Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1013234, 12 pages
https://doi.org/10.1155/2018/1013234

Research Article

A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples

Kanwal Yousaf[1], **Zahid Mehmood**[1], **Tanzila Saba**[2], **Amjad Rehman**[3], **Muhammad Rashid**[4], **Muhammad Altaf**[5] and **Zhang Shuguang**[6]

[1]*Department of Software Engineering, University of Engineering and Technology, Taxila 47050, Pakistan*
[2]*College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia*
[3]*College of Computer and Information Systems, Al-Yamamah University, Riyadh 11512, Saudi Arabia*
[4]*Department of Computer Engineering, Umm Al-Qura University, Makkah 21421, Saudi Arabia*
[5]*Department of Mathematics, University of Engineering and Technology, Taxila 47050, Pakistan*
[6]*Department of Statistics and Finance, University of Science and Technology of China, Hefei 23026, China*

Correspondence should be addressed to Kanwal Yousaf; kanwal.yousaf@uettaxila.edu.pk

Received 15 January 2018; Accepted 17 April 2018; Published 24 May 2018

Academic Editor: Seyed M. Buhari

Mobile technology is very fast growing and incredible, yet there are not much technology development and improvement for Deaf-mute peoples. Existing mobile applications use sign language as the only option for communication with them. Before our article, no such application (app) that uses the disrupted speech of Deaf-mutes for the purpose of social connectivity exists in the mobile market. The proposed application, named as vocalizer to mute (V2M), uses automatic speech recognition (ASR) methodology to recognize the speech of Deaf-mute and convert it into a recognizable form of speech for a normal person. In this work mel frequency cepstral coefficients (MFCC) based features are extracted for each training and testing sample of Deaf-mute speech. The hidden Markov model toolkit (HTK) is used for the process of speech recognition. The application is also integrated with a 3D avatar for providing visualization support. The avatar is responsible for performing the sign language on behalf of a person with no awareness of Deaf-mute culture. The prototype application was piloted in social welfare institute for Deaf-mute children. Participants were 15 children aged between 7 and 13 years. The experimental results show the accuracy of the proposed application as 97.9%. The quantitative and qualitative analysis of results also revealed that face-to-face socialization of Deaf-mute is improved by the intervention of mobile technology. The participants also suggested that the proposed mobile application can act as a voice for them and they can socialize with friends and family by using this app.

1. Introduction

Historically the term *deaf-mute* referred to the person who was either deaf using sign language as a source of communication or both deaf and unable to speak. This term continues to be used to refer to the person who is deaf but has some degree of speaking ability [1]. In deaf community, the word *deaf* is spelled in two separate ways. The small "d" deaf represents a person's level of hearing through audiology and being not associated with the other members of the deaf community whereas the capital "D" Deaf indicates the culturally Deaf people who use sign language for communication [2].

According to world federation of the deaf (WFD) over 5% of the world's population (≈360 million people) has disabling hearing loss including 328 million adults and 32 million children [3]. The degree of hearing loss is categorized into mild, moderate, severe, or profound levels [4]. Hearing loss of a person has a direct impact on his/her speech and language development. People with severe or profound hearing loss have higher voice handicap index (VHI) scores than those who suffer from mild hearing loss [5]. A person with mild hearing loss has less problems in speech development as he/she might not be able to hear certain sounds and the speech clarity is not affected that much. A person with severe or profound hearing

loss can have a severe problem in speech development and usually relies on sign language as a source of communication.

Deaf people face many irritations and frustrations that limit their ability to do everyday tasks. Research indicated [6] that Deaf people, especially Deaf children, have high rates of behavioral and emotional issues in relation to different methods of communication. Most people with such disabilities become introverts and resist social connectivity and face-to-face socialization. The inability to speak with family and friends can cause low self-esteem and may result in social isolation of Deaf person. It is not only that they lack social interactions but communication is also a major barrier to Deaf-mute healthcare [7]. In such conditions, it becomes difficult for the caretaker to interact with the deaf person.

Different medical treatments are available for the deaf community in order to get rid of their deafness but the cost of these treatments are expensive [8]. A report of world health organization (WHO) 2017 [9] states that there are different types of costs associated with hearing loss, which are as follows: (1) direct costs: they include the cost associated with hearing loss incurred by healthcare systems; some other types of direct costs include the education support for such children; (2) indirect costs: they include the loss of productivity and usually refer to the cost of individual being unable to contribute to the economy; and (3) intangible costs: they refer to the stigma experienced by the families that are experiencing the hearing loss. This report concludes that unaddressed hearing loss poses substantial costs to the healthcare system and to the economy as a whole.

Many communication channels are available, through which Deaf-mute people can deliver their messages, e.g., notes, helper pages, sign language, books with letters, lip reading, and gestures. Despite these channels, there are many problems which are encountered by Deaf-mutes and normal people during communication. The problem is not confined only to a Deaf-mute person who is unable to hear or speak, but another problem is lack of awareness of Deaf culture by normal people. Majority of hearing people have either no/little knowledge or experience of sign language [10]. There are also more than 300 sign languages and it is hard for a normal person to understand and become used to these languages [11]. The above-mentioned problems can be solved by involving the assistive technology as it can be used as an interpreter for converting the sign languages into text or speech for better communication between the Deaf community and hearing individuals [12]. Other technologies such as speech technologies can assist in different ways to help people with hearing loss by improving their autonomy [13]. A common example of speech technology is speech recognition, also termed as automatic speech recognition (ASR). It is the process of converting the speech signal into sequences of words with the help of an algorithm [14]. The ASR process comprises three steps, i.e., (1) feature extraction, (2) acoustic model generation, and (3) recognition phase [15, 16]. For feature extraction, MFCC is the most commonly used technique [17, 18]. The success of MFCC makes it the standard choice in the state-of-the-art speech recognizers such as HTK [19].

The main purpose of this research paper is to use a mobile-based assistive technology for providing a simple and cost-effective solution for Deaf-mute with little or complete speech development. The proposed system used HTK based speech recognizer to identify the speech of Deaf-mute and provide a communication platform for them. The next two sections explain the related work and proposed methodology of our system. Section 4 states the experimental setup and results of the proposed system.

2. Related Work

The Deaf community is not a monolithic group; it has a diversity of groups which are as follows [20, 21]:

(1) Hard-of-hearing people: they are neither fully deaf nor fully hearing, also known as culturally marginal people [22]. They can obtain some useful linguistic information from speech.

(2) Culturally deaf people: they might belong to deaf families and use sign language as the primary source of communication. Their voice (speech clarity) may be disrupted.

(3) Congenital or prelingual deaf people: they are deaf by birth or become deaf before they learn to talk and are not affiliated with Deaf culture. They might or might not use sign language based communication.

(4) Orally educated or postlingual deaf people: they have been deafened in their childhood but developed the speaking skills.

