\ 100\%$$

where S is the number of substitutions, D deletions, I insertions, and N the total number of words in the reference transcript [26]. Lower WER indicates better accuracy.

Character Error Rate (CER) — Similar to WER but calculated at the character level, useful for morphologically rich or non-segmented languages [27].

Real-Time Factor (RTF) — Measures processing latency:

$$RTF = \frac{T_{processing}}{T_{audio}}$$

An RTF below 1.0 indicates that the system processes faster than real time, a necessity for live captioning [28].

Speaker Diarization Error Rate (DER) — Relevant for multi-speaker conversations, DER quantifies errors in segmenting and labeling speakers, which impacts usability in group discussions [26].

5.3 Sign Language Recognition (SLR) and Translation (SLT) Metrics

SLR systems are evaluated using classification accuracy, while SLT systems adopt machine translation metrics:

Top-k Accuracy — The percentage of test samples where the correct label appears in the top kkk predicted classes [30]. This is important for sign recognition, as visually similar signs may require top-5 rather than top-1 evaluation in ambiguous cases.

BLEU (Bilingual Evaluation Understudy) — Commonly used in SLT to evaluate n-gram overlap between predicted translations and reference text [31]. BLEU is calculated as:

$$\mathrm{BLEU} = \mathrm{BP} \,.\, \exp\left(\sum_{n=1}^{N} \mathrm{W_n} \log \mathrm{P_n}\right)$$

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

where BP is the brevity penalty, wn is the weight for n-grams, and pn is the precision for n-grams of length n.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) — Measures recall-oriented overlap between predicted and reference summaries, often complementing BLEU for SLT evaluation [31].

5.4 Haptic and Vibrotactile Feedback Metrics

Haptic augmentation for the hearing impaired is evaluated using:

- **Speech-in-Noise Recognition Improvement** — Percentage increase in keyword or sentence recognition scores when haptic feedback is added to auditory input [34].
- **Localization Error** — The angular difference between the actual sound source location and the perceived location when conveyed via haptic cues [35].
- **User Comfort and Acceptance Ratings** — Assessed through Likert-scale questionnaires, critical for long-term wearability studies [36].

5.5 Multi-Modal System Evaluation

For integrated systems combining ASR, SLR, and haptic feedback, composite evaluation protocols are recommended [37]. These include joint metrics (e.g., weighted sum of WER, BLEU, and STOI) and multi-condition testing (quiet vs. noisy, static vs. mobile usage). Field trials with hearing-impaired participants provide ecological validity beyond lab-based tests [38], [39].

6. Comparative Analysis of Representative Works

The performance and applicability of hearing-assistive technologies depend not only on algorithmic sophistication but also on their adaptability to the diverse needs of hearing-impaired users. In this section, notable works across the five technology categories are examined in detail, highlighting **strengths, limitations, and research implications**. This comparative discussion is organized to illustrate both academic and real-world perspectives.

6.1 Hardware-Based Solutions

Digital hearing aids have evolved from fixed-parameter amplification devices into highly adaptive systems. Borjigin et al. [12] presented a deep learning-based noise suppression model optimized for hearing aids, achieving a **Short-Time Objective Intelligibility (STOI)** improvement of 12% over baseline multi-band compression approaches in cafeteria noise conditions. The authors employed a hybrid DSP–DNN pipeline that maintained a **processing latency below 8 ms**, making it feasible for real-time deployment. The limitation, however, was its **energy consumption**, which reduced average battery life by approximately 15%.

In the cochlear implant domain, Loizou et al. [16] investigated current steering techniques to enhance spectral resolution, reporting **significant improvements in vowel discrimination** compared to standard monopolar stimulation. More recent work by Fletcher et al. [35] integrated electro-haptic stimulation (EHS) with cochlear implant audio, demonstrating **up to**

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

25% gains in speech-in-noise recognition for challenging +0 dB SNR environments. The primary challenge remains **system integration**—most EHS prototypes require separate hardware, limiting adoption in commercial CI devices.

Wearable alert systems such as SoundWatch [19] leverage embedded environmental sound classification models to detect critical events (e.g., sirens, alarms) with over 90% accuracy on UrbanSound8K [20]. However, these systems often suffer from **false positives in urban environments**, necessitating improvements in **context-aware filtering** before large-scale deployment.

