ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 10, October 2024

Addressing Hearing Impairments Through Machine Learning: A Review of Sound Detection and Assistive Technologies

[1] Nada Barnawi, [2] Mohammed Alnuem

[1] Researcher, Nada Barnawi, Information Systems Department, King Saud University, Saudi Arabia, Riyadh
[2] Professor, Mohammed Alnuem, Information Systems Department, King Saud University, Saudi Arabia, Riyadh
Corresponding Author Email: [1] Nbarnawi@ksu.edu.sa, [2] Malnuem@ksu.edu.sa

Abstract— ***This paper studies recent assistive technologies and AI sound detection systems that have been developed to support both the safety and communication of individuals who are deaf. With a highlight of the critical role of machine learning in all four stages, which are processing, feature extraction, real-time sound event detection, and deep learning applications, the paper is able to give an overall understanding of the sound wave detection system. Additionally, it addresses challenges such as real-time processing, polyphonic sound detection, and the integration of speech recognition technologies to improve communication and situational awareness. By synthesizing recent research and applications, this paper seeks to demonstrate the transformative impact of these technologies on the quality of life and safety for those with hearing impairments.***

Index Terms— ***Machine learning-based, Deaf and hard of hearing, Sound detection, Assistive Technologies, Speech Recognition, and Realtime.***

I. INTRODUCTION

In daily life, humans face several dangers that often require constant vigilance and caution to deal with surrounding risks in general. This necessitates attention, identification of potential hazards, determination of the type of risk, and subsequent action accordingly. Risks vary in terms of response speed and the level of danger they pose. Some, like responding to fire alarms in a building, demand swift movement and evacuation, potentially even saving lives.

One of the most significant obstacles that may impede a swift response is the loss of hearing. Deaf or hard of hearing individuals face additional risks, such as an inability to respond to potential warnings and alerts. They heavily rely on others in times of danger, leading to a loss in response speed and the ability to promptly handle similar risks. Furthermore, beyond difficulty in hearing and identifying risks, they struggle with oral communication in their daily lives, especially in times of potential danger.

Recent data from the World Health Organization (WHO) highlights a growing global trend in hearing impairment. Currently, over 430 million people, more than 5% of the global population, suffer from hearing loss, whether congenital or acquired. By 2050, this number is expected to exceed 700 million, meaning one in ten people will experience disabling hearing loss [1]. This group is identified as the second-largest among individuals with disabilities in the General Census of Population and Housing [2].

Humanity has leveraged the tremendous and rapid advancements in computer science, especially in the field of artificial intelligence, to provide tools and resources to assist individuals with hearing impairments in their daily lives. This has created a significant leap in changing their lifestyle and habits for the better. Additionally, it has contributed to raising awareness about the dangers surrounding them. For instance, artificial intelligence technologies have provided the ability to identify and recognize surrounding sounds, whether they are hazardous or not, alerting and pinpointing the source of danger and dealing with it.

So, in this paper, it will serve as the literature review, elucidating the key concepts central to understanding how machine learning can contribute and utilization to enhance comfort and safety environment for deaf and hard of hearing people. It commences with a concise overview of various assistive technologies. Also, provides an exploration of sound detection technologies, and comprehensively covers sound detection within machine learning, encompassing preprocessing, feature extraction, model training and classification, real-time aspects, and associated challenges and constraints. In addition, a grasp of speech recognition and automatic speech recognition is established. To conclude this comprehensive review, presents an examination of research papers and applications offering insightful reviews.

II. ASSISTIVE TECHNOLOGIES TO HEARING DISABILITIES

Assistive technologies (AT) have been among the most significant advancements in the past 20 years for helping individuals with disabilities, including those with hearing impairments. According to ISO 9999:2016 and UNE-ISO 9999:2017, AT encompasses any product—whether a device, piece of equipment, instrument, or software—specifically designed to improve the engagement and functionality of individuals with disabilities. These technologies assist, support, train, measure, or substitute for bodily functions, aiming to prevent impairments, activity limitations, or participation restrictions [3].

For those with hearing impairments, AT includes hearing aids, communication systems, low-tech devices, cochlear implants, and specialized software and hardware that enhance hearing and communication abilities. These tools not only promote independence and well-being but also prevent secondary health issues and provide socioeconomic benefits by reducing healthcare costs and stimulating economic growth [3].

In communication for the deaf and hard of hearing, significant technological progress has been made. Text telephones (TTY) and telecommunications devices have been crucial in enabling telephone communication for this community, with communications assistants helping to bridge the gap between hearing-impaired and standard telephone users by relaying messages [2]. Recent innovations include sign recognition systems for interpreting sign language gestures and Personalized Emergency Response Systems that enhance safety through sensor-based alarms [4,5]. Mobile applications have also become a breakthrough, empowering the hearing-impaired to live more autonomously and communicate effectively within their communities [2]. These technologies cater to the diverse needs of individuals with hearing impairments across various areas of life.

III. SOUND DETECTION TECHNOLOGIES

Sound detection technology has seen growing interest due to recent advancements that have made it more reliable and precise. These developments enable machines to mimic human hearing and understand context, allowing for applications such as smartphone alerts [6], diagnosing conditions like coughing [7], improving security by identifying threats, aiding in cataloging audio archives [8], and enhancing safety monitoring on construction sites [9]. These technologies provide critical situational awareness, especially in environments beyond visual or attentive reach.

Sound detection involves identifying sound events in a continuous audio signal and analyzing them using various methodologies to extract relevant information, depending on the intended application. This process typically involves multiple techniques related to audio signal processing or machine learning. For example, traditional computational analysis systems extract specific acoustic features from an input signal, which can then be categorized and detected using supervised classifiers like neural networks. Developing sound detection applications requires defining several factors, such as the nature of the application, technological constraints, desired complexity and precision, and data availability [10].

Sound event detection systems are usually customized for specific tasks and environments, requiring a combination of different techniques and processes. The implementation of sound detection technologies involves integrating various methods tailored to specific purposes. The following section will explore the core processes and key techniques used in sound detection technologies, particularly those involving machine learning, in different environments.

