Cognitive Computation (2019) 11:489–502
https://doi.org/10.1007/s12559-018-9607-4

Cognitively Inspired Feature Extraction and Speech Recognition for Automated Hearing Loss Testing

Shibli Nisar[1] · Muhammad Tariq[1,2] · Ahsan Adeel[3,4,5] · Mandar Gogate[3] · Amir Hussain[6,7]

Received: 5 April 2018 / Accepted: 23 October 2018 / Published online: 13 February 2019

Abstract
Hearing loss, a partial or total inability to hear, is one of the most commonly reported disabilities. A hearing test can be carried out by an audiologist to assess a patient's auditory system. However, the procedure requires an appointment, which can result in delays and practitioner fees. In addition, there are often challenges associated with the unavailability of equipment and qualified practitioners, particularly in remote areas. This paper presents a novel idea that automatically identifies any hearing impairment based on a cognitively inspired feature extraction and speech recognition approach. The proposed system uses an adaptive filter bank with weighted Mel-frequency cepstral coefficients for feature extraction. The adaptive filter bank implementation is inspired by the principle of spectrum sensing in cognitive radio that is aware of its environment and adapts to statistical variations in the input stimuli by learning from the environment. Comparative performance evaluation demonstrates the potential of our automated hearing test method to achieve comparable results to the clinical ground truth, established by the expert audiologist's tests. The overall absolute error of the proposed model when compared with the expert audiologist test is less than 4.9 dB and 4.4 dB for the pure tone and speech audiometry tests, respectively. The overall accuracy achieved is 96.67% with a hidden Markov model (HMM). The proposed method potentially offers a second opinion to audiologists, and serves as a cost-effective pre-screening test to predict hearing loss at an early stage. In future work, authors intend to explore the application of advanced deep learning and optimization approaches to further enhance the performance of the automated testing prototype considering imperfect datasets with real-world background noise.

Keywords Hearing loss · Speech recognition · Machine learning · Automation · Cognitive radio

Introduction

Hearing loss is one of the most commonly reported disabilities in the world. According to the World Health Organization (WHO), approximately 360 to 538 million people worldwide (i.e., 5% of the total world population) are suffering from hearing loss [1]. Hearing impairment is a hidden disability with no painful symptoms. Identification and diagnoses of hearing impairment at an early stage can help reduce its negative consequences including headaches, muscle tension, increased stress, insecurity, sadness, social isolation, and depression [2]. Hearing impairment is categorized into six different classes depending on the level of hearing loss, as shown in Table 1 [1, 3, 4].

In the literature, a wide range of approaches have been proposed to identify and diagnose hearing impairment such as pure-tone testing [5], speech testing [6], middle ear testing [7], auditory brainstem response (ABR) [8], and otoacoustic emissions (OAEs). However, pure tone and speech audiometry are the most widely used approaches by audiologists [9–11]. In pure tone audiometry (PTA), hearing is measured over a range of pure tones in each ear. Frequencies vary at octave intervals from low pitches (125 Hz)

✉ Ahsan Adeel
ahsan.adeel@deepci.org

1 National University of Computer and Emerging Sciences, H-11/4, Islamabad, Pakistan

2 Princeton University, Princeton, NJ 085447, USA

3 University of Stirling, Stirling FK9 4LA, UK

4 deepCI, Edinburgh, EH16 5XW, UK

5 School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton, UK

6 Edinburgh Napier University, School of Computing, Edinburgh EH10 5DT, UK

7 Taibah Valley, Taibah University, Madinah, Saudi Arabia

Table 1 Hearing impairment categories

Hearing impairment category	Better ear hearing level (dB HL)	Hearing in a quiet environment	Hearing in a noisy environment
Normal	0–20	Does not have problems unless sound is near poorer hearing ear	May have real difficulty taking part in a conversation
Mild	21–40	Does not have problems hearing what is said	May have real difficulty taking part in a conversation
Moderate	41–55	May have difficulty hearing a normal voice	Has difficulty hearing and taking part in conversation
Moderately severe	56–70	Can hear loud speech	Has great difficulty hearing and taking part in conversation
Severe	71–90	Can hear loud speech directly in one's ear	Has very great difficulty hearing and taking part in conversation
Profound	90+	Can not hear any speech	Can not hear any speech

to high pitches (8,000 Hz). In [12], a PTA was for both ears (if PTA $<$ 20 dB, then the overall hearing is considered within the normal limits and with PTA above 90 dB, hearing loss is considered in the profound range). PTA test is typically administered in two ways; air conduction and bone conduction [13, 14]. Air conduction audiometry involves the use of headphones, whereas the bone conduction threshold is carried out using a vibrating device on a person's skull. The type of hearing loss can be found by comparing the results of both air conduction and bone conduction audiometry (e.g., conductive, sensorineural, or mixed hearing loss) [15–17].

However, PTA-based hearing assessment provides only a partial picture of the patient's auditory status. In order to confirm the PTA results and to measure the patient's ability to recognize speech stimuli, it is necessary to conduct speech audiometry. Speech recognition threshold (SRT) is one of the commonly used measures of speech audiometry [18–20]. Surveys conducted in the USA estimated that 99.5% [21] and 83% [22] of audiologists used SRT as part of their audiological assessment process. In conjunction with PTA, it helped in determining the degree and type of hearing loss. Speech audiometry also provides information regarding discomfort or tolerance to speech stimuli and information on word recognition abilities [23, 24]. In SRT, the lowest threshold is defined as such that a patient recognizes speech stimuli with 50% accuracy.

In recent literature, researchers have proposed several automated hearing loss testing methods [25–29]. The main objective of the mentioned literature was to accurately diagnose the hearing impairment by minimizing the absolute error rate and maximizing the accuracy. However, the approach is restricted to air conduction audiometry, and thus complete assessment of a patient is not possible without access to other testing modalities such as bone conduction and speech audiometry. It is to be noted that most of the aforementioned automated methods suffer from problems such as ambient noise, inaccurate results at low frequencies, difficulty in distinguishing conductive and sensorineural hearing losses, low reliability due to absence of speech audiometry, etc. The proposed automated hearing loss testing approach presented in this paper has addressed the aforementioned limitations by conducting tests in a noise proof cabin (minimizing the effect of ambient noise), carrying out both pure tone and speech audiometry (for reliability and accuracy), utilizing speech recognition for both audiometry tests, and carrying out both air and bone conduction tests (identifying both types of hearing losses i.e., conductive and sensorineural hearing loss). To the best of the authors' knowledge, this is the first effort to automate pure tone and speech audiometry based on speech recognition.

The rest of the paper is organized as follows: "Proposed Methodology" discusses the proposed model. "Materials" presents materials and experimental setup. "Classifier" covers the classifiers used in this work. "Results" discusses results obtained from the proposed model and comparison with state-of-the-art techniques. "Discussion" covers the detailed discussion. "Conclusion" presents some remarks and future directions.