(5) Late-deafened adults: they have had the opportunity to adjust their communication techniques as their progressive hearing losses.

Each group of a Deaf community has a different degree of hearing loss and use a different source of communication. Table 1 illustrates the details of Deaf community groups with their degree of hearing loss and source of communication with others.

Hearing loss or deafness has a direct impact on communication, educational achievements, or social interactions [23]. Lack of knowledge about Deaf culture is documented in society as well as in healthcare environment [24]. Kuenburg et al. also indicated that there are significant challenges in communication among healthcare professionals and Deaf people [25]. Improvement in healthcare access among Deaf people is possible by providing the sign language supported visual communication and implementation of communication technologies for healthcare professionals. Some of the implemented technology-based approaches for facilitating Deaf-mutes with easy-to-use services are as follows.

2.1. Sensor-Based Technology Approach. Sensors based assistance can be used for solving the social problems of Deaf-mute by bridging the communication gap. Sharma et al. used wearable sensor gloves for detecting the hand gestures of sign language [26]. In this system, flex sensors were used to record the sign language and to sense the environment. The hand gesture of a person activates glove, and flex sensors on glove convert those gestures into electrical signals. The signals

Table 1: Mapping of Deaf community groups with a degree of hearing loss and communication source [3, 20, 21].

Deaf Community Groups	Degree of Hearing Loss	Communication Source
Hard-of-Hearing People	Mild to Severe	Speech/Sign Language
Culturally Deaf People	Profound	Sign Language
Congenital or Pre-lingual Deaf People	Profound	Sign Language
Orally Educated or Post-lingual Deaf People	Severe to Profound	Speech/Sign Language
Late-Deafened Adults	Moderate to Profound	Speech/Sign Language

are then matched from the database and converted into corresponding speech and displayed on LCD. The cost-effective sensor-based communication device [27] was also suggested for Deaf-mute people to communicate with the doctor. This experiment used a 32-bit microcontroller, LCD to display the input/output, and a processing unit. The LCD displays different hand sign language based pictures to the user. The user selects relevant pictures to describe the illness symptoms. These pictures then convert into patterns and pair with words to make sentences. Vijayalakshmi and Aarthi used flex sensors on the glove for gesture recognition [28]. The system was developed to recognize the words of American Sign Language (ASL). The text output obtained from sensor-based system is converted into speech by using the popular speech synthesis technique of hidden Markov model (HMM). The HMM-based-text-to-speech synthesizer (HTS) was attached to the system for converting the text obtained from hand gestures of people into speech. The HTS system involved training phase for extraction of spectral and excitation parameters from the collected speech data and was modeled by context-dependent HMMs. The synthesis phase of HTS system was used for the construction of HMM sequence by concatenating context-dependent HMMs. Similarly, Arif et al. used five flex sensors on a glove to translate ASL gestures for Deaf-mute into the visual and audio output on LCD [29].

2.2. Vision-Based Technology Approach. Many vision-based technology interventions are used to recognize the sign languages of Deaf people. For example, Soltani et al. developed a gesture-based game for Deaf-mutes by using Microsoft Kinect which recognizes the gesture command and converts it into text so that they can enjoy the interactive environment [7]. Voice for the mute (VOM) system was developed to take input in the form of fingerspelling and convert into corresponding speech [30]. The images of fingerspelling signs are retrieved from the camera. After performing noise removal and image processing, the fingerspelling signs are matched from the trained dataset. Processed signs are linked to appropriate text and convert this text into required speech. Nagori and Malode [31] proposed the communication platform by extracting images from the video and converting these images into corresponding speech. Sood and Mishra [32] presented the system that takes images of sign language as input and displays speech as output. The features used in vision-based approaches for speech processing are also used in different object recognition based applications [33–39].

2.3. Smartphone-Based Technology Approach. Smartphone technology plays a vital role in helping the people with impairments to get themselves interacted socially and to overcome their communication barriers. Smartphone technology approach is more portable and effective as compared to sensor or vision technology. Many of the new smartphones are furnished with advanced sensors, high processors, and high-resolution cameras [40]. A real-time emergency assistant "iHelp" [41] was proposed for Deaf-mute people where they can report any kind of emergency situation. The current location of the user is accessed through built-in GPS system in a smartphone. The information about the emergency situation is sent to the management through SMS and then passed on to the closest suitable rescue units, and hence the user can get rescue through the use of iHelp. MonoVoix [42] is an Android application that also acts as a sign language interpreter. It captures the signs from a mobile phone camera and then converts them into corresponding speech. Ear Hear [43] is an Android application for Deaf-mute people. It uses sign language to communicate with normal people. The speech-to-sign and sign-to-speech technology are used. For a hearing person to interact with Deaf-mute, the text-to-speech (TTS) technology inputs the speech signal, and a corresponding sign language video is played against that input through which the mute can easily understand. Bragg et al. [44] proposed a sound detector. The app is used to detect the red alert sounds and alert the deaf-mute person by vibrating and showing a popup notification.

3. Proposed Methodology

Nowadays many technology devices such as smartphone-enabled devices prefer speech interfaces over visual ones. The research [49] highlighted that off-the-shelf speech recognition system cannot be used to detect the speech of deaf or hearing loss people as these systems contain a higher ratio of word error rate. This research recommended using human-based computations to recognize the deaf speech and using text-to-speech functionality for speech generation. In this regard, we proposed and developed an Android based application named as vocalizer to mute (V2M). The proposed application acts as an interpreter and encourages two-way communication between Deaf-mute and normal person. We refer to normal person as the one who has no hearing or vocal impairment or disability. The main features of the proposed application are listed below.

3.1. Normal to Deaf-Mute Person Communication. This module takes text or spoken message of a normal person as an input and outputs a 3D avatar that performs sign language for a Deaf-mute person. ASL based animations of an avatar

are stored in a central database of application. Each animation file is given 2–5 tags. The steps of normal to Deaf-mute person communication are as follows:

(1) The application takes text/speech of normal person as an input.

(2) The application converts the speech message of a normal person into text by using the Google Cloud Speech Application Program Interface (API) as this API detects normal speech better compared to Deaf persons' speech.

(3) The application matches the text to any of the tags associated with an animation file and displays the avatar performing corresponding sign for Deaf-mute.

3.2. Deaf-Mute to Normal Person Communication. Not everyone has knowledge of sign language so the proposed application uses disrupted speech of a Deaf-mute person. This disrupted form of speech is converted into recognizable speech format by using speech recognition system. HMM-based speech recognition is a growing technology as evidenced by the rapidly increasing commercial deployment. The performance of HMM-based speech recognition has already reached a level that can support viable applications [50]. For this purpose, HTK [51] is used for developing speech recognition system as this toolkit is primarily designed for building HMM-based speech recognition systems.

3.2.1. Speech Recognition System Using HTK. ASR system is implemented by using HTK version 3.4.1. The speech recognition process in HTK follows four steps to obtain the recognized speech of Deaf-mute. The steps are training corpus preparation, feature extraction, acoustic model generation, and recognition as illustrated in Figure 1.