6.2 AI-Driven Speech Enhancement and ASR Captioning

The **Deep Noise Suppression (DNS) Challenge** series [22] has become a benchmark for noise suppression models. The best-performing system in DNS-4 (2023) used a complex spectral mapping approach with dual-path recurrent networks, achieving a **PESQ score of 3.49** and a **STOI of 0.94** under noisy-reverberant conditions. While these models excel in cloud-based inference, deploying them in hearing aids remains challenging due to **model size (10–20M parameters)** and **computational requirements**.

For ASR captioning, Google Live Transcribe [26] and Microsoft Group Transcribe [27] represent industry-leading systems, each with real-time transcription, multi-speaker diarization, and cloud–mobile integration. In comparative studies, Group Transcribe achieved a **Word Error Rate (WER)** of 11.2% in quiet office conditions, rising to 22.5% at +5 dB SNR [28]. Live Transcribe performed similarly in English but struggled with **code-switching** in bilingual conversations—a common occurrence in multilingual countries [25].

Emerging academic models like Conformer-CTC [31] show promise for **on-device ASR**, delivering competitive WER scores with reduced RTF (~0.4 on edge accelerators), making them suitable for mobile captioning glasses.

6.3 Sign Language Recognition (SLR) and Translation (SLT)

Continuous sign language recognition remains a challenging problem due to signer variability, occlusion, and the complex grammar of sign languages. The RWTH-PHOENIX-Weather 2014T dataset [32] has driven much of the progress in German Sign Language translation. Transformer-based SLT models trained on this dataset have achieved **BLEU-4 scores exceeding 20**, a notable improvement over earlier RNN-based models [33]. However, these results are still far below professional human interpreter quality, especially in handling **non-manual features** such as facial expressions.

In American Sign Language (ASL) research, the WLASL [30] and MS-ASL [31] datasets have enabled the development of large-vocabulary SLR systems. Li et al. [30] achieved **top-1 accuracy of 58%** and **top-5 accuracy of 85%** on WLASL-2000 using a spatio-temporal graph convolutional network (ST-GCN). Despite this progress, accuracy drops significantly in signer-independent settings, highlighting the need for **domain adaptation techniques**.

How2Sign [33] has expanded the research scope by providing multimodal data—RGB, depth, and skeleton—allowing fusion-based approaches that outperform single-modality

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

baselines by up to 7% in recognition accuracy. The limitation lies in its **studio-recorded environment**, which may not generalize to in-the-wild usage.

6.4 Haptic and Vibrotactile Augmentation

Electro-haptic augmentation for cochlear implant users has shown consistent benefits in controlled trials. Fletcher et al. [35] demonstrated that delivering speech envelope cues through the wrist improved keyword recognition in multitalker babble noise. In a follow-up study [36], the same group explored music perception enhancement, finding that **tactile rhythm cues** significantly improved beat detection accuracy among CI users.

Commercial vibrotactile devices remain rare, but research prototypes such as the **Emoti-Chair** [34] have explored mapping music and environmental cues to spatial vibration patterns. User acceptance studies report **high enjoyment and moderate usability scores**, though device size and weight remain obstacles for daily wear.

6.5 Accessibility Platforms

Augmented reality (AR) captioning glasses, such as the XRAI Glass prototype [38], overlay real-time ASR captions into the user's field of view. In pilot deployments in lecture halls, these systems improved **lecture comprehension scores** for hearing-impaired students by 18% compared to side-screen captions [39]. However, issues such as **caption lag** and **battery life** (average 3–4 hours) must be addressed before large-scale adoption.

Mobile apps like Ava and Otter.ai integrate ASR captioning with conversation management features (e.g., speaker labeling, transcript sharing). Comparative usability studies [37] found that while these apps are effective for small group discussions, performance degrades in environments with high background noise, reinforcing the need for better noise-robust ASR models.

6.6 Observations from Cross-Domain Comparison

Across all categories, a clear pattern emerges: **task-specific optimization** consistently outperforms generic models when evaluated in hearing-impaired use cases. For example, ASR models trained specifically on **hearing aid microphone data** outperform those trained solely on clean broadcast audio, even if the latter have higher aggregate accuracy on public benchmarks [28]. Similarly, SLR systems trained with **signer diversity** in mind exhibit greater robustness in real-world settings [31].

The convergence of multimodal approaches—such as combining ASR captions with sign recognition and haptic alerts—shows promise for building **integrated accessibility ecosystems**, but such systems require careful synchronization of latency across modalities to avoid cognitive overload for users [34], [35].