Proposed Methodology

The block diagram of the proposed model is shown in Fig. 1. For pure tone testing, the subject is asked to respond to the heard tones. Similarly, for SRT test, the user is asked to repeat spondee words uttered by the machine. User response is captured through the speech signal, and the system identifies right and wrong guesses uttered by the user to calculate SRT of a subject. For speech recognition, the proposed system uses an adaptive filter bank with weighted Mel-frequency cepstral coefficients (MFCCs)-based feature

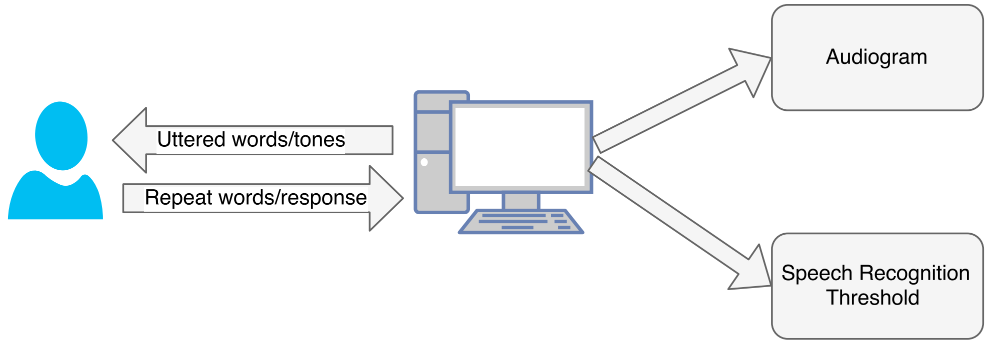

Fig. 1 Schematic diagram of proposed model

extraction method. Similarly, for classification, different machine learning algorithms are used such as support vector machine(SVM) [30], *k*-nearest neighbor (*k*-NN) [31], ensemble classifier [32], and hidden Markov model (HMM) [33]. More details are comprehensively presented in the subsequent sections.

Pure Tone Test

Pure tone audiometry performed by the audiologist manually, usually follows the modified Hughson-Westlake procedure [34]. This procedure is listed in ANSI S3.21-1978 (R-1992) standard [35]. The procedure starts with a signal that is easily audible by the subject. If the subject responds to that signal, the intensity is reduced by a fixed step until it is not audible to the subject. The intensity is increased by a fixed size until the subject responds to it. At this point, whenever the subject responds, the intensity is decreased by a fixed step. Otherwise, the intensity is increased by a fixed step.

The intensity at which a listener responds two to three times, is recorded as the threshold. This study has automated the modified Hughson-Westlake procedure. The complete algorithm of proposed pure tone audiometry based on speech recognition is listed in Algorithm 1. Algorithm 1 explains the workings of an automated pure tone audiometry. The upper amplitude threshold is set to 120 dB and the initial starting amplitude of tone signal is set as 45 dB.

Speech Test

In speech audiometry, an audiologist plays or utters familiar words to the patient and then observes the response of the subject through speech or some visual cues. In the proposed automated system, the subject listens to words uttered by the machine, and then repeats the same words. If the score is greater than 50% i.e., more than three words are correctly repeated out of six, the intensity is reduced by a fixed step until the score is reported as less than or equal to 50%. The intensity is increased by a small step until the listener's score is greater than 50%. Following this, whenever the score is greater than 50%, the level will be decremented, and whenever the score is less than 50%, the speech level will be incremented. When the score is equal to 50% i.e., three out of six words are correctly repeated, then the counter will be incremented by unity. When the counter reaches 3, the threshold is recorded for SRT as the median of three values that incremented the counter. The complete algorithm of the proposed speech recognition threshold based on speech recognition is listed in Algorithm 2 that shows the workings of automated speech audiometry.

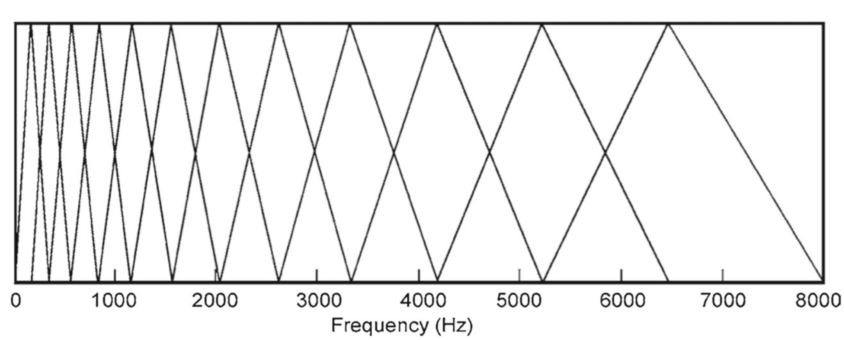

Fig. 2 Static filter bank

The upper amplitude threshold is set to 120 dB and the initial starting amplitude of tone signal is set as 45 dB.

Algorithm 1 Complete algorithm for pure tone audiometry

```
Require: Frequency = [1000 500 250 125 2000 4000 8000], status[3]=0, j=0, upper limit = 120 dB, initial amplitude = 45 dB
 1: procedure
 2:     for i = 1 to 7 do
 3:         if tone is heard then
 4:             while tone is heard do
 5:                 Decrease amplitude by 15 dB
 6:             end while
 7:             while tone is not heard do
 8:                 Increase amplitude by 5 dB
 9:             end while
10:             if tone is heard then
11:                 if j==2 then
12:                     j=0;
13:                     status[j++]=1
14:                 else
15:                     status[j++]=1
16:                 end if
17:                 status[0]+status[1]+status[2]>=2
18:                 Mark threshold & go to step 2
19:             else
20:                 Decrease tone by 10 dB
21:             end if
22:             if tone is heard then
23:                 go to step 10
24:             else
25:                 go to step 6
26:             end if
27:         else
28:             while tone is not heard && threshold < do
                upper limit
29:                 Increase amplitude by 20 dB
30:             end while
31:             if  tone is heard then
32:                 go to step 4
33:             else threshold >= upper limit
34:                 No threshold found
35:             end if
36:         end if

37:     end for
38: end procedure
39: return Thresholds ∀ frequencies
```