(a) Training Corpus Preparation. The training corpus consists of recordings of speech samples obtained from Deaf-mute in .wav format. The corpus contains spoken English alphabets (A–Z), English digits (0 to 9), and 15 common sentences used in daily routine life, i.e., good morning, hello, good luck, thank you, etc. The utterance of one participant is separated from the others due to the variance in speech clarity among Deaf-mute people. The training utterances of each participant are labeled to simple text file (.lab). This file is used in acoustic model generation phase of the system.

(b) Acoustic Analysis. The purpose of the acoustic analysis is to convert the speech sample (.wav) into a format which is suitable for the recognition process. The proposed application used MFCC approach for acoustic analysis. MFCC is the feature extraction technique in speech recognition [52]. Main advantages of using MFCC are (1) low complexity and (2) better performance with high accuracy in recognition [53]. The overall working of MFCC is illustrated in Figure 2 [19].

The features of each step of MFCC are listed below.

(1) Pre-Emphasis. The first step of MFCC feature extraction is done by passing the speech signal through a filter. The pre-emphasis filter is the first-order high-pass filter. It is responsible for boosting the higher frequencies of a speech signal.

$$x'(n) = x(n) - \alpha x(n-1) \quad 0.9 \le \alpha \le 1.0, \tag{1}$$

where α represents the pre-emphasis coefficient, $x(n)$ is the input speech signal, and $x'(n)$ is the output speech signal with a high-pass filter applied to the input. Pre-emphasis is important because the components of speech with high frequency have small amplitude w.r.t components of speech with low frequency [54]. The silent intervals are also removed in this step by using the logarithmic technique for separating and segmenting speech from noisy background environments [55].

(2) Framing. Framing process is used to split the pre-emphasized speech signal into short segments. The voice signal is represented by N frame samples and the interframe distance or frameshift is M ($M < N$). In the proposed application, the frame sample size (N) = 256 and frameshift (M) = 100. The frame size and frameshift (in milliseconds) are calculated as

$$\begin{aligned} \text{FrameSize (ms)} = f_n = \frac{1}{N * M} = 25.6\,\text{ms}, \\ \text{Frame\_Shift} = 10\,\text{ms}. \end{aligned} \tag{2}$$

(3) Windowing. The speech signal is a nonstationary signal but it is stationary for a very short period of time. The window function is used to analyze the speech signal and extract the stationary portion of a signal. There are two types of windowing:

(i) Rectangular window,

(ii) Hamming window.

Rectangular window cuts the signal abruptly so the proposed application used Hamming window. Hamming window shrinks the values towards zero at the boundaries of the speech signal. The value of Hamming window ($w(n)$) is calculated as

$$w(n) = \begin{cases} 0.54 - 0.46 * \cos\left(\dfrac{2\pi n}{N-1}\right) & 0 \le n \le N-1 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The windowing at time n is calculated by

$$y_t(n) = w(n) * s(n). \tag{4}$$

(4) Discrete Fourier Transform (DFT). The most efficient approach for computing Discrete Fourier Transform is to use Fast Fourier Transform algorithm as it reduces the computation complexity from $\Theta(n^2)$ to $\Theta(n \log n)$. It converts the N discrete samples of speech from the time domain to the frequency domain as calculated by

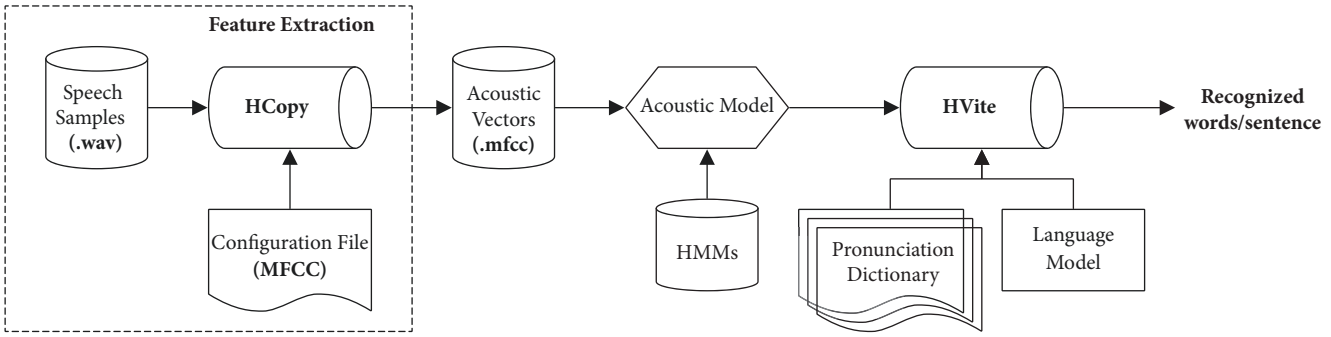

Figure 1: Speech recognition process using MFCC and HTK.

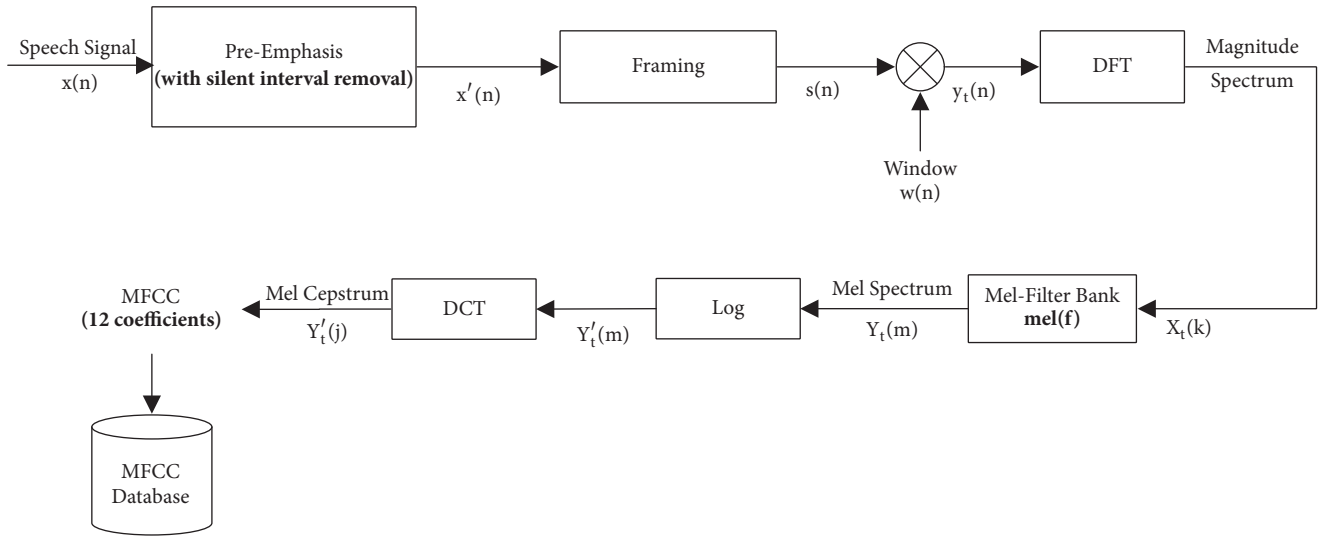

Figure 2: Block diagram of MFCC feature extraction technique.

$$X_t(K) = \sum_{n=1}^{N} s(n) * w(n) * e^{-j2\pi(k/N)n} \quad 1 \le K \le k = \sum_{n=1}^{N} y_t(n) * e^{-j2\pi(k/N)n}, \tag{5}$$

where $X_t(K)$ is the Fourier transform of $y_t(n)$ and k is the length of the DFT.