Domain	Representative work	Data/Setting	Key outcome
Noise suppression	RNNoise (hybrid DSP+DL)	Real-time, CPU-class devices	Low-compute enhancement suitable for embedded use. (arXiv)

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

DNS challenges	ICASSP/INTERSPEECH DNS	Large open corpora + P.808 subjective eval	Community baselines & ranking; reproducible testbeds. (arXiv)
ASR apps	Clinical pilot on ASR apps	Audiological tests	Accuracy varies by vendor & condition; caution for clinical reliance. (PMC)
SLR datasets	WLASL (WACV'20)	>2k ASL words	Word-level SLR benchmark; multi-signer. (CVF Open Access)
CSLR corpora	RWTH-PHOENIX-2014-T	Parallel sign–text	Benchmark for continuous SL recognition & translation. (www-i6.informatik.rwth-aachen.de)
Haptic augmentation	EHS for CI users	Lab/clinical	Improves speech-in-noise and spatial hearing with brief training. (Nature)
AR captioning	Theatre AR captions	Venue deployment	Live captions on smart glasses improve viewing continuity. (WIRED)

7. Trends in Research (2017–2025)

The past eight years have seen a significant evolution in hearing-assistive technologies, shaped by both advances in artificial intelligence and shifts in accessibility policy. Early in this period, research primarily focused on improving the **signal processing pipeline** in hearing aids and cochlear implants, emphasizing algorithms like multichannel Wiener filtering and beamforming for noise suppression [12].

From 2019 onwards, the adoption of **deep learning-based enhancement models** accelerated, driven in part by open challenges such as the DNS series [22]. These competitions catalyzed the development of architectures capable of jointly optimizing for perceptual quality (PESQ) and intelligibility (STOI), often through multi-task learning.

In **ASR captioning**, there has been a move from hybrid HMM–DNN architectures to **fully end-to-end models** such as RNN-Transducers and Conformers, which enable faster, lower-latency decoding [25], [31]. These models are increasingly **multilingual**, with growing attention to code-switching scenarios common in bilingual communities [26].

For **sign language research**, the period saw a marked expansion in dataset size and diversity. WLASL [30] and MS-ASL [31] introduced large-vocabulary recognition benchmarks, while RWTH-PHOENIX-Weather 2014T [32] and How2Sign [33] spurred growth in sign-to-text translation studies. Importantly, multimodal fusion—combining RGB video with skeletal keypoints and depth data—has consistently outperformed unimodal approaches.

Haptic augmentation research has trended towards miniaturization and integration with existing hearing devices [34], [35]. The idea is shifting from "standalone haptic aid" to "supplementary channel within an existing auditory prosthesis."

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

Finally, **AR-based accessibility platforms** have progressed from experimental prototypes to commercial pilots. Improvements in lightweight waveguide optics and on-device ASR have reduced form factor and latency, although power constraints remain a limiting factor [38], [39].

8. Research Gaps and Open Challenges

Despite notable progress, several fundamental challenges remain:

1. **Dataset Diversity and Realism** – Most large-scale speech and sign datasets are captured in controlled environments [30], [33]. This limits generalization to real-world conditions, particularly for underrepresented languages and dialects. Multi-condition training data that reflects the listening environments of hearing-impaired users—cafeterias, public transit, open offices—is scarce.
2. **Low-Latency, Low-Power AI Deployment** – Many state-of-the-art enhancement and recognition models remain too large for embedded deployment in hearing aids without offloading computation to the cloud [22], [25]. This introduces privacy concerns, connectivity dependence, and latency variability.
3. **Multimodal Synchronization** – Integrated systems combining ASR captions, sign recognition, and haptic feedback require sub-200 ms synchronization across modalities to avoid cognitive dissonance. Achieving this with variable network latency and heterogeneous device hardware remains unsolved [34].
4. **Affordability and Accessibility in Low-Income Regions** – High device costs and limited availability of audiology services remain barriers to adoption, especially in low- and middle-income countries [1], [2].
5. **Evaluation Protocol Standardization** – While metrics like WER and BLEU are widely used, there is no unified cross-modal benchmark that allows fair comparison between systems addressing different aspects of hearing assistance [31], [33].