Algorithm 2 Complete algorithm for speech audiometry

```
Require: Recorded spondee words, counter=0, status[3]=0, upper limit = 120 dB, A=45 dB
 1: procedure
 2:     Play(spondee, A)
 3:     if score > 50 then
 4:         while score > 50 do
 5:             Decrease amplitude by 15 dB
 6:             A=A-15
 7:         end while
 8:         while score =< 50 do
 9:             Increase amplitude by 5 dB
10:             A=A+5
11:         end while
12:         if score = 50 then
13:             counter++
14:             status[j++]=A
15:             if counter ==3 then
16:                 SRT = median(status[3])
17:                 break END TEST
18:             else
19:                 play(spondee,A)
20:                 go to step 12
21:             end if
22:         else
23:         end if
24:         if  score ¿ 50 then
25:             Decrease tone by 10 dB,
26:             A=A-10
27:             play(spondee,A)
28:             go to step 12
29:         else
30:             Increase tone by 5 dB
31:             A=A+5
32:             play(spondee,A)
33:             go to step 12
34:         end if
35:     else
36:         while s docore =< 50 && threshold < upper
            limit do
37:             Increase amplitude by 20 dB
38:         end while
39:         if  score > 50 then
40:             go to step 4
41:         else threshold >= upper limit
42:             No threshold found
43:         end if
44:     end if
45: end procedure
46: return Speech Recognition Threshold
```

Feature Extraction

The two main modules of any speech recognition method are feature extraction and classification. In the literature, researchers have proposed several different feature extraction techniques. For example, authors in [36] proposed an enhanced feature extraction process for automatic speech recognition with fractal dimensions to address the limitation of MFCCs-based feature extraction for automatic speech recognition (ASR). Similarly, authors in [37] investigated a low-variance multitaper spectrum estimation method to compute the MFCC features for robust speech and speaker recognition systems. MFCC is one of the most widely used feature extraction technique in speech recognition. In MFCC, the spectrum of speech signal is obtained after taking the Fourier transform and the spectrum power is approximated and scaled to the response of the human ear. In traditional MFCC, a static Mel-filter bank is used to get energies at each filter as shown in Fig. 2. Since spectrum of speech varies with utterance of different words, therefore, static Mel-filter bank is not a suitable choice. Our proposed system uses an adaptive filter bank with weighted MFCCs for feature extraction (inspired by the principle of spectrum sensing in cognitive radio), where feature extraction first senses the spectrum in order to design the adaptive filter bank of relevant frequency bands [38, 39]. The proposed adaptive filter bank with weighted MFCCs improves the speech recognition results as compared to the state-of-the-art static MFCC. The block diagram of the proposed model is shown in Fig. 3.

Following are the steps to calculate adaptive filter bank with weighted MFCC (AWMFCC) coefficients. The short time Fourier transform (STFT) of the signal x[n] is given by

$$x[k] = \sum_{n=0}^{N-1} x(n)w(k-n)exp\left(\frac{2\pi j}{N}kn\right), \tag{1}$$

with N as length of the frame, and the Hamming window

$$w[n] = 0.54 - 0.46\cos\left(2\pi\frac{n}{N-1}\right). \tag{2}$$

The power spectrum is

$$s[k] = (real(X[k]))^2 + (imag(X[k]))^2. \tag{3}$$

In AWMFCC, the Mel-filter bank constitutes after sensing the spectrum of input speech. The first step involves spectrum sensing that determines the orientation of the signal on the spectrum using normalized power spectral density $\hat{f}$. Expectation and standard deviation of normalized power spectral density ˆf are computed using Eqs. 4 and 5, respectively.

$$\mu = \sum_{i}^{N} \hat{f}_i \cdot A_i, \tag{4}$$

$$\sigma = \sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}\left(\hat{f}_i - \mu\right)^2}. \tag{5}$$

It is assumed that the spectrum of speech is Gaussian distribution. With this assumption, 99.7% of the signal lies within $\mu \pm 3\sigma$. Hence, the Mel-filter bank constitutes in such a way to cover the spectrum from $\mu \pm 3\sigma$ only.

The Mel spectrum is given by

$$m[l] = \sum_{k=0}^{N/2} S[k]mel]k], \tag{6}$$

with *l=0,...,L-1*. where L is the number of Mel-weighted filters.

Mel filter is a series of L bandpass filters designed to simulate the bandpass filtering

$$Mel(f) = 2595\log_{10}\left(1 + \frac{f}{700}\right). \tag{7}$$

Discrete cosine transform (DCT) is applied to the natural logarithm of the Mel spectrum to obtain an adaptive filter bank coefficients. The derivative and double derivative of adaptive MFCC produce delta and double delta features that improve the overall accuracy of the speech recognition system. However, this approach increases the dimension of the feature vector leading to higher computational complexity overheads. In order to improve the accuracy while keeping the complexity intact, weighted

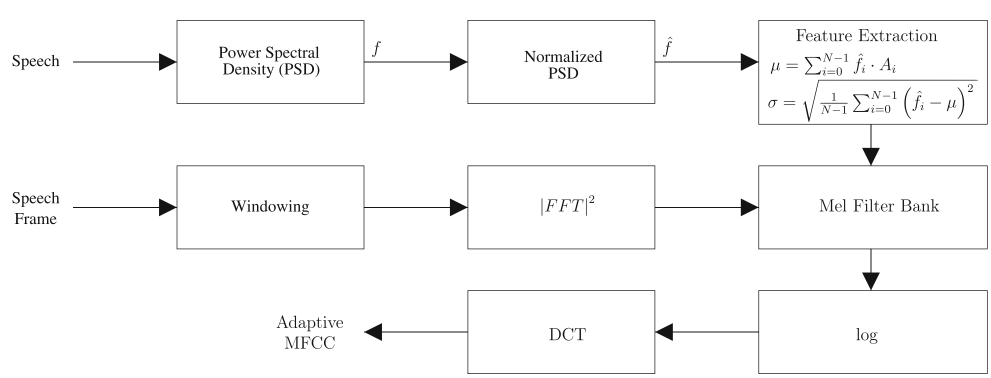

Fig. 3 Block diagram of adaptive filter bank in MFCC

mel frequency cepstral coefficients *wc(n)* are used. *wc(n)* is defined as:

$$wc(n) = c(n) + a\Delta c(n) + b\Delta\Delta c(n), \tag{8}$$

where a and b are weights assigned to delta and double delta features, respectively. *c(n)* is the adaptive MFCC coefficients. Since these derivative features contribute slightly less than *c(n)*, the weights are constrained to be $b<a<1$. The final feature vector *wc(n)* is 13-dimensional thus reducing the complexity overhead at recognition stage. The entire process of adaptive filter bank with weighted MFCC is listed in Algorithm 3.