(5) Mel-Filter Bank Processing. Human ears act as band-pass filters; i.e., they focus on only certain frequency bands and have less sensitivity at higher frequencies (roughly >1000 Hz). A unit of pitch (mel) is defined for separating the perceptually equidistant pair of sounds in pitch into an equal number of mels [56] and it is calculated as

$$Y_t(m) = \text{mel}(f) = 2595 * \log_{10}\left(1 + \frac{f}{700}\right). \tag{6}$$

(6) Log. This step takes the logarithm of each of the mel-spectrum values. As human ear has less sensitivity to the slight difference in amplitude at higher amplitudes as compared to lower amplitudes. Logarithm function makes the frequency estimates less sensitive to the slight difference in input.

(7) Discrete Cosine Transform (DCT). It converts the frequency domain (log mel-spectrum) back to the time domain by using DCT. The result of the conversion is known as mel frequency cepstrum coefficient (MFCC) [57]. We calculated the mel frequency cepstrum by

$$Y'_t(j) = \sum_{m=1}^{M} \log\left(\left|Y_t(m)\right|\right) \cos\left(j(m - 0.5)\frac{\pi}{M}\right), \quad k = 1, \ldots, J. \tag{7}$$

In the proposed methodology, the value of $J = 12$ because a 12-dimensional feature parameter is sufficient to represent the voice feature of a frame [17]. The extraction of cepstrum via DCT results in 12 cepstral coefficients for each frame. These set of coefficients are called acoustic vectors (.mfcc). The acoustic vector (.mfcc) files are used for both the training and

Table 2: Details of a configuration file (config.txt).

Description	Parameters
Input Source File Format ($x(n)$)	SOURCEFORMAT = WAV
Output of Speech Sample	TARGETKIND = MFCC_0
Pre-emphasis Coefficient (α)	PREEMCOEF = 0.97
Frameshift (M)	TARGETRATE = 100000
Window Size	WINDOWSIZE = 250000
Using Hamming Window ($w(n)$)	USEHAMMING = T
No. of Filter Bank Channels (f)	NUMCHANS = 26
No. of the Cepstral Coefficients	NUMCEPS = 12
Save the Output File Compressed	SAVECOMPRESSED = T

testing speech samples. The HTK-HCopy runs for conversion of input speech sample into acoustic vectors. The configuration parameters, used for MFCC feature extraction of the speech sample, are listed in Table 2.

(c) Acoustic Model Generation. It provides a reference acoustic model with which the comparisons are made to recognize the testing utterances. A prototype is used for the initialization of first HMM. This prototype is generated for each word of the Deaf-mute dictionary. The HMM topology comprises 6 active states (observation functions) and two nonemitting states (the initial and the last state with no observation function) which are used for all the HMMs. Single Gaussian observation functions with diagonal matrices are used as observation functions and are described by a mean vector and variance vector in a text description file known as prototype. This predefined prototype file along with acoustic vectors (.mfcc) of training data and associated labels (.lab) is used by the HTK tool HInit for initialization of HMM.

(d) Recognition Phase. HTK provides a Viterbi word recognizer called HVite, and it is used to transcript the sequence of acoustic vectors into a sequence of words. HVite uses the Viterbi algorithm in finding the acoustic vectors as per MFCC model. The testing speech samples are also prepared in the same way of preparing the training corpus. In the testing phase, the speech sample is converted into series of acoustic vectors (.mfcc) using the HTK-HCopy tool. These input acoustic vectors along with HMM list, Deaf-mute pronunciation dictionary, and language model (text labels) are taken as an input by HVite to generate the recognized words.

3.3. Messaging Service for Deaf-Mute and Normal Person. The application also provides messaging feature to both Deaf-mute and normal people. A person can choose between the American Sign Language or English keyboard for sending the messages. The complete flowchart of "V2M" is illustrated in Figure 3.

4. Experimental Results and Discussions

4.1. Experimental Setup. The proposed application V2M required a camera, a mobile phone for the installation of the V2M app, laptop (acting as a server), and an instructor to guide the Deaf-mute student. The complete scenario is shown in Figure 4.

A total of 15 students from Al-Mudassir Special Education Complex Baharwal, Pakistan, participated in this experiment and the participated students were between the ages of 7 and 13 with some speech training in school. The instructor guided all students in using the mobile application. The experiment consisted of two phases.

4.1.1. Speech Testing Phase. In this phase, instructor selected a "register voice" option from a menu of the app and entered a word/sentence or question (label) in a text field of the "register sample" dialog box, for which the training speech samples of participants were taken (see Figure 5(b)). At first, the instructor needed sign language for asking the participants to speak a word/sentence or an answer. The system took 2 to 4 voice samples of each word/sentence. Whenever the participant registered his/her voice, the system acknowledged by a visual support (as in Figure 5(c)). For testing, the researcher asked questions via the V2M app, and it displayed an avatar that performed sign language for a Deaf-mute participant in order to understand the questions (see Figure 5(d)). In response, the participant selected the microphone icon (as shown in Figure 5(e)) to speak his/her answer. The app processed and compared the recorded speech sample with the registered samples. After the comparison, it returned the text and spoke out the answer of the participant (see Figure 5(f)).

4.1.2. Message Activity Phase. The participants took minimal support from an instructor in this phase. They easily composed and sent the messages by selecting sign language keyboard (see Figure 5(g)).

4.2. Qualitative Feedback. Researchers formalized questionnaire survey to evaluate the effectiveness of the Deaf-mute application. The survey comprised 12 questions for participants to answer and the reason for this short-length selection of questions was not to overwhelm Deaf-mute students with longer interviews. Secondly, these students had no experience of using any Deaf-mute based application. The qualitative feedback is summarized into following categories (paraphrased from the feedback forms).

Familiarity with Existing Mobile Apps. All participants have not heard or used any mobile applications which are dedicated to Deaf-mutes.