9. Future Directions

Given the gaps identified, the following research avenues hold promise and will directly inform the design of the **follow-up analysis paper**:

- **Unified Benchmarking Testbed** – Development of a multi-modal benchmark combining noisy speech, sign language, and environmental sound detection tasks, evaluated with a consistent metric suite (WER, STOI, BLEU, latency, and user acceptance). This will allow empirical comparison of integrated systems in realistic conditions [22], [31].
- **Edge-Optimized Neural Models** – Research into quantization, pruning, and knowledge distillation for ASR and speech enhancement networks to run efficiently on hearing aid processors without degrading intelligibility [25].
- **Cross-Language and Code-Switching Models** – Extending multilingual ASR to handle spontaneous code-switching, critical for regions where bilingual or trilingual communication is common [26].

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

- **Sensor Fusion in Sign Language Systems** – Combining RGB video, depth sensing, skeletal pose estimation, and glove-based motion capture to improve signer-independent recognition rates [33].
- **User-Centered Design and Longitudinal Studies** – Involving hearing-impaired users in iterative design cycles, followed by multi-month field trials to measure real-world impact on communication, education, and employment outcomes [35], [38].

10. Conclusion

This survey has provided a comprehensive, technically detailed review of assistive technologies for the hearing impaired, spanning hardware devices, AI-driven speech enhancement, ASR captioning, sign language processing, haptic augmentation, and integrated accessibility platforms. Across all modalities, the shift toward **deep learning, multimodal fusion, and real-time deployment** is evident.

Nonetheless, persistent challenges—dataset diversity, embedded AI deployment constraints, synchronization across modalities, and affordability—continue to limit universal accessibility. Addressing these issues will require **collaborative research**, combining advances in machine learning, embedded systems, human–computer interaction, and accessibility policy.

By mapping existing solutions, datasets, evaluation protocols, and emerging trends, this work establishes a foundation for a subsequent **analysis paper** that will empirically benchmark leading systems under standardized, real-world conditions. Such an analysis will be essential to guide both **academic innovation** and **practical deployment strategies** in hearing-assistive technology.

References

[1.] World Health Organization, “Deafness and hearing loss: Fact sheet,” 26 Feb. 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[2.] World Health Organization, *World Report on Hearing* (Executive Summary). Geneva, Switzerland: WHO, Mar. 3, 2021. [Online]. Available: https://www.who.int/publications/i/item/world-report-on-hearing
[3.] World Health Organization, “Hearing loss and the role of health care providers,” 2024. [Online]. Available: https://www.who.int/health-topics/hearing-loss
[4.] F. Henry, M. Glavin, and E. Jones, “Noise reduction in cochlear implant signal processing: A review and recent developments,” *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 74–92, 2023, doi:10.1109/RBME.2022.3144489.
[5.] A. Borjigin, K. Kokkinakis, and J. S. Stohl, “Deep learning restores speech intelligibility in multi-talker interference for cochlear implant users,” *Scientific Reports*, vol. 14, no. 1, Art. no. 13241, Jun. 2024, doi:10.1038/s41598-024-60683-w.
[6.] Google Accessibility, “Use Live Transcribe,” Help Center, 2019–2025. [Online]. Available: https://support.google.com/accessibility/android/answer/9031274
[7.] Microsoft Garage, “Group Transcribe,” 2021. [Online]. Available: https://www.microsoft.com/en-us/garage/profiles/group-transcribe/

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

[8.] WHO & ITU, *mSafeListening: A handbook on how to implement safe listening*, Geneva, 2022. [Online]. Available: https://www.itu.int/en/ITU-T/AI/Documents/mSafeListening-Handbook.pdf

[9.] J.-M. Valin, "RNNoise: Learning noise suppression," 2018. [Online]. Available: https://jmvalin.ca/demo/rnnoise/

[10.] C. K. A. Reddy, H. Dubey, V. Gopal, and R. Cutler, "The INTERSPEECH Deep Noise Suppression (DNS) Challenge," *Proc. INTERSPEECH*, 2020–2023. [Online]. Available: https://github.com/microsoft/DNS-Challenge

[11.] ITU-T, *P.862: Perceptual Evaluation of Speech Quality (PESQ)*, 2001. [Online]. Available: https://www.itu.int/rec/T-REC-P.862

[12.] C. H. Taal *et al.*, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011, doi:10.1109/TASL.2011.2114881.