IV. MACHINE LEARNING-BASED SOUND DETECTION

Machine learning has significantly advanced numerous domains, as highlighted by Rebala et al. They have succinctly defined the core processes of sound detection within the broader machine learning context. They characterize machine learning as a computer science discipline dedicated to automating solutions for intricate problems that defy conventional programming methods. Traditional programming necessitates meticulous design and code implementation, posing challenges for tasks such as character recognition or sound event detection. In contrast, machine learning algorithms acquire knowledge from labeled data, bypassing the need for explicit rules. These algorithms excel at solving complex problems with greater accuracy and objectivity compared to rules crafted by humans. Their approach involves creating a model from a dataset and subsequently predicting labels for new data points [11].

Therefore, in order to improve the detection of sound events across various applications, it is contemporary to employ machine learning methods and techniques, whether in real-time or not. It is also essential to grasp the key challenges and constraints associated with machine learning-based sound detection. So, the following sections will discuss more about these topics.

A. Processes and Technics for Sound Detection:

To gain a deeper understanding of the fundamental processes and methodologies utilized in machine learning for sound detection, the two referenced papers [12, 13] provide insights into the core stages of sound detection technologies or sound event detection systems. These stages comprise three key components: (1) Preprocessing, (2) Feature Extraction, and (3) Model Training and Classification, various papers will demonstrate a range of machine-learning techniques for sound detection, with supervised learning emerging as the dominant approach in tackling the sound event detection task.

1- Preprocessing step

The initial stage in sound event detection entails the application of various techniques to enhance the quality of the audio data before feature extraction. This step is necessary because raw audio data cannot be directly employed as input for machine learning-based classification. The rationale behind this necessity lies in the presence of signal redundancy, which must be addressed. Preprocessing activities typically encompass tasks such as noise reduction, equalization, low-pass filtering, and segmenting the original audio signal into audio and silent events to facilitate subsequent feature extraction [12, 13].

Data Collection: Effective sound detection systems hinge on understanding sound data and context, especially in supervised learning, where the training data must closely resemble the intended application scenario. As sound event detection can cover a broad range of sound classes and environments, no single dataset or acoustic model fits all scenarios. Instead, multiple datasets are curated to address specific challenges, with dataset size often reflecting the complexity of the labels. Access to diverse datasets is crucial for training models that can adapt to different environments and sound events [13].

Three research papers explore gunshot-related sound detection systems using various datasets, achieving different accuracy levels. The first study uses a sound event recognition model trained with 692 samples from sources like Freesound.org and YouTube, achieving an accuracy of 77.32% in indoor event classification [14]. The second paper introduces a hybrid algorithm for detecting gunshots in indoor settings, with data collected from online sources and real-world locations like shopping malls and universities, achieving accuracy between 91.65% and 94.97%, depending on the classifier [15]. The third paper presents a novel approach for recognizing environmental sounds across various settings, using a dataset of 1000 sounds across 10 categories, including gunshots, and achieving up to 92.22% classification accuracy [16].

Audio Signal Preprocessing is pivotal for the effectiveness of machine learning algorithms, particularly in creating generalized predictive models for classification tasks [17]. This phase involves a variety of techniques to prepare the audio signal for feature extraction, with the choice of methods depending on the sensitivity and accuracy requirements of the specific application or system.

Normalization is an essential technique in data analysis, especially when handling data from different sources with varying measurement scales. Dalwinder and Birmohan define data normalization as a process that resizes or transforms raw data so that each feature contributes uniformly. For example, when data includes both age (in years) and height (in centimeters), normalization ensures consistency in the magnitude of these parameters [17, 18].

In signal processing, *Noise Reduction* and *Filtering* are two fundamental methods. Noise refers to unwanted disturbances within a specific frequency range, such as electric waves and random variations [19]. Noise reduction is crucial for cleaning data that is susceptible to noise interference [18]. Filtering is closely related and focuses on eliminating noise, often using low-pass or high-pass filters to remove high-frequency noise or highlight specific data features [20]. These methods are particularly useful in medical fields like Electrocardiogram (ECG) signal analysis. ECG signals, which record heart rates, are vital for investigating abnormal heart functions, such as arrhythmias and conduction disturbances. Filters like low-pass, high-pass, and Butterworth filters are employed to preprocess these signals by effectively removing high-frequency noise, with Butterworth filters being particularly effective [21].

Another important preprocessing step is *Silence Removal*, which addresses the presence of complete silence at the beginning, end, or within audio signals. This process applies a specific threshold to remove unvoiced portions that lack relevant data, retaining only the voiced sections [22]. In studies related to cough sound detection, silence removal is crucial as it allows the focus to remain on sound events, thereby improving detection accuracy [23, 24, 25].

In practice, sounds often overlap or occur sequentially, requiring effective segmentation for classification. *Segmentation* prepares audio signals for feature extraction by dividing them into distinct segments based on temporal proximity and setting thresholds to determine the relevance of sound segments [26].

The final preprocessing method is *Feature Windowing*, which treats non-stationary signals as quasi-stationary by sliding a window over the entire signal for comprehensive analysis [27]. Unlike traditional acoustic analysis systems that divide sound recordings into fixed-sized windows, contemporary methods adapt the window size to the signal's characteristics [28, 27].

Segmentation and feature windowing are closely related but serve different purposes. While segmentation focuses on creating meaningful signal segments, feature windowing concentrates on extracting features within these windows. For instance, Baughman et al. used a peak detection method to identify specific sounds in a tennis match recording, isolating the sound of interest within a single window. This technique is valuable for classifying acoustic events using machine learning. However, if the sound of interest spans two different windows, classification accuracy may be reduced [28].

Ultimately, these preprocessing methods lead to feature extraction, where the extracted features are stored in a database for training the classifier, a topic to be discussed in the next step.