Algorithm 3 Complete algorithm for feature extraction

Require: signal x, sampling frequency (16KHz), linear filter (10), log spacing (100), log filters (12), cepstral coefficients (20), window size (400), FFT points (512), liftering (22).

```
procedure
    Framing x̂ := Extraction(x)
    Fourier Transform ẑ := fft(ŷ)
    Windowing ŷ := Hamming(x̂)                          (2)
    PSD f := periodogram(x)
    Normalized PSD f̂ := f/sum(f)
    μ := Expectation of f̂                               (4)
    σ := Standard Deviation of f̂                        (5)
    Construction-filter-bank := Mel-bank(μ, σ)           (7)
    Logrithmic p̂ := log(ẑ)
    Inverse Fourier MFCC‾ := Discrete Cosine Transform(p̂)
    Assign weights to MFCC‾
    Return MFCC‾
end procedure
```

Ensure: Adaptive Weighted Mel Frequency Cepstral Coefficients

Materials

Instructions

Pure tone test is an unusual test. The subject encounters nothing similar in their daily lives. Therefore, good instructions are essential to achieve best results. It is recommended to start the hearing test with the listener's better ear [40, 41]. If the listener notices no difference between right and left ears, then the right ear should be the starting ear by default [41]. In order to make the procedure fully automatic, the machine will ask the patient; which ear you listen better, right, left, or I don't know? The machine will start the test from the ear specified by the patient. If option three is uttered then the test will start from the right ear. Before starting the PTA test, the patient will be guided to utter "YES" if the tone is heard no matter how soft it is. On the contrary, if no tone is heard by the subject then "NO" must be uttered. Same procedure is repeated for both air and bone conduction. The instructions given below are recorded in the machine and before starting the test, the subject has to listen to the complete instruction set:

Do you hear better out of one ear than the other ? (If yes, speak out which ear. If no, then start with the right ear). In your <better or right> ear, you will hear some faint tone and then in your <other or left> ear, you will hear a chain of tones (of same frequency) and then silence. Listen to the tone and when you hear it <uttered "YES">. The tone will generally get fainter and fainter each time they are presented. <Speak out "YES"> whenever you think you hear the tone. The pitch of the tones will change, first going lower in pitch and then going higher in pitch. The test of your <other or left> ear will not begin until your <better or right> ear has been tested for all of the frequencies. If you are certain that you hear the tone <speak out "YES">, for as long as you hear the tone. If no tone is heard <uttered "NO">. A simple <"YES" when a tone is heard> and <"NO" when no tone is heard>. Same procedure is repeated for both air and bone conduction. Please don't remove the earphones until <"test is over"> is played by the machine.

Setup for Experiments

Reliable test requires a noise proof environment and proper equipment. In order to achieve noise proof environment, complying the *ISO 8253-1:2010* standards, a cabin is established in National University Systems and Simulation Lab (NUSyS). All tests were conducted in the noise proof cabin. Pure tones of different frequencies and level were generated using MATLAB R2016b with chirp command. Equipment used to carry out the proposed method is listed in Table 2.

Since no free standard corpus of spondee words are is available online, the first step was to develop the spondee words corpus to build, train and test the proposed model. The following steps were taken to develop a standard corpus. The recordings were taken using the Sony PCM-M10. The speech atmosphere and surroundings were kept

Table 2 Equipment used

Air conduction	TDH 39 [42]
Bone conduction	AS600IG Trekz titanium [43]
Machine	Hp EliteBook core i7
Recorder	Sony PCM-M10 [44]
MATLAB	R2016b

in mind, in order to ensure noise free surroundings. In case of any disturbance or unexpected noise, the recorder was paused and the respected dataset was recorded again. The recorder was set at a sampling rate of 44.1 KHz. All the recordings were saved in the .WAV format. A total of 50 students aging from 18 to 35 years were asked to utter the spondee words (72 letters) given in Table 3 for the creation of corpus. All participants signed the informed consent form. The spondee corpus constitutes total 3600 utterances. Similarly, each student was asked to utter "Yes" and "No" twice. Both the corpses were used to train and test the classifiers on 70% and 30% of the corpus data respectively.

Participants

Sixty individuals voluntarily participated in this study, including 40 males and 20 females, aged 18–70 years. All participants signed the informed consent form. Patients were selected to cover the hearing impairment of all categories with both conductive and sensorineural hearing loss. The conventional pure tone and speech audiometry examination was carried out first. Following these tests (performed by a trained audiologist), subjects were sent to the noise proof cabin and the proposed automated approach was used to identify and diagnose hearing impairments.

Comparison Metric Between Conventional and Proposed Audiometry Test

The results obtained from the proposed hearing test based on speech recognition were compared with the results generated by the expert audiologist. The degree of agreement between the proposed and conventional audiograms was calculated in terms of mean absolute and average error for each frequency as well as for the overall system. The absolute error and average error are defined as:

$$\Delta dB = |\text{Actual}_{audiologist} - \text{Approximate}_{proposed}|, \quad (9)$$

$$\text{Average} = \frac{\sum \Delta dB}{\text{number of } \Delta\text{dB}}. \quad (10)$$

Table 3 Spondee words uttered by participants

Letters
sidewalk, birthday, cupcake, airplane, playground, baseball
railroad, baseball, playground, cowboy, cupcake, sunset
sunset, hot-dog, outside, scarecrow, jump rope, backyard
rainbow, toothbrush, ice-cream, doorbell, ice cream, airplane
schoolroom, backyard, jump rope, bedroom, playground, sunset
highchair, sunshine, football, blue jay, cowboy, outside
sidewalk, birthday, cupcake, hot-dog, bedroom, sunshine
schoolroom, rainbow, outside, doorbell, sunshine, football
rainbow, jackknife, cowboy, hairbrush, doorbell, schoolroom
sunset, schoolroom, football, cupcake, playground, rainbow
sunset, hot-dog, football, cupcake, outside, sidewalk
rainbow, jackknife, cowboy, hairbrush, doorbell, schoolroom

Classifier

The classification in the proposed method is in fact the matching of features extracted from the tested words and the features saved in the database in the training phase. In this work, four classifiers are studied i.e., SVM, KNN, AdaBoost, and HMM. In machine learning, SVM is widely used for feature matching and classification [45]. The principle of SVM is to maximize the functional margin between the nearest training data of a distinct class and construct an optimal hyper plane [46]. KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions) [47]. KNN has been used in statistical estimation and pattern recognition in the beginning of 1970s as a non-parametric technique. The distance function used in this work is Euclidean distance [48]. Recently, ensemble classifier is widely used in different domains of machine learning. Two most popular methods for constructing an ensemble are Bagging [49] and Boosting [50]. The first practical version of Boosting is Adaptive Boosting (AdaBoost) [51, 52]. It is a popular ensemble algorithm that improves the simple boosting algorithm by using an iterative process. The main concept behind this algorithm is to concentrate more on those patterns, which are difficult to classify. The amount of focus is quantified by weights that are assigned to every pattern in training set. In each iteration the weights of correctly classified instances are decreased while the weights of misclassified instances are increased. As a result, the weak learner is forced to focus on the instances in the training set that are difficult to classify by performing additional iterations and creating more classifiers. Weights are assigned to each classifier depending upon the overall accuracy of the classifiers. Higher weights are assigned to more account classifiers, which help in the classification of new patterns. Dietterich [53], provides a detailed informal reasoning, from computational, statistical,

and representational viewpoint. One of the most powerful statistical tools used in speech is HMM. HMM is widely used because of its simple networks that generate a sequence of vectors using number of states [33, 54]. Modeling the short-term spectra associated with each state usually is a mixture of Gaussian distribution. Parameters of the

Table 4 Reliability analysis of a conventional audiogram versus proposed automated audiogram for each frequency and ear

Left		Right	
Frequency (Hz)	Δ dB	Frequency (Hz)	Δ dB
125	4.3	125	3.5
250	4	250	3.8
500	5.1	500	4.9
1000	5.2	1000	5.1
2000	5.4	2000	5.3
4000	4.5	4000	5.2
8000	5.5	8000	5
Avg Δ dB = 4.86		Avg Δ dB = 4.69	

(ΔdB = | proposed - audiologist |)

model are usually the state transition probabilities, variance, means, etc [55].

