Convert Arabic Letters Voice into Gesture

Shaker K. Ali1 and Sabreen k. Saud2
1Computer sciences and mathematics college, University of Thi_Qar, Thi_Qar, Iraq.
2College of Education for Pure, University of Thi_Qar, Thi_Qar, Iraq.
1shaker@utq.edu.iq and 2 sabreenkhalid55@gmail.com

Abstract. This paper suggests an approach to solve the problem of social communication between blind and dumb by converting voices of 28 Arabic letters (ﺃ……….ﻱ) into gesture (images) by extraction features by using Mel-frequency Cepstral coefficients (MFCC) and classify the types of letters by using J48, KNN, and Naive byes (NB). Several features are extracted from speech voice of Arabic letters voices. The dataset collected by recorded voices from twenty different persons, each person recorded ten voices for each twenty eight letters so the total dataset are 5600 voices (200 voices for each 28 letters). Mel-frequency Cepstral coefficients are extracted from 5600 voices of letters which convert the voices into a signal and extract features vector to classify later by using J48, KNN and NB algorithms, which may vary in time or speed signals. The experimental results show that the best accuracy of speech recognition algorithm by using the J48 algorithm with a performance ratio of 100% while KNN is the 94.023% and Naive byes is the 20.012%.

Keywords: Mel frequency cepstral coefficients (MFCC), Vector quantization, J48, KNN and NB, K-mean.

1. Introduction

Communication is the way of exchanging thoughts, opinions, information, or messages among the people by writing, speech, or signs. Communication bridges the gap between the people. There are different ways to communicate, where communication is usually oral among people i.e. talking to each other while dumb people cannot communicate with others as ordinary people do. They cannot speak, while the people who are deaf are able to speak, but unable to hear and the blind are unable to see but they can speak and hear the voice [1]. From the perspective of biology, sound can be considered as a signal that contain many or single tone which emanate from living organisms which own sounding member utilized to communicate between the individuals or other genus which express what is wanted by an action or speech and referred to as sense which result from such vibrations hearing. Voice is considered as the base for various experiences which are obtained via the individuals and the sound speed in the normal antenna’s centre is determined at 343 meters for each second or 1224 kilo-meters.
for each hour. The sound’s speed is associated to material’s density and the hardness factor related to materials where sound moves [2]. Voice can be considered as signal related to infinite information. Speech signal’s digital processing is of high importance for high-speed in addition to the adequate automatic voice recognition technologies. Today, it is utilized for disabled individuals, telephony military, and healthcare [3]. There are 2 areas related to the voice recognition: speaker recognition and speech recognition, however the research was limited to focus on the field of speech recognition [4].

1.1 Speech recognition:
The main way of communication between individuals is speech, also it is considered as the most effective and natural approach to exchange information between humans [5]. Speech Recognition, also referred to as (computer speech recognition or automatic speech recognition) can be defined as the capability of programs or machines in identifying phrases and words from spoken languages and convert them into certain format that is readable by machines. The major aim of this method is to develop the systems and methods for inputting the speech to the machines [6]. Speech is considered as distinctive signals that carry and convey many levels related to knowledge sources, non-linguistic as well as linguistic information. Human speech is defined as complicated acoustic wave which is the result of the output coming from the effort of a speaker. Speech consist of sentences which contain words [7].

1.2 Types of Speech recognition
Speech can be defined as vocalization (speaking) related to words or single word which are representing single meaning. Utterances could be multiple sentences, single sentence, many words, or one word [10].
- Isolated Word: Typically, the isolated word recognizer demands each one of the utterances to have quiet on the two sides related to sample window. It does not indicate that it does accept single words, yet it does demand one utterance at the same time. This can be applied for conditions in which the users are demanded for providing just single word commands or responses, yet it is extremely inefficient for multiple word inputs. Also, it can be implemented easily due to the fact that the boundaries of words are understandable and words have tendency towards being plainly pronounced and that is the main benefit of such type [10].
- Connected Word also means one word received and analyzed by the system at the same time in the same way as isolated words but here the silence between the words is reduced so that they appear to be connected or as intermittent sentence [11].
- Continuous speech the sentences used in this type of speech are considered to be the most difficult to apply in speech recognition. It is difficult to define the limits of each word on the one hand and the lack of precision in the pronunciation of words when they are in a sentence [11].
- Spontaneous speech This speech type is no rehearsed and it is considered to be natural. ASR system with spontaneous speech must have the ability of handling various natural speech features like the words which are run together in addition to the minor stutters. Spontaneous (un-rehearsed) speech might consist of nonwords, mis-pronunciations, and false-starts [10].