2- Feature Extraction step

Feature extraction is vital in audio content analysis as it involves creating a numerical representation, or feature vector, that captures key acoustic characteristics of audio segments [12, 13]. This vector is foundational for various audio analysis and information extraction algorithms [29], as it condenses extensive data into a more manageable format while retaining essential information [30].

Time-Frequency analysis techniques are commonly used in feature extraction to focus on the signal's frequency domain, breaking the signal into overlapping frames to track frequency distribution changes over time [30]. These techniques, such as Mel frequency cepstral coefficients (MFCCs), Log-Mel energies, and Spectrograms, divide the signal into smaller segments and calculate the frequency content for each. The resulting magnitude spectrum shows energy distribution over frequency for each segment, allowing for the computation of a concise set of features that capture fundamental spectral characteristics. This compact feature set is preferred for machine learning algorithms, as it maintains informativeness while reducing complexity, making it widely applicable in signal processing tasks [29].

Different sound extraction features are used depending on the system and its performance needs. Common methods include log-mel energies, MFCCs, spectrograms, and constant-Q filterbank-based features. For instance, Jain et al. developed ProtoSound, an interactive system that enhances sound awareness for deaf or hard-of-hearing individuals. ProtoSound personalizes sound recognition models using user recordings and log-mel spectrogram features, significantly improving performance when integrated with deep convolutional neural networks (CNN) [31].

Deep Neural Networks (DNNs) have been used to classify cough sounds by extracting MFCC features. Liu et al. [32] reported a DNN with MFCC features achieving 90.1% accuracy for positive cases and 85% for negative cases, while Amoh and Odame [33] attained 86.8% accuracy for positive cases and 92.7% for negative cases using a similar approach.

In speech-based emotion recognition, extensive reviews have compared different approaches. Studies using the RAVDESS, Emo-DB, and IEMOCAP datasets found that Log-Mel spectrogram features outperform MFCCs, challenging their prevalent use in this field [34, 35]. Other feature types are also effective for sound event recognition. For example, Sing et al. analyzed constant-Q filter bank-based time-frequency representations, offering superior frequency resolution at low frequencies compared to MFSC. Wang et al. used a model based on discrete Fourier parameters, where frequency harmonics and their derivatives effectively distinguished emotion classes. Additionally, Badshah et al. introduced a method combining spectrograms and CNN for sound event recognition, achieving promising results in emotion prediction [36, 37, 38].

3- Model Training and Classification step

During this phase, the system learns to correlate extracted audio signal features with specific class labels to develop a model for categorizing audio recordings into predefined classes. For instance, a sound scene classification system might categorize recordings as "home," "street," or "office" [12, 13].

Classification, a machine learning method, assigns input patterns to predefined categories using a classifier. This process involves two main phases: training the classifier with samples representing each class and then categorizing unknown inputs into these classes. Different classification techniques use various algorithms and rules, which can impact accuracy based on the specific application [39]. Four notable classification algorithms for sound detection are:

1. Convolutional Neural Networks (CNNs): CNNs, commonly used in deep learning, excel in image classification but require a specialized approach for sound. Sound is converted into spectrogram images, which CNNs can then analyze. CNNs consist of layers that process sound data by generating feature maps to identify patterns [40, 41]. They are effective but require longer training times [8]. For example, CNNs have been used to classify bird sounds in normal and threatened conditions by analyzing spectrograms [40].
2. Deep Neural Networks (DNNs): Unlike standard neural networks with a single hidden layer, DNNs have multiple hidden layers, mimicking the human brain's visual recognition model. This depth allows DNNs to achieve high precision by progressively recognizing complex information [29]. Research by Li et al. on DNN hyperparameters for speech recognition and audio analysis demonstrates their adaptability and high performance [8].
3. Recurrent Neural Networks (RNNs): RNNs are designed for sequence modeling, capturing temporal dependencies by considering both current input and previous hidden states. This allows RNNs to handle sequences of varying lengths and contexts. Arsenali et al. utilized RNNs for sound event classification, achieving high accuracy and sensitivity with their optimized model [42, 43].
4. Decision Trees: Decision trees, a non-linear classification method, use a hierarchical structure to eliminate classes sequentially until the correct class is reached. They are efficient for problems with many classes. The Ordinary Binary Decision Tree is a common variant, and research by Saifan et al. explored its application in sound engine classification [39].

Based on the different stages of sound detection technology, the entire process can be categorized and summarized into three primary stages: Preprocessing, Feature Extraction, and Model Training and Classification. These stages encompass the core methodologies and techniques employed in sound detection systems. Table 1 provides a detailed overview of these processes, highlighting the key steps, methods, and relevant research associated with each stage. This structured approach allows for a clearer understanding of how sound detection technology operates and the advancements made at each stage.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 10, October 2024

Table I: Summary of Key Processes in Sound Detection Technology.

Process Stage	Description	Key References	Key Techniques/Methods
Preprocessing	Enhances audio data quality before feature extraction.	[12, 13], [14], [15], [16], [17, 18], [19, 20], [21], [22], [23, 24, 25], [26], [27, 28]	Noise Reduction, Equalization, Filtering, Silence Removal, Segmentation, Feature Windowing
Feature Extraction	Converts audio data into a numerical representation (feature vector) that captures key acoustic characteristics.	[12, 13], [29], [30], [31], [32], [33], [34, 35], [36, 37, 38]	MFCCs, Log-Mel Energies, Spectrograms, Time-Frequency Analysis
Model Training & Classification	Develops models to classify audio recordings into predefined classes based on extracted features.	[12, 13], [39], [40, 41], [42, 43]	CNNs, DNNs, RNNs, Decision Trees

The table categorizes sound detection technology into three main stages, highlighting key processes and methods employed in each stage, alongside relevant research, to present a clear overview of advancements in sound detection technology.

B. Real-time Sound Detection:

Real-time processing refers to the immediate handling of data as it becomes available, with two main requirements: the processing must be completed faster than the data's duration, and delays should be minimized to ideally zero. In computer operating systems, "real-time" also involves precise scheduling to handle events within a set timeframe [44].