Results

To test the proposed system, 10 patients from each hearing impairment category as shown in Table 1 were requested to participate in the testing process. A total of 60 patients were included in this study. Findings of the validity analysis of the results for each frequency is presented in Table 4. When assessing the mean difference between the pure tone audiogram and proposed hearing test results, the accuracy is very satisfactory and achieved the mean absolute difference (Δ dB) of less than 5.6 dB for each frequency as shown in Table 4. The overall average error is found to be less than 4.9 dB. Similarly, the comparison of SRT obtained by the expert audiologist with that of the proposed hearing tests were very satisfactory and achieved a mean absolute difference of less than 4.4 dB.

Reliability analysis of a conventional audiogram versus the proposed audiogram for each frequency and ear of a subject is compared. Results are summarized in Table 5 and shown in Fig. 4 for the air conduction test. Since the difference in hearing level is more than 10 dB between ears, the result concluded asymmetric hearing loss. Similarly, reliability analysis of a conventional audiogram versus the proposed audiogram for each frequency and ear of the subject are compared. Results are summarized in Table 6 and shown in Fig. 5 for both air and bone conduction tests. The gap between air and bone conduction (AB gap) is less than 10 dB. Hence, it is concluded that the type of hearing loss is sensorineural.

The proposed method was tested with 10 patients, each with different hearing impairment categories. Table 7 summaries the accuracy achieved by different classifiers i.e., SVM, KNN, AdaBoost, and HMM. The table clearly shows that HMM outperformed SVM, KNN, and AdaBoost classifiers by achieving highest accuracy rate. It can be noted that the classification accuracy is almost as high as it can be therefore, any margin for improvement has been reduced. However, in future, we intend to consider more realistic real-life challenging scenarios including imperfect datasets with real-world background noise, where HMM and deep learning are expected to preform significantly better. Table 8 summarizes the accuracy achieved by the proposed method with HMM classifier. The confusion matrix shows correct identification of hearing loss. It is found that the overall accuracy achieved from the proposed method is 96.67%.

Case Study

In order to integrate the information and interpret the outcome of the proposed model, one complete case study is discussed in this section. The subject audiogram generated from the proposed model is shown in Fig. 6, which includes both air and bone conduction test. The x(left ear) and

Table 5 Reliability analysis of a conventional audiogram versus proposed automated audiogram for each frequency and ear

Proposed results			Audiologist results		
Frequency (Hz)	Right Ear (dB)	Left Ear (dB)	Frequency (Hz)	Right Ear (dB)	Left Ear (dB)
125	20	40	125	15	35
250	10	50	250	10	50
500	20	45	500	25	40
1000	10	45	1000	15	45
2000	15	50	2000	15	50
4000	15	55	4000	20	50
8000	25	55	8000	20	50
	Normal	Moderate		Normal	Moderate
Asymmetric hearing loss			Asymmetric hearing loss		

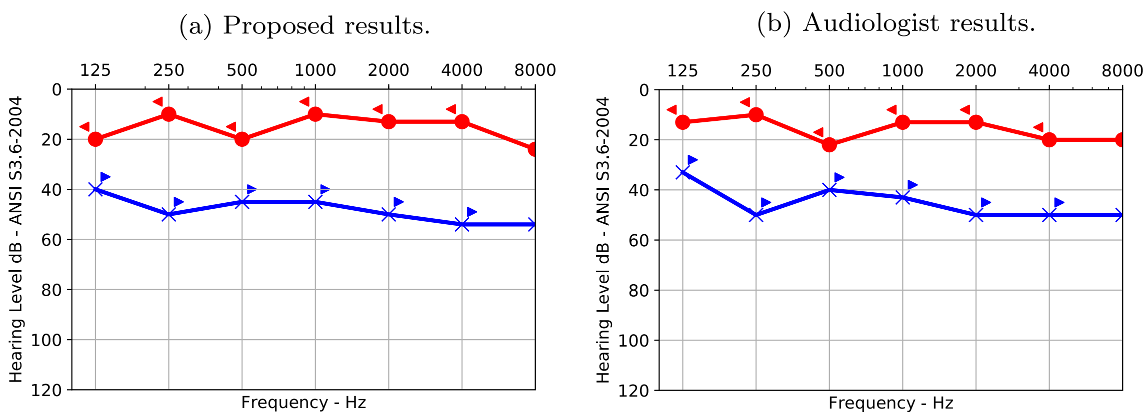

Fig. 4 Proposed and expert audiologist air conduction audiogram

Table 6 Reliability analysis of a conventional audiogram versus proposed automated audiogram for each frequency and ear for bone conduction

Proposed results			Audiologist results		
Frequency (Hz)	Right Ear (< dB)	Left Ear (> dB)	Frequency (Hz)	Right Ear (< dB)	Left Ear (> dB)
125	10	30	125	10	30
250	10	45	250	10	45
500	15	40	500	25	35
1000	10	45	1000	10	40
2000	10	50	2000	10	40
4000	10	50	4000	15	45
Sensorineural hearing loss			Sensorineural hearing loss		