1.3 Human Speech Production
Speech is a tool of communication. The speech production is considered as day-to-day mean in humans, also it is considered to be an uncomplicated mechanism, yet there is a high complexity in its internal approach. The first stage of speech production is breathing, since it includes 2 processes, the first one is to inhale, while the other is to exhale. Throughout the inhaling process, the air will enter the lungs, while in exhaling process, the air is going to flow out of organism. Throughout the exhaling process, the air is going to flow out through lungs, trachea, larynx, vocal folds, mouth, lips, nasal cavity, and so on. Articulatory movement is the movement of such organs. The articulatory movement control can be referred to as the motor control, also it is done via the brain [4]. The vocal cords in their normal position are open and approaching each other in case of talking according to the sound to be issued [5].
1.4 Research problem

According to statistics provided by the World Health Organization, there are 285 million people around the world suffering from blindness, one million are dumb and many suffer from other physical disabilities.[1] The problem is how to help people with physical disabilities communicate easily with one another (blind and dumb). This research focuses on this problem and tries to develop a new system that enables the blind and dumb communicate with each other by turning speech into a signal.

1.5 Literature Review

Parwinder Pal Singh, Pushpa Rani 2014 [10] this study has introduced a method for extracting the features for the speech signals related to spoken words with the use of MFCC. It is considered as non-parametric frequency domain method that operates on the basis of the human auditory perception system. Initially, all the voice samples related to the words will be taken as input and through the use of praat tool denoise all the samples. After that, the coefficients will be extracted via the use of MFCC since such coefficients are collectively representing short-term power spectrum that is related to the sound. This study effectively denoised input samples in the same time as extracting MFCC coefficients, also it took into account the Delta energy function and concluded that there is a possibility of increasing the MFCC coefficient based on the requirements. Acceleration and velocity could be added for extracting twelve MFCC coefficients. The features have been extracted on the basis of the information which has been involved in speech signal. A file of extensions (.wav) has been used to store the extracted features.

Anjali Bala, Abhijeet Kumar and et al 2010 [32] Voice is considered as a signal related to infinite information. The digital processing regarding the speech signals is of high importance for adequate and high-speed recognition technologies. Today, it is applied by disabled individuals, telephony military, and healthcare, thus, digital signal processes like feature matching as well as feature extraction are the most recent issues for studying the voice signals. For the purpose of extracting significant information from speech signals, making decisions related to processes, in addition to obtaining results, data should be manipulated and analyzed. The major approaches which have been applied to extract features related to signals is finding MFCCs, since they are considered to be collectively representing short-term power spectrum related to the sound, on the basis of the linear cosine transform that is related to log power spectrum on non-linear mel scale of frequency. This study presents MFCC for extracting features as well as DTW for comparing test patterns.

Ms. Rupali S Chavan, Dr. Ganesh. S Sable 2013 [14] Speech can be considered as a major and a natural communication form in humans. There are a lot of factors associated to speech, such as speaker identification, speech synthesis, speech recognition, speech verification, and so on. The main aim of this research is studying the system of speech recognition with the use of HMM. The main aim of speech recognition is determining the present speech on the basis of spoken information. HMM is used for pattern training, MFCC is used for feature extraction. It is indicated that MFCC is applied majorly for speech’s feature extraction since it is robust to noise, while HMM is optimum in modeling methods since it increases the speed and preciseness of recognition.

Om Prakash Prabhakar and Navneet Kumar Sahu 2013 [31] Speech is defined as the major communication mode in humans. Nowadays, the technologies of technologies are provided commercially for unlimited, yet wide range of tasks. Such technologies are enabling the machines to reliably and correctly respond to the voices of humans, and offer significant services. This study provides a summary related to the main technological perspectives related to the major development of speech recognition as well as providing overview method created in each one of the speech recognition stages. The basics related to speech recognition have been discussed, also its new developments have been examined. The AST system’s performance depends on adopted feature extraction approach and speech recognition method for specific language is put to comparison. Recently, the demand for speech recognition studies has increased. The effectiveness of HMM method
accompanied by MFCC features is more appropriate for such demands and provide optimum recognition results. The approaches are going to allow creating more powerful systems, used worldwide in the future.

K. M. Shiva Prasad*, G. N. Kodanda Ramaiah2 , et al 2017 [13]. Human speech is defined as distinctive signal that carry and convey multiple levels related to the knowledge source, non-linguistic and linguistic information. Speech signals might be considered as information bearing signals that are evolving as functions related to single independent variable such as time. Speech can be defined as complicated acoustic wave that result from the output related to the effort of speaker. Speech consists of sentences which are made of words. Generally, words are composed of phoneme sequences referred to as syllables. The speech analysis is of high importance in speech recognition and synthesis. Speech analysis is also referred to as the feature extraction. There are various approaches of speech analysis, each one of them have certain advantages and disadvantages, no single approach is defined as optimum for speech analysis or recognition. Speech analysis or speech signal front ends enable the extraction of speech features. LPC, PLP, and MFCC are majorly applied techniques of feature extraction in the speech analysis. The speech’s nonlinear nature made the selection of LPC not excellent for speech estimation, MFCC and PLP are derived from logarithmically spaced filter banks combined with human auditory system, thus they have more optimum response in comparison to LPC. Rasta PLP approach is uncomplicated computationally effective and have high robustness for studying state spectral factors.