Signal classification can be divided into three categories:

1. Analog Signal Processing: Deals with signals that have not been digitized, such as those from radios or older televisions.
2. Continuous Signal Processing: Focuses on signals with continuous amplitude variations, including modeling continuous systems, system function adjustments, and time-based filtering.
3. Discrete Signal Processing: Handles signals sampled and quantized at specific intervals, represented as a sequence of numbers. Discrete-time signals are crucial for real-time applications like sound event detection and speech recognition due to their ability to capture and process temporal dynamics [45].

Non-real-time signal processing involves manipulating pre-gathered and digitized signals without real-time constraints, while real-time processing demands precise timing from both hardware and software [46]. Key strategies for real-time sound detection include:

1. On-line Processing: Manages live data streams, computing results as the data is recorded or transmitted, typically using short buffers.
2. Incremental Processing: Optimizes on-line processing by reducing the delay between data input and analytical results.

Digital Signal Processing (DSP) significantly enhances real-time sound detection by digitally representing and analyzing signals. DSP involves breaking down signals, applying mathematical operations like filtering and Fourier transforms, and integrating results for effective analysis. Recent advancements in DSP technology have enabled real-time applications where analog methods are impractical [46].

An example of DSP application is in real-time arrhythmia classification, involving three stages: preprocessing to reduce noise, feature extraction using techniques like Wavelet Transform, and classification with algorithms such as probabilistic neural networks. DSP is essential in minimizing noise, extracting features, and classifying arrhythmias [47].

C. Challenges and Limitations:

Creating automatic systems for sound event detection is a complex task with several challenges, particularly related to sound characteristics, data collection, and annotation. Addressing these challenges is crucial for the effectiveness of machine learning techniques [13]. Recent advancements offer potential solutions to these issues, which are detailed below.

A major challenge in sound event detection is handling overlapping sound events, a task known as polyphonic sound event detection. This involves identifying all coinciding sounds simultaneously [48]. To tackle this, supervised classification methods like RNNs [49, 50] and CNNs [51] are commonly used. These methods predict the presence of each sound event on a frame-by-frame basis, helping to manage the complexity of real-world sound environments.

Data challenges, such as insufficient samples and class imbalance, also impact classifier performance. Data augmentation, which involves synthetically increasing the data through techniques like pitch shifting, noise removal, compression, and time stretching, is a key strategy to improve the system's performance by enhancing data representation [52].

Traditional audio processing often separates feature representation from classifier design, which can lead to suboptimal features. Deep Neural Networks (DNNs) address this by integrating feature extraction and classification, optimizing both processes simultaneously. For example, in speech recognition, lower DNN layers adapt to speaker

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 10, October 2024

characteristics while upper layers focus on class discrimination [42].

Another challenge is weakly labeled data, where only the presence or absence of events is known, but not their exact timings. Multiple Instance Learning (MIL) is a useful approach for dealing with such data. MIL treats the entire audio clip as a "bag" for classification when annotations are partial, allowing for effective sound event detection despite weak labels [53, 54].

V. SPEECH RECOGNITION

Recent advancements in speech recognition have enhanced communication across languages and interactions with various devices. These improvements stem from three main factors: increasing computational power from multi-core processors and GPU clusters, access to extensive datasets, and the rise of mobile, wearable, and smart home technologies [55].

Speech technology impacts both human-to-human (HHC) and human-to-machine communication (HMC). In HHC, research has focused on assisting individuals with speech impairments through applications like Google's Speech API, which converts speech to text with high accuracy [56]. Studies also address transcription for the deaf and hard of hearing and explore deep learning models for language translation [57][58].

In HMC, advancements include voice-activated smart home systems, though these often overlook less common languages like Romanian. To address this, new acoustic and grammar models for Romanian have been developed, alongside remote speaker recognition techniques [59][60]. Additionally, the field covers voice search and interaction with mobile devices and entertainment systems [61].

This paper emphasizes automatic speech recognition as a critical component in spoken language technology, focusing on its role in real-time audio transcription.

A. Automatic Speech Recognition:

In this book [55], the architecture of automatic speech recognition (ASR) is elucidated, comprising four primary components: Signal processing and feature extraction, acoustic model (AM), language model (LM), and hypothesis search. Signal processing readies audio input by converting it into feature vectors. The AM evaluates the likelihood of feature sequences, while the LM estimates word sequence probabilities. The hypothesis search combines AM and LM scores to yield the recognition result. The AM must address challenges such as variable-length feature vectors and acoustic variability, stemming from factors like speaker characteristics, speech style, noise, and accents. Real-world ASR systems encounter further complexities, including extensive vocabularies, spontaneous speech, and multilingual contexts. While traditional ASR systems employed features like MFCC and RASTA-PLP with GMM-HMM models trained via maximum likelihood criteria, recent advances include discriminative training methods such as MCE and MPE in the 2000s and the adoption of DNNs and discriminative hierarchical models like CD-DNN-HMM, substantially improving accuracy due to enhanced computational capabilities and more extensive training data.

VI. RELATED APPLICATIONS FOR DEAF AND HARD-OF-HEARING INDIVIDUALS

This section will present research papers and applications aimed at improving the lives of deaf and hard-of-hearing individuals. Numerous studies focus on leveraging recent technologies, such as sound detection and speech recognition, to enhance their safety and overall quality of life.

A. Related Research Papers

The Enssat application [64] utilizes Google Glass as a wearable device to support individuals who are deaf or hard of hearing in both Arabic and English. This application primarily focuses on functions such as sound detection and speech recognition. The sound detection feature of Enssat utilizes the microphone of either a mobile phone or Google Glass to identify ambient sounds. When a sound is detected, a snippet of it is recorded and sent to another thread for identification. The system then compares this recorded sound with a set of stored sounds using the MusicG library to determine the degree of similarity. Regarding speech recognition, Enssat offers real-time transcription of speech. It employs Google's Speech-to-Text service to convert audio files into text. The challenge lies in effectively processing continuous speech and segmenting it into snippets for accurate transcription. Additionally, the Enssat application provides translation capabilities, enabling the translation of both spoken words and text captured in images. To achieve this, the system utilizes Google's Translation API to deliver real-time translation services.