Ease of Use and Enjoyment. All participants enjoyed using the app. They liked the idea of using an avatar for performing sign language. Out of 15 students, 12 students performed the given tasks quite easily and 3 students have not used or interacted with mobile devices before. Initially, they found this app difficult but it became easier for them after app functions were performed 2-3 times in front of them. Overall they found this app user-friendly and interactive.

Application Interface. Participants liked the interface of the app. They learned the steps of app quite fast and they also

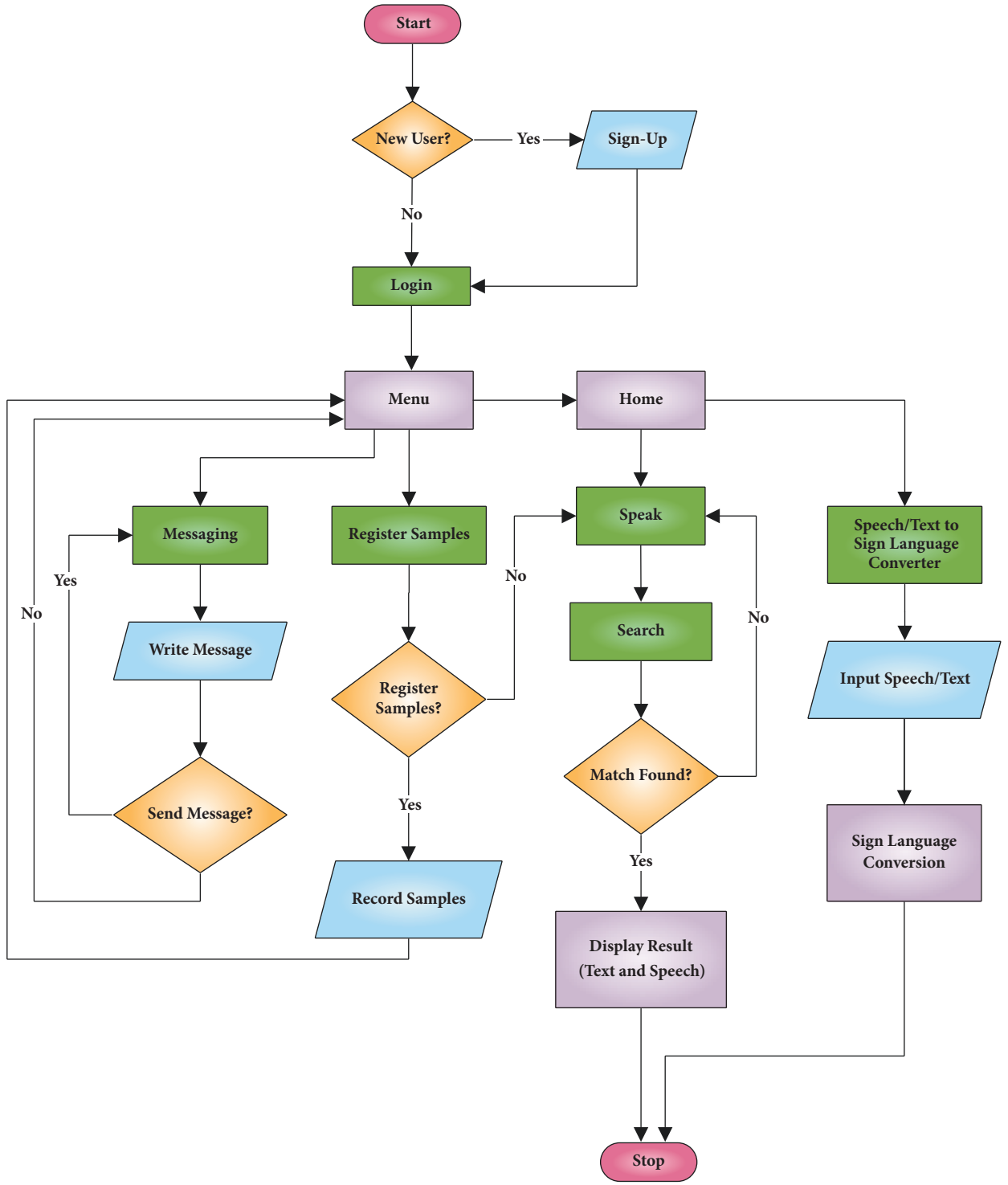

Figure 3: Flowchart of V2M application.

liked the idea of an avatar performing greeting gesture at home screen.

Source of Communication. All participants were using sign language as a primary source of communication. They recommended the intervention of mobile application as a source of communication for them. They acknowledged that the mobile app can be used to convey the message of deaf-mute to a normal person.

4.3. Results and Comparative Analysis. The application training and testing corpus are obtained from the speech samples

Figure 4: Experimental setup: a participant performing registration of speech sample task.