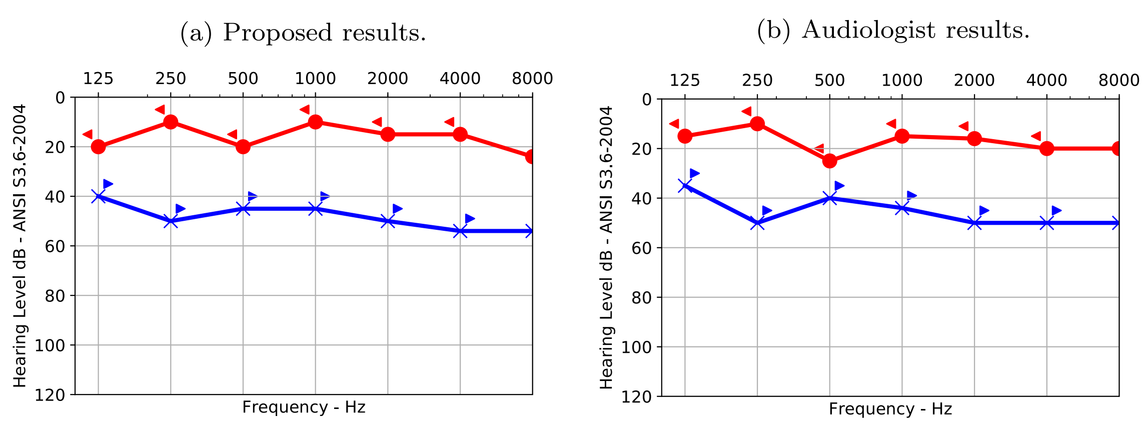

Fig. 5 Proposed and expert audiologist audiogram

Table 7 The % accuracy achieved with different classifiers

Classifier	Accuracy
KNN	78.33
SVM	81.67
AdaBoost	95
HMM	96.67

o(right ear) indicate the softest levels that an individual can hear (thresholds) via air conduction audiometry using headphones, while < and > indicate the thresholds obtained via bone conduction audiometry for right and left ear, respectively. In this case, the gap between air and bone conduction thresholds is not significant (less than 10 dB), therefore, the model drop out the conductive loss (i.e., a conductive hearing loss would indicate that something is preventing sound from being conducted through outer and middle ear systems causing better thresholds to be obtained by bone conduction than via air conduction.). Because air and bone conduction thresholds are essentially same for all test frequencies, the model concludes that there is no blockage involves in the outer and middle ear. Hence, the hearing loss lies within the sensory system and therefore identifies sensorineural hearing loss. Table 10 summaries the air-bone conduction gap (compared with the threshold value i.e., 10 dB). In this case, the AB gap is less than 10 dB. Hence, the type of hearing loss identified by the proposed model is sensorineural. Table 10 shows the threshold of hearing for each ear and frequency generated by the proposed model and expert audiologist. The overall hearing loss for each ear is also reported in Table 10. Table 9, summaries the subject hearing sensitivity for hearing and recognizing spondees at a level of 20 dB in the right ear and 45 dB in the left ear, that is compared with the expert audiologist result. The normal range for SRT is from −10 dB to 25 dBHL; therefore, the subject SRTs are normal for right and moderate hearing loss for left ear. The model also averages the thresholds of 500 Hz, 1000 Hz, and 2000 Hz. The audiogram reveals pure tone averages (PTA) of 18.33 dB and 51.66 dB for the right and left ears, respectively. Since, average PTA value is within 10 dB of the SRT

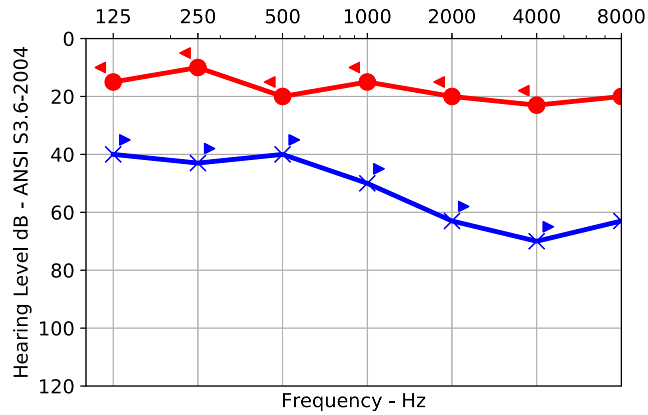

Fig. 6 Audiogram generated from proposed method

obtained for each ear, one can conclude with good certainty that the SRT and PTA findings agreed with each other. Therefore, the test appears to be reliable, otherwise the test will have to be repeated. The mean absolute difference between the SRT obtained from the proposed model and expert audiologist is less than 2.6 dB in this case (Table 10.).

Discussion

More than 5% of the world's population is suffering from hearing loss, comprising 10% children and 90% adults [1]. One third of the individuals over 65 years of age suffer from some form of hearing impairment [56]. Diagnoses of hearing loss at early stage is very important. Because proper treatment will help to improve hearing and reduce negative consequences of hearing loss such as insecurity, sadness, emotional, mental, social well-being, physical, social isolation, and depression. Preliminary hearing test is carried out by the audiologist to assess a patients' auditory system. A shortage of trained healthcare professionals and associated infrastructure and resource limitations mean that hearing health services are unavailable to the majority of the world population. Automation of hearing loss test is very important to help patient in early diagnoses and to cater the above problems faced by the patients.

Table 8 The % accuracy achieved by the proposed models with HMM classifier

	Normal	Mild	Moderate	Moderate severe	Severe	Profound	Accuracy
Normal	10	0	0	0	0	0	100
Mild	0	10	0	0	0	0	100
Moderate	0	1	9	0	0	0	90
Moderate severe	0	0	1	9	0	0	90
Severe	0	0	0	0	10	0	100
Profound	0	0	0	0	0	10	100
Overall accuracy	96.67						

Table 9 Reliability analysis of a conventional SRT versus proposed automated SRT

	PTA	$SRT_{Proposed}$	$SRT_{audiologist}$	Δ dB
RE	18.33	20	25	5
LE	51.66	45	45	0

In recent years, some promising works have been reported for automating the hearing loss test. In [25], the validity of the smartphone hearing application is assessed with a clinical audiogram. The absolute difference between the smartphone test and a pure tone audiogram was found to be less than 8.8 dB. A clinical study was carried out in [27], pointing out that ambient noise hampers the hearing thresholds of lower frequencies. It is also found that the ambient noise had an effect at both low and high frequencies. In [26], the author studied a group of patients with moderate hearing loss and demonstrated an iPod-based hearing screening test i.e., UHear. The study showed 98% sensitivity (i.e., correctly diagnosing the hearing loss), however application results overestimated by 8 dB when compared to the conventional audiometry results carried out in a sound booth. Comparison of the application test with conventional audiometry is carried out in [28]. The comparison of 325 subjects between 6 and 10 years of age was carried out, where 83% specification and 63% sensitivity is found. Foulad et al. [29], compared the self-administrated test with conventional audiometry. The tests were performed in a quiet room and found 94% of the threshold values within 10 dB of the threshold values obtained with formal audiometry in 42 subjects. The technology is restricted to air conduction audiometry, and thus comprehensive evaluation may be limited without access to other testing modalities such as bone conduction and speech audiometry. In [57], a novel method for measuring pure tone thresholds using an automated thresholds measurement is reported. Similarly this paper is restricted to air conduction audiometry only.

There are some limitations regarding the performance of the automated hearing test listed above. First, the users can drive the test at any place because of the self-administration of the test. This will introduce a lot of noise such as ambient noise. As a result, it will hamper the thresholds obtained. There may be inaccurate results, particularly for low frequencies because of ambient noise. Thus, this is less reliable than the tests performed by the audiologists. The hearing test does not distinguish between conductive and sensorineural hearing loss. Hence the existing methods are not capable of identifying the nature of hearing loss, whether it is due to the outer, middle, or inner ear.