A different document explores the development of a specialized virtual assistant catering to individuals with hearing impairments, highlighting its proficiency in identifying and categorizing various sounds. The proposed remedy encompasses a sound classification module, a gesture recognition module, and a multilingual translation module. The sound classification module is engineered to recognize and categorize diverse sounds, such as those produced by vehicles, to notify users of potential hazards. It leverages audio data from the UrbanSound8K dataset and employs a deep neural network for sound recognition. The gesture recognition module translates gestures from Indian Sign Language into text and audio, facilitating communication between non-deaf individuals and those with hearing impairments. The multilingual translation module converts the generated text into various regional languages, offering translation services for hearing-impaired individuals in India. This solution is seamlessly integrated into an Android application and has undergone a comparative analysis with existing apps, assessing factors such as response time, accuracy, output predictions, and alert systems [6].

Saifan presents the Deaf Assistant Digital System [39], a solution utilizing smartphones to provide alerts for individuals with hearing impairments across various situations. Employing vibrations and visual notifications on the smartphone screen, the system ensures effective communication with the user. The paper delves into the technical intricacies of the speech and sound recognition engines, detailing the utilization of deep auto-encoder-based low-dimensional feature extraction from FFT spectral envelopes. This approach enables the identification of diverse sounds and words. Additionally, the paper highlights the application of Praat, a computer program for speech analysis, synthesis, and manipulation, in extracting sound features. The system's matching engines play a crucial role in comparing recognized words and sounds with predefined cautionary words and sound alerts. Upon identifying a match, these engines trigger appropriate actions such as vibration and visual effects to alert individuals with hearing impairments.

In this paper [65], the significance of sound detection is examined, and diverse technologies and systems designed for this objective are explored. The paper emphasizes the growing presence of comprehensive sound detection systems in the market. In the realm of sound detection technology, a proof-of-concept for a sound detection algorithm based on Gaussian Mixture Model (GMM) is discussed. Additionally, the paper briefly touches upon the application of Gaussian Mixture Models for speaker identification and verification in the context of speech recognition.

The iHelp application [66] introduces a real-time mobile emergency assistance system designed to aid deaf-mute individuals or elderly individuals living alone in promptly and effectively reporting emergencies. This system employs mobile application software installed on smartphones, enabling users to report emergencies through SMS, even in the absence of internet access. By doing so, it optimizes the dispatching of rescue units and enhances the overall success rate of emergency rescue operations. The system is comprised of three key components: the report subsystem, dispatch system, and rescue subsystem.

B. Related Mobile Applications

The Sound Alert App functions as a tool for capturing and informing users of significant environmental and household occurrences. It can identify various sounds like doorbells, phone rings, microwave beeps, alarms, and intercoms without the need for pre-recording. What distinguishes this solution is its seamless integration with existing building infrastructure and alarm systems, providing a cost-effective alternative to flashy lights and expensive hardware setups. This app proves especially beneficial for individuals who are hard-of-hearing, deaf, elderly, or heavy sleepers. By activating "Detection Mode," the app's intelligent algorithm continuously monitors the environment through the smartphone's microphone. Additionally, it can sync with Pebble Watch for extra notification options, including vibration, flashing lights, and on-screen icons with event names [67]. Another similar app, The Deaf and Hearing Impaired (APK), is designed to assist deaf or hearing-impaired individuals by alerting them through vibration and flashlight signals when a loud sound occurs nearby [68].

The Android app Live Transcribe & Sound Notifications enhances accessibility for individuals with hearing impairments. It enables real-time transcription of spoken words in over 80 languages and dialects, allowing users to customize word additions. The app also notifies users of various sounds, including potentially risky situations. Users have the flexibility to adjust settings, save transcriptions for three days, and search within saved transcriptions. Developed in collaboration with Gallaudet University, a leading institution for the deaf and hard of hearing, the app is compatible with Android 6.0 and newer devices. It incorporates features such as vibrating when the user's name is spoken and supports external microphones for improved audio reception [69].

Rogervoice, a revolutionary call transcription application, has transformed phone communication for individuals who are deaf or hard of hearing. By offering real-time call subtitles in over 80 languages, it enables users to independently connect with family, friends, medical professionals, and customer service helplines. The app is designed to be user-friendly, allowing calls to be initiated either from contacts or by entering numbers manually. Conversations are transcribed instantly, and users have the option to respond through speech or typing, with a voice synthesizer delivering text messages. It's important to note that Rogervoice does not support emergency calls or premium-rate numbers. Subscriptions are required for calls to individuals who do not use the application, and pricing details can be found on the website [70].

In the pursuit of modernizing communication within the realm of security and improving the efficiency of security personnel in managing emergency reports, the General Directorate of Public Security has introduced the "Kulluna Amn" mobile application. This application is designed to actively involve citizens and residents in the security framework and has been launched with the direct support and guidance of HRH Prince Mohammed bin Nayef, Deputy Prime Minister and Minister of Interior. The app empowers users to report unusual incidents, which are then transmitted to the thirty-nine operation rooms situated across the kingdom. Users can furnish details about incidents, including photos and GPS location, choose the incident category, and even pinpoint the nearest police or traffic department based on their current geographical coordinates. "Kulluna Amn" signifies a noteworthy stride in enhancing emergency response and encouraging public participation in upholding security [71].

IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 10, October 2024

The TapSOS application serves as a crucial tool for reaching Emergency Services in situations where verbal communication is challenging or unsafe, especially for individuals with hearing impairments. Users establish profiles containing essential information, which is communicated to Emergency Call Handlers through visual icons. A medical profile provides valuable information for First Responders. The GPS feature automatically identifies the user's location, allowing manual adjustments for precision. By responding to a series of questions aligned with Emergency Services protocols, users trigger alerts that are directly transmitted to the UK's 999 Emergency Call Handlers. This makes TapSOS an indispensable tool for non-verbal emergency communication [72].