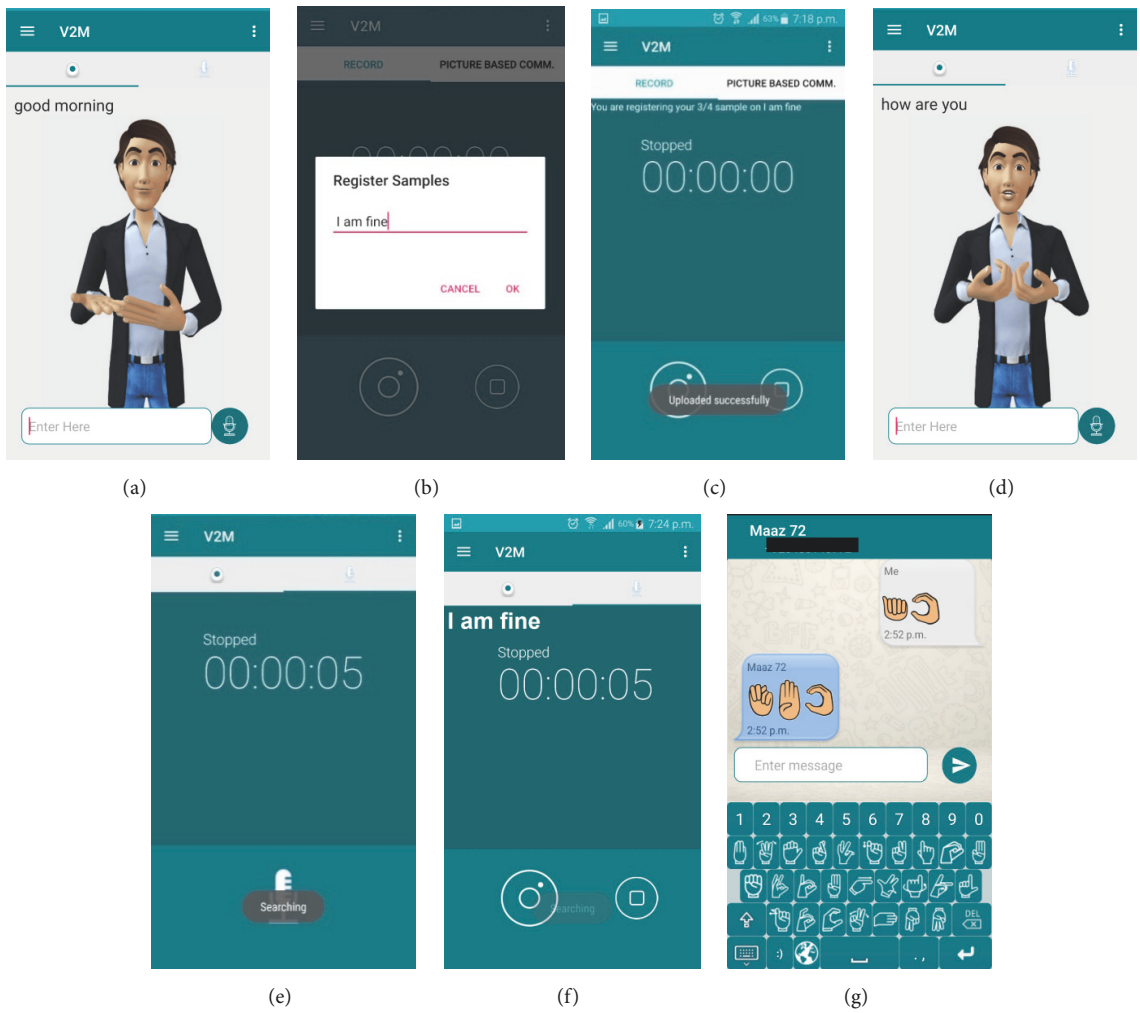

Figure 5: The working of V2M. (a) Avatar greets deaf-mute person. (b) Instructor registers text sample to ask participant for speaking it. (c) Participant recorded his/her speech samples. (d) Avatar asks a question to the Deaf-mute person. (e) Participant recorded his/her answer and app is processing the speech signal. (f) V2M displays and speaks the answer after matching the speech signal. (g) Sign language-based message service.

of Deaf-mutes. Training corpus is comprised of English alphabets (A–Z), English digits (0 to 9), and 15 common sentences used in daily routine life, i.e., good morning, hello, good luck, thank you, etc. All participants uttered each alphabet, digit, and statement 2–4 times. The total training utterances are 2440. The HTK speech recognizer was used in training process and speech recognition. HMM was used at the backend of speech recognizer HTK. For testing, each participant was asked 10 questions to answer. There are a total of 390 testing utterances. The application recorded the answer

(speech sample), processed it, and displayed (text/speech) result for normal person understanding. The accuracy of simulation results of proposed application is calculated by using precision and recall. For the V2M app, the precision is calculated by a fraction of correctly identified speech signals to a total number of speech samples whereas recall is a percentage of the number of relevant results. Precision, recall, and accuracy are calculated by using the following formulas:

$$\begin{aligned}
\text{precision} &= \frac{\text{true positive (tp)}}{\text{true positive (tp)} + \text{false positive (fp)}} \\
\text{recall} &= \frac{\text{true positive (tp)}}{\text{true positive (tp)} + \text{false negative (fn)}} \\
\text{accuracy} &= \frac{\text{true positive (tp)} + \text{true negative (tn)}}{\text{true positive (tp)} + \text{true negative (tn)} + \text{false positive (fp)} + \text{false negative (fn)}}.
\end{aligned} \tag{8}$$

True positive (tp) refers to words that are uttered by the person and detected by the system.

False positive (fp) refers to words not uttered by the person but detected by the system.

False negative (fn) refers to words that are uttered by the person but the system does not detect it.

True negative (tn) refers to everything else.

The experimental results of the proposed methodology in terms of precision, recall, and accuracy parameters are illustrated in Table 3.

It is observed from Table 3 that the number of speech samples has direct impact on precision and recall of the application. Overall average precision is 56.79% and recall is 46.79% when registered sample count in all statements is 2 ($N = 2$) for each participant. However, the average precision is 93.16% and recall is 83.19% for registered sample count 3 ($N = 3$). The average accuracy in terms of precision and recall is above 97% when registered sample count in all statements is 4 ($N = 4$) for each participant. The $F1$-score of best precision and recall is calculated:

$$\begin{aligned}
F1\ \text{score}_{(N=4)} &= 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \\
&= 2 * \frac{0.9861 * 0.979}{0.9861 + 0.979} = 0.98.
\end{aligned} \tag{9}$$

Hence it is deducted that the precision of application decreases by taking the limited number of speech samples ($N \leq 2$) of the deaf-mute. The application outperforms when the number of speech samples for each statement is greater than 2 ($2 < N \leq 2$). The speech recognition methodology of proposed application is compared with other speech recognition systems as shown in Table 4.

5. Conclusion

Deaf people face many irritations and frustrations that limit their ability to do everyday tasks. Deaf children have high rates of behavioral and emotional issues in relation to different methods of communication. The main inspiration behind the proposed application is to remove the communication barrier for Deaf-mutes especially children. This app uses the speech or text input of normal person and translates it into sign language via 3D avatar. It provides speech recognition system for the distorted speech of Deaf-mutes. The speech recognition system uses MFCC feature extraction technique to extract the acoustic vectors from speech samples. The HTK toolkit is used to convert these acoustic vectors into recognizable words or sentences by using pronunciation dictionary and language model. The application is able to recognize Deaf-mute speech samples of English alphabets (A–Z), English digits (0 to 9), and 15 common sentences used in daily routine life, i.e., good morning, hello, good luck, thank you, etc. It provides message service for both Deaf-mutes and normal people. Deaf-mutes can use customized sign language keyboard for composing the message. The app also can convert the received sign language message to text for a normal person. The proposed application was also tested on 15 children aged between 7 and 13 years. The accuracy of proposed application is 97.9%. The qualitative feedback of children also highlighted that it is easy for Deaf-mutes to adapt the mobile technology and mobile app can be used to convey their message to a normal person.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors are thankful to Maaz Khalid and Mustabshira Zia for their valuable assistance in this study. The authors also gratefully acknowledge Al-Mudassir Special Education Complex Baharwal, Pakistan, for providing them a platform to test the proposed technique of this article. The authors are appreciative of the hard work and dedication of instructors and children who participated in this study. This work was financially supported by the Machine Learning Research

Table 3: Precision and recall of proposed application with speech samples (N = 2, 3, and 4).