All of the above problems are addressed in the proposed model presented in this work. All the tests are carried out in a noise proof cabin to minimize the effect of noises such as ambient noise etc. The proposed model carries out both pure tone and speech audiometry, which makes the results more reliable and accurate. Both the audiometry are carried out using speech recognition. Hence, this make the overall system more natural and pervasive. The proposed model also carries out both air and bone conduction tests, which are capable of identifying both types of hearing loss i.e., conductive and sensorineural hearing loss. In the literature the diagnosis of hearing loss is done using the air conduction test only [25, 26, 28, 29]. In this paper, in order to improve the diagnosis rate, bone conduction and speech threshold tests are carried out for the first time along with air conduction tests. Air conduction test based hearing assessments provide only a partial picture of the patient's auditory status. In order to confirm the result of air conduction and measure the patient's ability to hear, it is necessary to assess bone conduction and conduct a speech threshold test. Both of these tests are included in

Table 10 Reliability analysis of a conventional audiogram versus proposed automated audiogram for each frequency and ear

Proposed results			Audiologist results		
Frequency (Hz)	Right Ear (dB)	Left Ear (dB)	Frequency (Hz)	Right Ear (dB)	Left Ear (dB)
125	15	40	125	15	35
250	10	45	250	10	50
500	20	40	500	25	40
1000	15	50	1000	20	45
2000	20	65	2000	15	50
4000	25	70	4000	20	60
8000	20	65	8000	20	55
	Normal	Moderate		Normal	Moderate
Asymmetric hearing loss			Asymmetric hearing loss		
Sensorineural hearing loss			Sensorineural hearing loss		

Table 11 Comparison of proposed model

References	Absolute error	Tone audiometry		Speech audiometry
		Air conduction	Bone conduction	
[25]	8.8 dB	✓		
[26]	8 dB	✓		
[29]	10 dB	✓		
This work	4.9 dB	✓	✓	✓

this work by computing an automated hearing loss test. Table 11 summarized the comparison between the proposed model with the most robust models recently reported in the literature.

In order to correctly recognize the spoken speech words, a novel feature extraction method is proposed in this work. The proposed system uses an adaptive filter bank with weighted MFCC for feature extraction (inspired by the principle of spectrum sensing in cognitive radio), where the feature extraction method first senses the spectrum in order to design the adaptive filter bank at relevant frequency bands. By introducing this novel feature extraction method the absolute error rate achieved in this work is much lower than those in the state-of-the-art methods.

Conclusion

A hearing test can be carried out by the audiologist to assess the patient's auditory system. However, long delays due to appointment requirements, practitioner fees, and unavailability of required equipment/qualified practitioners (particularly in remote areas) are some of the major limitations. In this paper, a novel idea that automatically identifies the hearing impairment based on a cognitively inspired feature extraction and speech recognition is presented. To the best of our knowledge, this is the first attempt to automate pure tone and speech audiometry testing. In the proposed method, the user is asked to repeat specific words uttered by the machine. The system captures the user response through the speech signal, and identifies right and wrong guesses uttered by the user, in order to generate an audiogram and speech recognition threshold automatically. For feature extraction, the proposed system uses an adaptive filter bank with weighted MFCCs to design the adaptive filter bank of relevant frequency bands. The feature extraction method is inspired by the principle of spectrum sensing in cognitive radio and follows the idea of learning and adaptation. For classification, machine learning algorithms including SVM, KNN, AdaBoost, and HMM are employed, where HMM outperformed AdaBoost, SVM, and KNN. Comparative performance evaluation (when compared with the expert audiologist test) demonstrated the potential of our automated method achieving an overall absolute error of less than 4.9 dB and 4.4 dB for the pure tone and speech audiometry tests respectively. The overall accuracy achieved by the proposed method, in terms of correctly classifying the hearing impairment category is 96.67%. Hence, it is concluded that the proposed method potentially offers a second opinion to audiologists, and serves as a cost-effective pre-screening test to predict hearing loss at an early stage. In the future, authors intend to explore the application of advanced deep learning and optimization approaches to further enhance the performance of the automated testing prototype considering more realistic real-life challenging scenarios including imperfect datasets with real-world background noise.

Acknowledgments The authors would like to acknowledge the support of Prof. Hidayat Ullah from Khyber Teaching Hospital (ENT department) in Pakistan, for his kind support with pure tone and speech audiometry. The authors would like to acknowledge audiologists, Kamran Mulk from Rehman Medical Institute and Muhammad Saeed from Khyber Teaching Hospital for highlighting pure tone and speech audiometry-related issues. The authors would like to acknowledge the support of Dr. Muhammad Arsalan Khan from Khyber Teaching Hospital for inviting participants (patients). Lastly, the authors would like to gratefully acknowledge the support of deepCI and Taibah Valley (Taibah University, Madinah, Saudi Arabia).

Funding Information This research was supported by deepCI and Taibah Valley (Taibah University, Saudi Arabia).

Compliance with Ethical Standards This manuscript has not been published in whole or in part elsewhere, which has also not currently being considered for publication in another journal. All authors have been personally and actively involved in substantive work leading to the manuscript, and will hold themselves jointly and individually responsible for its content.

Conflict of interests The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Organization WH, et al. Millions of People in the World have Hearing Loss that can be Treated or Prevented. Awareness is the Key to Prevention. 2013.