This section explores various applications and research papers dedicated to enhancing the lives of deaf and hard-of-hearing individuals. Leveraging technologies like sound detection and speech recognition, these innovations aim to improve safety and overall quality of life. From specialized virtual assistants to real-time emergency assistance systems and mobile apps like Sound Alert App and Live Transcribe & Sound Notifications, the focus is on providing timely alerts, communication aids, and accessibility features tailored to the needs of this community. Table 2. Summarizing the related applications and research papers aimed at improving the lives of deaf and hard-of-hearing individuals.

Table II: Summary of related applications and research papers

Category	**Application/ Research Paper**	**Description**	**Reference**
Related Research Papers	Enssat Application	Utilizes Google Glass for sound detection and speech recognition, offers real-time transcription and translation.	[64]
	Specialized Virtual Assistant	Features sound classification, gesture recognition, and multilingual translation for individuals with hearing impairments.	[6]
	Deaf Assistant Digital System	Uses smartphones to provide alerts via vibrations and visual notifications, with sound and speech recognition.	[39]
	Sound Detection Technology	Discusses a proof-of-concept for sound detection using Gaussian Mixture Models (GMM).	[65]
	iHelp Application	Mobile emergency assistance system for deaf-mute or elderly individuals, includes reporting emergencies via SMS.	[66]
Related Mobile Applications	Sound Alert App	Identifies and alerts users to significant sounds, integrates with building infrastructure, and supports Pebble Watch.	[67]
	The Deaf and Hearing Impaired (APK)	Alerts users to loud sounds nearby through vibrations and flashlight signals.	[68]
	Live Transcribe & Sound Notifications	Provides real-time transcription of spoken words and notifications of various sounds in over 80 languages.	[69]
	Rogervoice	Real-time call transcription application with subtitles in over 80 languages, enabling phone communication.	[70]
	Kulluna Amn	Mobile app for reporting incidents to security forces, includes GPS location and incident details.	[71]
	TapSOS	Emergency communication app for non-verbal communication, provides visual icons and GPS location for emergency services.	[72]

This table reviews technologies and research designed to improve the lives of deaf and hard-of-hearing individuals, focusing on sound detection and speech recognition.

VII. CONCLUSION

In summary, incorporating advanced technologies especially in the areas of machine learning and artificial intelligence shows a lot of potentials to improve the safety and the quality of life of hearing-impaired individuals. This paper has analyzed assistive technologies, sound recognition, and sonification for the enhancement of the functionality of patients suffering from hearing loss. Hearing aids, cochlear implants, and mobile apps breakdown communication barriers and promote self-sufficiency, effectively changing the status quo of everyday routines. On the other hand, Machine learning-centered technologies for sound detection provide advanced means of real-time situational awareness for safety, security, and environmental monitoring.

The key findings of this review highlight the transformative impact of these technologies. Augmented hearing aids and cochlear implants are essential; both assist in increasing overall auditory fulfillment, while empowering a sense of independence. Sound detection systems driven by machine learning based models, generate alerts in time and also provide contextual information when required that specifically addresses the safety problems encountered to people with hearing impairments. Further, the evolution of speech recognition technologies enable communication across disparate contexts enabling greater context-aware support for integration and cross interaction within different environments.

Nevertheless, there is still room for improvement such as resolving these conflicting sound events and expanding the data sets which already point to conclusions in more research work. Future studies should address these issues and examine other possibilities to use these technologies more efficiently. These innovations possess the possibility of improving the quality of life of the hearing impaired by averting potential threats or achieving safe compliance in the access environment as well as in communication enabling ease of movement within the environment.

REFERENCES

[1] World Health Organization. Deafness and hearing loss. 2023. Available at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (Accessed: 14 September 2023).

[2] Abdallah, E.E., Fayyoumi, E. Assistive technology for deaf people based on Android platform. Procedia Computer Science. 2016; 94: 295–301. doi:10.1016/j.procs.2016.08.044.

[3] World Health Organization. Improving access to assistive technology. 2018. Available at: http://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_21-en.pdf (Accessed: 14 September 2023)

[4] Kbar, G., et al. Assistive technologies for hearing, and speaking impaired people: A survey. Disability and Rehabilitation: Assistive Technology. 2016; 12(1): 3–20. doi:10.3109/17483107.2015.1129456.

[5] Advanced Solutions International, Inc. Educational Technology. The Delta Kappa Gamma Bulletin. Summer 2012. Available at: https://www.media.mit.edu/~mres/papers/educational-technology-2012.pdf (Accessed: 14 September 2023).

[6] Ozarkar, S., et al. AI for accessibility: Virtual assistant for hearing impaired. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2020. doi:10.1109/icccnt49239.2020.9225392.

[7] Alqudaihi, K.S., et al. Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities. IEEE Access. 2021; 9: 102327–102344. doi:10.1109/access.2021.3097559.

[8] Li, J., et al. A comparison of deep learning methods for environmental sound detection. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. doi:10.1109/icassp.2017.7952131.

[9] Lee, Y.-C., et al. Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents. Automation in Construction. 2020; 113: 103127. doi:10.1016/j.autcon.2020.103127.

[10] Virtanen, T., Plumbley, M.D., Ellis, D. Introduction to Sound Scene and Event Analysis. Computational Analysis of Sound Scenes and Events. 2017: 3–12. doi:10.1007/978-3-319-63450-0_1.

[11] Rebala, G., Ravi, A., Churiwala, S. Machine learning definition and basics. An Introduction to Machine Learning. 2019: 1–17. doi:10.1007/978-3-030-15729-6_1.

[12] Babaee, E., et al. An Overview of Audio Event Detection Methods from Feature Extraction to Classification. Applied Artificial Intelligence. 2017; 31(9–10): 661–714. doi:10.1080/08839514.2018.1430469.