Testing Statement	Speech Samples N = 2			Speech Samples N = 3			Speech Samples N = 4		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy
S:1	37.5%	30%	20%	91.6%	78.5%	73.3%	100%	93.3%	93.3%
S:2	62.5%	41.6%	33.3%	85.7%	92.4%	80%	100%	100%	100%
S:3	66.6%	30.7%	26.67%	100%	80%	80%	100%	93.3%	93.3%
S:4	60%	54.54%	40%	100%	86.6%	86.67%	92.8%	100%	100%
S:5	80%	61%	53.3%	92.8%	86.6%	86.67%	100%	100%	100%
S:6	57.1%	33.3%	26.67%	100%	73.3%	73.3%	100%	93.3%	93.3%
S:7	53.8%	77.7%	46.67%	84.6%	84.6%	73.3%	100%	100%	100%
S:8	45.45%	55.5%	33.3%	100%	80%	80%	100%	100%	100%
S:9	30%	37.5%	20%	100%	86.7%	86.67%	100%	100%	100%
S:10	75%	46.1%	40%	76.9%	83.3%	66.67%	93.3%	100%	100%
Average	*56.79%*	*46.79%*	*46.67%*	*93.16%*	*83.19%*	*78.67%*	*98.61%*	*97.9%*	*97.9%*

Table 4: Comparison of proposed methodology with state-of-the-art ASR systems.

ASR Systems	Methodology	Accuracy	
Proposed Methodology (N = 4)	MFCC + HTK (8-state HMM)	97.9%	
MSIAC (Liu et al., 2017) [15]	MFCC + GMM	Experiment 1	89.50%
		Experiment 2	85.92%
TAMEEM V1.0 (Abushariah, 2017) [45]	MFCC + Sphinx 3	92.36%	
Speaker Identification System (Leu and Lin, 2017) [46]	MFCC + GMM	Experiment 1	94.08%
		Experiment 2	84.88%
Telugu Speech Signals (Mannepalli et al., 2016) [47]	MFCC – GMM	92%	
AMAZIGH LANGUAGE (Elouahabi et al., 2016) [48]	MFCC + HTK (6-state HMM)	80%	

Group, Prince Sultan University, Riyadh, Saudi Arabia [RG-CCIS-2017-06-02]. The authors are grateful for this financial support and the equipment provided to make this research successful.

References

[1] W. Tin, Z. Lin, -. Swe, and N. K. Mya, "Deaf mute or Deaf," *Asian Journal of Medical and Biological Research*, vol. 3, no. 1, pp. 10–19, 2017.

[2] I. W. Leigh and J. F. Andrews, *Deaf People and Society: Psychological, Sociological and Educational Perspectives*, Psychology Press, 2016.

[3] WHO, "Deafness and Hearing Loss," *Fact Sheet*, vol. 300, 2017.

[4] J. G. Clark, "Uses and abuses of hearing loss classification," *Asha*, vol. 23, no. 7, pp. 493–500, 1981.

[5] B. H. Jacobson, A. Johnson, C. Grywalski et al., "The Voice Handicap Index (VHI): development and validation," *American Journal of Speech-Language Pathology*, vol. 6, no. 3, pp. 66–70, 1997.

[6] P. Vostanis, M. Hayes, M. Du Feu, and J. Warren, "Detection of behavioural and emotional problems in deaf children and adolescents: Comparison of two rating scales," *Child: Care, Health and Development*, vol. 23, no. 3, pp. 233–246, 1997.

[7] F. Soltani, F. Eskandari, and S. Golestan, "Developing a gesture-based game for deaf/mute people Using microsoft kinect," in *Proceedings of the 2012 6th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012*, pp. 491–495, July 2012.

[8] D. G. Blazer, S. Domnitz, C. T. Liverman, C. on Accessible, A. H. H. C. for Adults, E. National Academies of Sciences, and Medicine, Hearing Loss: Extent, Impact, and Research Needs. 2016.

[9] WHO, "Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017, in Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report," Tech. Rep., 2017.

[10] T. Humphries, "Of Deaf-mutes, the Strange," in *Deaf World: A Historical Reader and Primary Sourcebook*, 2001.

[11] List of Sign Languages, n.d.

[12] J. Gugenheimer, K. Plaumann, F. Schaub et al., "The impact of assistive technology on communication quality between deaf and hearing individuals," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2017*, pp. 669–682, March 2017.

[13] A. Piquard-Kipffer, O. Mella, J. Miranda, D. Jouvet, and L. Orosanu, "Qualitative investigation of the display of speech recognition results for communication with deaf people," in *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, pp. 36–41, Dresden, Germany, September 2015.

[14] B. Chigier, "Automatic speech recognition," Google Patents, 1997.

[15] J.-C. Liu, F.-Y. Leu, G.-L. Lin, and H. Susanto, "An MFCC-based text-independent speaker identification system for access control," *Concurrency Computation*, vol. 30, no. 2, Article ID e4255, 2018.

[16] L.-S. Lee, J. Glass, H.-Y. Lee, and C.-A. Chan, "Spoken Content Retrieval - Beyond Cascading Speech Recognition with Text Retrieval," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 9, pp. 1389–1420, 2015.

[17] R. Vergin, D. O'Shaughnessy, and A. Farhat, "Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 7, no. 5, pp. 525–532, 1999.

[18] M. Holmberg, D. Gelbart, and W. Hemmert, "Automatic speech recognition with an adaptation model motivated by auditory processing," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 43–49, 2006.

[19] F. S. Al-Anzi and D. AbuZeina, "The Capacity of Mel Frequency Cepstral Coefficients for Speech Recognition," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 11, no. 10, pp. 1094–1098, 2017.

[20] J. L. Pray and I. K. Jordan, "The deaf community and culture at a crossroads: Issues and challenges," *Journal of Social Work in Disability and Rehabilitation*, vol. 9, no. 2, pp. 168–193, 2010.

[21] S. Barnett, "Communication with deaf and hard-of-hearing people: A guide for medical education," *Academic Medicine: Journal of the Association of American Medical Colleges*, vol. 77, no. 7, pp. 694–700, 2002.

[22] N. Glickman, "Cultural Identity, Deafness, and Mental Health," *Journal of Rehabilitation of the Deaf*, vol. 20, no. 2, pp. 1–10, 1986.

[23] R. E. Perkins-Dock, T. R. Battle, J. M. Edgerton, and J. N. McNeill, "A Survey of Barriers to Employment for Individuals Who Are Deaf," *Journal of the American Deafness & Rehabilitation Association (JADARA)*, vol. 49, no. 2, 2015.

[24] L. Sirch, L. Salvador, and A. Palese, "Communication difficulties experienced by deaf male patients during their in-hospital stay: findings from a qualitative descriptive study," *Scandinavian Journal of Caring Sciences*, vol. 31, no. 2, pp. 368–377, 2017.

[25] A. Kuenburg, P. Fellinger, and J. Fellinger, "Health Care Access Among Deaf People," *Journal of Deaf Studies and Deaf Education*, vol. 21, no. 1, pp. 1–10, 2016.

[26] M. V. Sharma, N. V. Kumar, S. C. Masaguppi, S. Mn, and D. R. Ambika, "Virtual talk for deaf, mute, blind and normal humans," in *Proceedings of the 2013 1st Texas Instruments India Educators' Conference, TIIEC 2013*, pp. 316–320, April 2013.