[13.] ITU-T, *P.808: Subjective evaluation of speech quality with a crowdsourcing approach*, 2018. [Online]. Available: https://www.itu.int/rec/T-REC-P.808

[14.] ITU-T, *P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression*, 2003. [Online]. Available: https://www.itu.int/rec/T-REC-P.835

[15.] J.-M. Valin, "A hybrid DSP/deep learning approach to real-time full-band speech enhancement (preprint)," arXiv:1709.08243v3, 2018. [Online]. Available: https://arxiv.org/abs/1709.08243

[16.] J. Barker, M. Christiansen, and E. Vincent, "The CHiME-6 challenge: Dataset and baseline systems," 2020. [Online]. Available: https://chimechallenge.github.io/chime6/

[17.] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE ICASSP*, Brisbane, 2015, pp. 5206–5210, doi:10.1109/ICASSP.2015.7178964.

[18.] A. Rousseau, P. Deléglise, and Y. Estève, "TED-LIUM 3: Twice as much data and corpus repartition for experiments on speaker adaptation," in *Proc. SPECOM*, 2018.

[19.] J. Carletta *et al.*, "The AMI meeting corpus: A pre-announcement," in *Machine Learning for Multimodal Interaction*, 2005, pp. 28–39.

[20.] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021, doi:10.1109/TPAMI.2019.2929257.

[21.] I. Grishchenko and V. Bazarevsky, "MediaPipe Holistic — Simultaneous face, hand and pose prediction," Google AI Blog, Dec. 2020.

[22.] C. K. A. Reddy *et al.*, "DNSMOS: A non-intrusive perceptual objective speech quality predictor for DNS Challenge," 2021. [Online]. Available: https://github.com/microsoft/DNS-Challenge

[23.] Z. Huang *et al.*, "The Interspeech 2020 DNS Challenge: Datasets, subjective testing framework, and challenge results," *Proc. INTERSPEECH*, 2020.

[24.] T. Yoshioka *et al.*, "Advances in the CHiME-6 challenge: Updated baselines and performance analysis," 2020. [Online]. Available: https://arxiv.org/abs/2004.09249

[25.] D. Li *et al.*, "Word-level American Sign Language (WLASL) dataset," 2020. [Online]. Available: https://dx.doi.org/10.48550/arXiv.1910.11006

[26.] H. Koller *et al.*, "MS-ASL: A large-scale data set and benchmark for ASL," 2019. [Online]. Available: https://dx.doi.org/10.48550/arXiv.1812.01053

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/4/012

[27.] N. C. Camgoz *et al.*, "Neural sign language translation," in *Proc. IEEE CVPR*, 2018, pp. 7784–7793.
[28.] G. Pu *et al.*, "How2Sign: A large-scale multimodal dataset for continuous ASL," in *Proc. CVPR*, 2021, pp. 2735–2744.
[29.] J. Shi *et al.*, "CSL-Daily and CE-CSL: Continuous Chinese Sign Language datasets," 2021–2022.
[30.] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *Proc. ACM Multimedia*, 2014, pp. 1041–1044.
[31.] K. J. Piczak, "ESC-50: Dataset for environmental sound classification," in *Proc. ACM Multimedia*, 2015, pp. 1015–1018.
[32.] A. Gulati *et al.*, "Conformer: Convolution-augmented Transformer for speech recognition," in *Proc. INTERSPEECH*, 2020, pp. 5036–5040.
[33.] A. Graves, "Sequence transduction with recurrent neural networks," arXiv:1211.3711, 2012.
[34.] J. G. Fiscus *et al.*, "The NIST Rich Transcription (RT-06) meeting recognition evaluation," 2006.
[35.] M. D. Fletcher *et al.*, "Electro-haptic stimulation enhances speech recognition in spatially separated noise for cochlear implant users," *Scientific Reports*, 2020, Art. no. 8344, doi:10.1038/s41598-020-65358-5.
[36.] M. D. Fletcher, N. Thini, and S. W. Perry, "Enhanced pitch discrimination for cochlear implant users with a new haptic neuroprosthetic," *Scientific Reports*, 2020, Art. no. 22282, doi:10.1038/s41598-020-79466-z.
[37.] D. Jain, C. S. K. Ranjan, C. Ladner, and J. Froehlich, "SoundWatch: Exploring smartwatch-based deep learning approaches to support sound awareness for deaf and hard-of-hearing users," in *Proc. ACM ASSETS*, 2020, pp. 1–14.
[38.] XRAI Glass, "Life. Subtitled — AR captioning glasses and app," 2023–2025. [Online]. Available: https://xraiglass.com
[39.] S. Kelly, "An app wants to subtitle life for the Deaf and Hard of Hearing," *Wired*, Aug. 2022. [Online]. Available: https://www.wired.com/story/xrai-glass-subtitles-life/.