2. Dalton DS, Cruickshanks KJ, Klein BE, Klein R, Wiley TL, Nondahl DM. The impact of hearing loss on quality of life in older adults. Gerontol. 2003;43(5):661–668.
3. Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. Health Technology Assessment-Southampton-. 2007;11(42).
4. Fagan J. Open access guide to audiology and hearing aids for otolaryngologists. 2014.
5. Association ASLH, et al. Guidelines for manual pure-tone threshold audiometry. 2005.
6. Hudgins CV, Hawkins J, Kaklin J, Stevens S. The development of recorded auditory tests for measuring hearing loss for speech. Laryngoscope. 1947;57(1):57–89.
7. Probst R, Lonsbury-Martin B, Martin G, Coats A. Otoacoustic emissions in ears with hearing loss. Amer J Otolaryngol. 1987;8(2):73–81.
8. Wilson DF, Hodgson RS, Gustafson MF. Auditory brainstem response testing. Laryngoscope. 1993;103(5):580–581.
9. Schlauch RS, Han HJ, Tzu-Ling JY, Carney E. Pure-tone–spondee threshold relationships in functional hearing loss: a test of loudness contribution. J Speech Language Hear Res. 2017;60(1):136–143.
10. Martin FN, Clark JG. Introduction to audiology. Boston: Allyn and Bacon; 1997.
11. Brandy WT. Speech audiometry. Handb Clin Audiol. 2002;5:96–110.
12. Franks JR. Hearing measurement. National Institute for Occupational Safety and Health. 2001; p. 183–232.
13. Carhart R. Clinical application of bone conduction audiometry. Arch Otolaryngol. 1950;51(6):798–808.
14. Stapells DR, Oates P. Estimation of the pure-tone audiogram by the auditory brainstem response: A review. Audiol Neurotol. 1997;2(5):257–280.
15. Loss CH. Sensorineural hearing loss. Diseases Ear Nose Throat. 2012.
16. Pensak ML, Adelman RA. Conductive hearing loss. Otolaryngology-head and neck surgery St Louis: Mosby Year Book. 1993.
17. Ramsay HA, Linthicum JF. Mixed hearing loss in otosclerosis: indication for long-term follow-up. Amer J Otol. 1994;15(4):536–539.
18. Sreedhar J, Venkatesh L, Nagaraja M, Srinivasan P. Development and evaluation of paired words for testing of speech recognition threshold in Telugu A preliminary report. J Indian Speech Lang Hear Assoc. 2011;25(2):128–136.
19. Van Tasell DJ, Yanz JL. Speech recognition threshold in noise: effects of hearing loss, frequency response, and speech materials. J Speech Lang Hear Res. 1987;30(3):377–386.
20. Association ASLH, et al. Determining threshold level for speech. 1988.
21. Martin FN, Champlin CA, Chambers JA. Seventh survey of audiometric practices in the United States. J-Amer Acad Audiol. 1998;9:95–104.
22. MD R. Audiological survey. 2000.
23. Schoepflin JR. Back to basics: speech audiometry. 2015.
24. Boothroyd A. Developments in speech audiometry. Br J Audiol. 1968;2(1):3–10.
25. Renda L, Selċuk ÖT, Eyigör H, Osma Ü, Yılmaz MD. Smartphone based audiometric test for confirming the level of hearing; Is it useable in underserved areas? J Int Adv Otol. 2016;12(1):61–6.
26. Szudek J, Ostevik A, Dziegielewski P, Robinson-Anagor J, Gomaa N, Hodgetts B, et al. Can Uhear me now? Validation of an iPod-based hearing loss screening test. Journal of Otolaryngology–Head & Neck Surgery. 2012; p. 41.
27. Wong TW, Yu T, Chen W, Chiu Y, Wong C, Wong A. Agreement between hearing thresholds measured in non-soundproof work environments and a soundproof booth. Occup Environ Med. 2003;60(9):667–671.
28. Kam ACS, Gao H, Li LKC, Zhao H, Qiu S, Tong MCF. Automated hearing screening for children: a pilot study in China. Int J Audiol. 2013;52(12):855–860.
29. Foulad A, Bui P, Djalilian H. Automated audiometry using Apple iOS-based application technology. Otolaryngol–Head Neck Surg. 2013;149(5):700–706.
30. Ananthi S, Dhanalakshmi P. SVM and HMM modeling techniques for speech recognition using LPCC and MFCC features. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Springer; 2015. p. 519–526.
31. Chen Ch. Handbook of pattern recognition and computer vision. Singapore: World Scientific; 2015.
32. Anagnostopoulos CN, Iliou T, Giannoukos I. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artif Intell Rev. 2015;43(2):155–177.
33. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77(2):257–286.
34. Carhart R, Jerger J. Preferred method for clinical determination of pure-tone thresholds. Journal of Speech & Hearing Disorders. 1959.
35. Franks JR. Hearing measurement. National Institute for Occupational Safety and Health. 2001; p. 183–232.
36. Ezeiza A, de Ipiña KL, Hernández C, Barroso N. Enhancing the feature extraction process for automatic speech recognition with fractal dimensions. Cogn Comput. 2013;5(4):545–550.
37. Alam MJ, Kenny P, O'shaughnessy D. Low-variance multitaper mel-frequency cepstral coefficient features for speech and speaker recognition systems. Cogn Comput. 2013;5(4):533–544.
38. Hei Y, Li W, Li M, Qiu Z, Fu W. Optimization of multiuser MIMO cooperative spectrum sensing in cognitive radio networks. Cogn Comput. 2015;7(3):359–368.
39. Nisar S, Khan OU, Tariq M. An efficient adaptive window size selection method for improving spectrogram visualization. Computational intelligence and neuroscience. 2016.
40. Dobie RA, Van Hemel S, Council NR, et al. Basics of Sound, the Ear, and Hearing. 2004.
41. Schoepflin JR. Back to Basics: Speech Audiometry. 2015.
42. Kapul A, Zubova E, Torgaev SN, Drobchik V, Vol. 881. Pure-tone audiometer. In: Journal of Physics: Conference Series. UK: IOP Publishing; 2017, p. 012010.
43. Behgam M, Grant SL. Echo cancellation for bone conduction transducers. In: 2014 48th Asilomar Conference on Signals, Systems and Computers. IEEE; 2014. p. 1629–1632.
44. Zhong W, Kong X, You X, Wang B. Recording Device Identification Based on Cepstral Mixed Features. 2015.
45. Hsu CW, Chang CC, Lin CJ, et al. A practical guide to support vector classification. 2003.
46. Shady Y, Zayed SHH. Speaker independent Arabic speech recognition using support vector machine. Department of Electrical Engineering, Shoubra Faculty of Engineering. Cairo: Benha University; 2009.
47. Priya TL, Raajan N, Raju N, Preethi P, Mathini S. Speech and non-speech identification and classification using KNN Algorithm. Proced Eng. 2012;38:952–958.
48. Bhatia N, et al. Survey of nearest neighbor techniques. arXiv:10070085. 2010.

49. Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123–140.
50. Freund Y, Schapire RE. Game theory, on-line prediction and boosting. In: Proceedings of the ninth annual conference on Computational learning theory. ACM; 1996. p. 325–332.
51. Freund Y, Schapire RE, et al. Experiments with a new boosting algorithm. In: icml; 1996. p. 148–156.
52. Rokach L. Ensemble-based classifiers. Artif Intell Rev. 2010;33(1):1–39.
53. Dietterich TG. Ensemble methods in machine learning. In: International workshop on multiple classifier systems. Springer; 2000. p. 1–15.
54. Vimala C, Radha V. Isolated speech recognition system for Tamil language using statistical pattern matching and machine learning techniques. J Eng Sci Technol (JESTEC). 2015;10(5):617–632.
55. Juang BH, Rabiner LR. Hidden Markov models for speech recognition. Technometrics. 1991;33(3):251–272.
56. Organization WH, et al. Deafness and hearing loss. 2015. http://www.who.int/mediacentre/factsheets/fs300/en/ (visited on 01/16/2016). 2014.
57. Eddins DA, Walton JP, Dziorny AE, Frisina RD. Comparison of pure tone thresholds obtained via automated audiometry and standard pure tone audiometry. J Acoust Soc Amer. 2012;131(4):3518–3518.