[27] K. Rojanasaroch and T. Laohapensaeng, "Communication aid device for illness deaf-mute," in *Proceedings of the 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2015*, June 2015.

[28] P. Vijayalakshmi and M. Aarthi, "Sign language to speech conversion," in *Proceedings of the 2016 International Conference on Recent Trends in Information Technology, ICRTIT 2016*, April 2016.

[29] A. Arif, S. T. H. Rizvi, I. Jawaid, M. A. Waleed, and M. R. Shakeel, "Techno-talk: An American Sign Language (ASL) Translator," in *Proceedings of the 3rd International Conference on Control, Decision and Information Technologies, CoDIT 2016*, pp. 665–670, April 2016.

[30] A. K. Tripathy, D. Jadhav, S. A. Barreto, D. Rasquinha, and S. S. Mathew, "Voice for the mute," in *Proceedings of the 2015 International Conference on Technologies for Sustainable Development, ICTSD 2015*, February 2015.

[31] N. P. Nagori and V. Malode, "Communication Interface for Deaf-Mute People using Microsoft Kinect," in *Proceedings of the 1st International Conference on Automatic Control and Dynamic Optimization Techniques, ICACDOT 2016*, pp. 640–644, September 2016.

[32] A. Sood and A. Mishra, "AAWAAZ: A communication system for deaf and dumb," in *Proceedings of the 5th International Conference on Reliability, Infocom Technologies and Optimization, ICRITO 2016*, pp. 620–624, September 2016.

[33] M. Yousuf, Z. Mehmood, H. A. Habib et al., "A Novel Technique Based on Visual Words Fusion Analysis of Sparse Features for Effective Content-Based Image Retrieval," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2134395, 13 pages, 2018.

[34] Z. Mehmood, S. M. Anwar, and M. Altaf, "A novel image retrieval based on rectangular spatial histograms of visual words," *Kuwait Journal of Science*, vol. 45, no. 1, pp. 54–69, 2018.

[35] Z. Mehmood, T. Mahmood, and M. A. Javid, "Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine," *Applied Intelligence*, vol. 48, no. 1, pp. 166–181, 2018.

[36] N. Ali, K. B. Bajwa, R. Sablatnig, and Z. Mehmood, "Image retrieval by addition of spatial information based on histograms of triangular regions," *Computers & Electrical Engineering*, vol. 54, pp. 539–550, 2016.

[37] Z. Mehmood, S. M. Anwar, N. Ali, H. A. Habib, and M. Rashid, "A Novel image retrieval based on a combination of local and global histograms of visual words," *Mathematical Problems in Engineering*, vol. 2016, Article ID 8217250, 12 pages, 2016.

[38] Z. Mehmood, F. Abbas, T. Mahmood, M. A. Javid, A. Rehman, and T. Nawaz, "Content-based image retrieval based on visual words fusion versus features fusion of local and global features," *Arabian Journal for Science and Engineering*, pp. 1–20, 2018.

[39] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, and M. T. Mahmood, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model," *PLoS ONE*, vol. 13, no. 4, e0194526, 2018.

[40] S. Ghanem, C. Conly, and V. Athitsos, "A Survey on Sign Language Recognition Using Smartphones," in *Proceedings of the the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 171–176, ACM, June 2017.

[41] L.-B. Chen, C.-W. Tsai, W.-J. Chang, Y.-M. Cheng, and K. S.-M. Li, "A real-time mobile emergency assistance system for helping deaf-mute people/elderly singletons," in *Proceedings of the IEEE International Conference on Consumer Electronics, ICCE 2016*, pp. 45-46, January 2016.

[42] R. Kamat, A. Danoji, A. Dhage, P. Puranik, and S. Sengupta, "MonVoix-An Android Application for the acoustically challenged people," *Journal of Communications Technology, Electronics and Computer Science*, vol. 8, pp. 24–28, 2016.

[43] G. Subhaashini, S. Divya, S. DivyaSuganya, and T. Vimal, "Ear Hear Android Application for Specially Abled Deaf People," *International Journal of Computer Science and Engineering*, vol. 3, no. 3, pp. 1108–1114, 2015.

[44] D. Bragg, N. Huynh, and R. E. Ladner, "A personalizable mobile sound detector app design for deaf and hard-of-hearing

users," in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2016*, pp. 3–13, October 2016.

[45] M. A. M. Abushariah, "TAMEEM V1.0: speakers and text independent Arabic automatic continuous speech recognizer," *International Journal of Speech Technology*, vol. 20, no. 2, pp. 261–280, 2017.

[46] F. Leu and G. Lin, "An MFCC-Based Speaker Identification System," in *Proceedings of the 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 1055–1062, March 2017.

[47] K. Mannepalli, P. N. Sastry, and M. Suman, "MFCC-GMM based accent recognition system for Telugu speech signals," *International Journal of Speech Technology*, vol. 19, no. 1, pp. 87–93, 2016.

[48] S. Elouahabi, M. Atounti, and M. Bellouki, "Amazigh Isolated-Word speech recognition system using Hidden Markov Model toolkit (HTK)," in *Proceedings of the International Conference on Information Technology for Organizations Development, IT4OD 2016*, April 2016.

[49] J. P. Bigham, R. Kushalnagar, T. K. Huang, J. P. Flores, and S. Savage, "On How Deaf People Might Use Speech to Control Devices," in *Proceedings of the the 19th International ACM SIGACCESS Conference*, pp. 383-384, October 2017.

[50] M. Gales and S. Young, "The application of hidden Markov Models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304, 2008.

[51] S. J. Young and S. Young, *The HTK hidden Markov model toolkit: Design and philosophy*, University of Cambridge, Department of Engineering, 1993.

[52] M. Anusuya and S. K. Katti, "Speech recognition by machine, a review," https://arxiv.org/abs/1001.2267.

[53] D. Gupta, P. Bansal, and K. Choudhary, "The State of the Art of Feature Extraction Techniques in Speech Recognition," in *Speech and Language Processing for Human-Machine Communications*, pp. 195–207, Springer, 2018.

[54] W. Han, C.-F. Chan, C.-S. Choy, and K.-P. Pun, "An efficient MFCC extraction method in speech recognition," in *Proceedings of the ISCAS 2006: 2006 IEEE International Symposium on Circuits and Systems*, pp. 145–148, May 2006.

[55] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction," in *Proceedings of the 4th International Conference on Signal Processing and Communication Systems, ICSPCS'2010*, December 2010.

[56] S. Memon, M. Lech, and L. He, "Using information theoretic vector quantization for inverted MFCC based speaker verification," in *Proceedings of the 2009 2nd International Conference on Computer, Control and Communication, IC4 2009*, February 2009.

[57] M. R. Hasan, M. M. G. Jamil, Rabbani., and M. S. Rahman, "Speaker identification using mel frequency cepstral coefficients," *Variations*, vol. 1, no. 4, 2004.