[13] Mesaros, A., et al. Sound event detection: A tutorial. IEEE Signal Processing Magazine. 2021; 38(5): 67–83. doi:10.1109/msp.2021.3090678

[14] Min, K., et al. Sound event recognition-based classification model for automated emergency detection in indoor environment. Advances in Informatics and Computing in Civil and Construction Engineering. 2018: 529–535. doi:10.1007/978-3-030-00220-6_63.

[15] Rahman, S.U., et al. Hybrid system for automatic detection of gunshots in indoor environment. Multimedia Tools and Applications. 2020; 80(3): 4143–4153. doi:10.1007/s11042-020-09936-w.

[16] Souli, S., Lachiri, Z. Audio sounds classification using scattering features and support vectors machines for medical surveillance. Applied Acoustics. 2018; 130: 270–282. doi:10.1016/j.apacoust.2017.08.002.

[17] Singh, D., Singh, B. Investigating the impact of data normalization on classification performance. Applied Soft Computing. 2020; 97: 105524. doi:10.1016/j.asoc.2019.105524.

[18] Caixinha, M., Nunes, S. Machine learning techniques in clinical vision sciences. Current Eye Research. 2016; 42(1): 1–15. doi:10.1080/02713683.2016.1175019.

[19] Fink, D. A new definition of noise: Noise is unwanted and/or harmful sound. noise is the new “secondhand smoke”. Proceedings of Meetings on Acoustics. 2019. doi:10.1121/2.0001186.

[20] Cadena, L., et al. Noise reduction techniques for processing of medical images. Noise Reduction Techniques for Processing of Medical Images. 2017. Available at: http://www.iaeng.org/publication/WCE2017/WCE2017_pp496-500.pdf (Accessed: 28 September 2023).

[21] Celin, S., Vasanth, K. ECG signal classification using various machine learning techniques. Journal of Medical Systems. 2018; 42(12). doi:10.1007/s10916-018-1083-6.

[22] Deshmukh, G., et al. Speech based emotion recognition using machine learning. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). 2019. doi:10.1109/iccmc.2019.8819858.

[23] Adhi Pramono, R.X., Anas Imtiaz, S., Rodriguez-Villegas, E. Automatic identification of cough events from acoustic signals. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019. doi:10.1109/embc.2019.8856420.

[24] Cohen-McFarlane, M., Goubran, R., Knoefel, F. Comparison of silence removal methods for the identification of audio cough events. 2019 41st Annual International Conference of

the IEEE Engineering in Medicine and Biology Society (EMBC). 2019. doi:10.1109/embc.2019.8857889.
[25] Hall, J.I., et al. The present and future of cough counting tools. Journal of Thoracic Disease. 2020; 12(9): 5207–5223. doi:10.21037/jtd-2020-icc-003
[26] Tsalera, E., et al. CNN-based segmentation and classification of sound streams under realistic conditions. Proceedings of the 26th Pan-Hellenic Conference on Informatics. 2022. doi:10.1145/3575879.3576020.
[27] Sharma, G., Umapathy, K., Krishnan, S. Trends in audio signal feature extraction methods. Applied Acoustics. 2020; 158: 107020. doi:10.1016/j.apacoust.2019.107020.
[28] Baughman, A., et al. Detection of tennis events from acoustic data. Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports. 2019. doi:10.1145/3347318.3355520.
[29] Tzanetakis, G. Manipulation, analysis and retrieval systems for audio signals. Ann Arbor, MI: ProQuest. 2004.
[30] Ramadan, M., Bani Younes, A., Moheidat, J. Robust sound detection & localization algorithms for robotics applications. AIAA SciTech 2019 Forum. 2019. doi:10.2514/6.2019-2046.
[31] Jain, D., et al. ProtoSound: A personalized and scalable sound recognition system for deaf and hard-of-hearing users. CHI Conference on Human Factors in Computing Systems. 2022. doi:10.1145/3491102.3502020.
[32] Liu, J.-M., et al. Cough detection using Deep Neural Networks. 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2014. doi:10.1109/bibm.2014.6999220.
[33] Amoh, J., Odame, K. Deep neural networks for identifying cough sounds. IEEE Transactions on Biomedical Circuits and Systems. 2016; 10(5): 1003–1011. doi:10.1109/tbcas.2016.2598794.
[34] Venkataramanan, K., Rajamohan, H.R. Emotion recognition from speech. arXiv.org. 2019. Available at: https://arxiv.org/abs/1912.10458 (Accessed: 01 October 2023).
[35] Pandey, S.K., et al. A survey on the use of audio data for medical diagnosis. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). 2019. doi:10.1109/smc.2019.8914461.
[36] Singh, D., et al. Deep learning techniques for sound classification. IEEE Transactions on Audio, Speech, and Language Processing. 2020; 28(2): 286–297. doi:10.1109/taslp.2019.2950487.
[37] Kunxia Wang et al. Speech emotion recognition using Fourier parameters. IEEE Transactions on Affective Computing. 2015; 6(1): 69–75. doi:10.1109/taffc.2015.2392101.
[38] Badshah, A.M. et al. Speech emotion recognition from spectrograms with deep convolutional neural network. 2017 International Conference on Platform Technology and Service (PlatCon) [Preprint]. doi:10.1109/platcon.2017.7883728
[39] Saifan, R.R., Dweik, W. and Abdel-Majeed, M. A machine learning based Deaf Assistance Digital System. Computer Applications in Engineering Education. 2018; 26(4): 1008–1019. doi:10.1002/cae.21952.
[40] Permana, S.D. et al. Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm. Journal of King Saud University - Computer and Information Sciences. 2022; 34(7): 4345–4357. doi:10.1016/j.jksuci.2021.04.013.
[41] Kattenborn, T. et al. Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. ISPRS Journal of Photogrammetry and Remote Sensing. 2021; 173: 24–49. doi:10.1016/j.isprsjprs.2020.12.010.
[42] Purwins, H. et al. Deep Learning for Audio Signal Processing. IEEE Journal of Selected Topics in Signal Processing. 2019; 13(2): 206–219. doi:10.1109/jstsp.2019.2908700.
[43] Arsenali, B. et al. Recurrent neural network for classification of snoring and non-snoring sound events. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Preprint]. doi:10.1109/embc.2018.8512251.
[44] Eyben, F. Real-time speech and music classification by large audio feature space extraction. Cham: Springer International Publishing. 2018.
[45] Dastres, R. and Soori, M. A review in Advanced Digital Signal Processing Systems. ResearchGate. Available at: https://www.researchgate.net/publication/350449625_A_Review_in_Advanced_Digital_Signal_Processing_Systems (Accessed: 05 October 2023).
[46] Kuo, S.M., Lee, B.H. and Tian, W. Real-time Digital Signal Processing: Fundamentals, implementations and applications. Chichester: John Wiley & Sons. 2017.
[47] Gutiérrez-Gnecchi, J.A. et al. DSP-based arrhythmia classification using wavelet transform and Probabilistic Neural Network. Biomedical Signal Processing and Control. 2017; 32: 44–56. doi:10.1016/j.bspc.2016.10.005
[48] Adavanne, S. et al. Sound event localization and detection of overlapping sources using convolutional recurrent neural networks. IEEE Journal of Selected Topics in Signal Processing. 2019; 13(1): 34–48. doi:10.1109/jstsp.2018.2885636.
[49] Parascandolo, G., Huttunen, H. and Virtanen, T. Recurrent neural networks for polyphonic sound event detection in real life recordings. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [Preprint]. doi:10.1109/icassp.2016.7472917.
[50] Adavanne, S. et al. Sound event detection in multichannel audio using spatial and harmonic features. arXiv.org. Available at: https://arxiv.org/abs/1706.02293 (Accessed: 05 October 2023).
[51] Phan, H. et al. Robust audio event recognition with 1-max pooling Convolutional Neural Networks. Interspeech 2016 [Preprint]. doi:10.21437/interspeech.2016-123.
[52] Abayomi-Alli, O.O. et al. Data augmentation and deep learning methods in sound classification: A systematic review. Electronics. 2022; 11(22): 3795. doi:10.3390/electronics11223795.
[53] Kong, Q. et al. Sound event detection of weakly labelled data with CNN-transformer and automatic threshold optimization. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2020; 28: 2450–2460. doi:10.1109/taslp.2020.3014737.
[54] Carbonneau, M.-A. et al. Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition. 2018; 77: 329–353. doi:10.1016/j.patcog.2017.10.009.
[55] Yu, D. and Deng, L. Automatic speech recognition: A deep learning approach. London: Springer. 2016.
[56] Anggraini, N. et al. Speech recognition application for the speech impaired using the Android-based Google Cloud Speech API. TELKOMNIKA (Telecommunication

Computing Electronics and Control). 2018; 16(6): 2733. doi:10.12928/telkomnika.v16i6.9638.

[57] Gaur, Y. The effects of automatic speech recognition quality on human transcription latency. Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15 [Preprint]. doi:10.1145/2700648.2811331.

[58] Gibadullin, R.F., Perukhin, M.Yu. and Ilin, A.V. Speech recognition and machine translation using Neural Networks. 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM) [Preprint]. doi:10.1109/icieam51226.2021.9446474.

[59] Caranica, A. et al. Speech recognition results for voice-controlled assistive applications. 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD) [Preprint]. doi:10.1109/sped.2017.7990438.

[60] Peng, S. et al. Remote speaker recognition based on the enhanced LDV-Captured speech. Applied Acoustics. 2019; 143: 165–170. doi:10.1016/j.apacoust.2018.08.007.

[61] S, K. and E, C. A review on Automatic Speech Recognition Architecture and approaches. International Journal of Signal Processing, Image Processing and Pattern Recognition. 2016; 9(4): 393–404. doi:10.14257/ijsip.2016.9.4.34.

[62] Alharbi, A., Alzuwaed, J. and Qasem, H. Evaluation of e-health (SEHA) application: A cross-sectional study in Saudi Arabia. BMC Medical Informatics and Decision Making. 2021; 21(1). doi:10.1186/s12911-021-01437-6.

[63] Al-Wathinani, A. et al. Increased emergency calls during the COVID-19 pandemic in Saudi Arabia: A national retrospective study. Healthcare. 2020; 9(1): 14. doi:10.3390/healthcare9010014.

[64] Alkhalifa, S. and Al-Razgan, M. ENSSAT: Wearable Technology application for the deaf and hard of hearing. Multimedia Tools and Applications. 2018; 77(17): 22007–22031. doi:10.1007/s11042-018-5860-5.

[65] Bragg, D., Huynh, N. and Ladner, R.E. A personalizable mobile sound detector app design for deaf and hard-of-hearing users. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility [Preprint]. doi:10.1145/2982142.2982171.

[66] Chen, L.-B. et al. A real-time mobile emergency assistance system for helping deaf-mute people/elderly singletons. 2016 IEEE International Conference on Consumer Electronics (ICCE) [Preprint]. doi:10.1109/icce.2016.7430516.

[67] Get alerts to sounds around you (no date) Sound Alert | Convert Sounds Into Visual And Sensory Notifications. Available at: http://www.soundalert.co/ (Accessed: 13 October 2023).

[68] The deaf and hearing impaired apk for Android download (2015) APKPure.com. Available at: https://apkpure.com/the-deaf-and-hearing-impaired/kr.ac.kaist.isilab.doyeob.humanityProject2 (Accessed: 13 October 2023).

[69] Live Transcribe & Notification - apps on Google Play (no date) Google. Available at: https://play.google.com/store/apps/details?id=com.google.audio.hearing.visualization.accessibility.scribe (Accessed: 13 October 2023).

[70] Caption all your phone calls instantly! (no date) Rogervoice. Available at: https://rogervoice.com/en/ (Accessed: 13 October 2023).

[71] كلنا أمن - التطبيقات على google play (no date) Google. Available at: https://play.google.com/store/apps/details?id=sa.gov.moi.securityinform&hl=ar (Accessed: 13 October 2023).

[72] Home (2023) TapSOS. Available at: https://tapsos.com/ (Accessed: 13 October 2023).