Dr. Jaffar Alkhier 2017 [8] The speech recognition is one of the most modern technologies, which entered force in various fields of life, whether medical or security or industrial techniques. three systems have been created for speech recognition. They differ from each other in the used methods during the stage of features extraction. While the first system used MFCC algorithm, the second system used LPCC algorithm, and the third system used PLP algorithm. All these three systems used HMM as classifier.

2. Proposed System
The proposed system used Cross-Validation as statistical method, which performs in two parts; training and testing part at the same time. Where the testing part used 10% of the recorded voices and 90% for training. The training part used 5040 voices and the testing part used 560 voices. The system consist from four stages, the first stage is the recording of sounds (dataset collection) and the second stage is feature extraction using Mel Frequency Cepstrum Coefficient algorithm which contains six steps; (the first step Pre-emphasis, the second step Framing, the third step Hamming windowing, the fourth step Fast Fourier Transform, the fifth step Mel scale filter bank and the sixth step Discreet Cosine Transform). The third stage is vector quantization with the use of K-means algorithm. The fourth stage is the classification stage using the following algorithms (J48, KNN and Naive Bayes). The Figure.1 shows the flowchart of proposed system.
2.1. Dataset

Dataset collected from 20 different persons (males and females) by recording voices from 20 different persons, each person recorded 10 voices for each 28 letters so the total dataset are 5600 voices (200 voices for each 28 letters), which have been recorded in same environment. All voices signals have been acquired under various conditions, like the record time length, and sound amplitude level. The recorded voices are stored in format “.WAV” extension. Postulates to user for choosing any sample of the speech for testing from recorded dataset.

2.2. Feature Extraction

The feature extraction is the basic portion of the system of speech recognition the feature extraction is the core of the system. Its function is to extract the features from input speech (signal). Extraction features compress the amount of input signal (i.e. the vector) with no causing of damages to the speech signal power [12]. At this stage, the speech signal is converted to a sequence of characteristic (feature vector) that represent information which is stored in the spoken speech. An important characteristic of feature extraction phase is the suppression of information which does not matter in order
to properly classify such as information about the speaker and information about the transmission channel such as the telephone. Extraction features play an important role in speech recognition. It is very difficult to get data from speech signal [8].

There are several techniques that are used to do the job such as MFCC (Mel Frequency Cepstral Coefficients), RASTA filtering, LPC (Linear Predictive Coding), and PLDA (Probabilistic Linear Discriminate Analysis) [13].

2.2.1 MFCC Feature Extraction

This technique is a very common technique in extracting the distinctive features of sound in speech recognition systems due to the accuracy of its results and the ability to partially eliminate the noise of the signal [14]. As well as the speed of application and less complex and more effective under different circumstances [15]. In this technique, the human hearing process is approached, i.e., an attempt to extract the signal characteristics in a manner consistent with the human hearing mechanism, since the human ear is sensitive to frequencies that are less than 1,000 HZ and weak to frequencies higher than 1,000 HZ [11]. The reason for the design of such filters is that the human ear is not sensitive to high frequencies and therefore can reduce the number of filters characteristic of these frequencies [16]. It has been discovered that MFCC is commonly utilized for extracting speech features due to its robustness to noise [17].

MFCC is used due to the following reasons [13].

- MFCC is the most significant features that are needed between various speech application types.
- Provides highly accurate results for cleaning speech
- MFCC may be considered as standard features in speech recognition systems

![Figure 2. Mel-Frequency Cepstrum Co-efficients](image)

In the MFCC, the voice signal passes through the following stages:

Step 1: **Pre-emphasis**

In the pre-emphasis, undesirable frequencies are produced from the environment during the sound recording process. These frequencies are low frequencies that are not considered to be spoken by the person and should therefore be omitted from the signal [4]. Each value in the speech signal is reassessed using the following Equation (1), [16].

\[
y(n) = x(n) - a * x(n - 1)
\]

where \(y(n)\) is the output signal, the value of \(a\) is usually between [0.9 - 1.0] and \(x(n - 1)\) is the input signal.

Step 2: **Framing**

The speech signal is a constantly changing signal, so a framing process must be applied. The framing process is divide the signal into several sections each section called frame so that every one of the frames may be analyzed independently and in a short time rather than analyzing the whole signal. The signal cannot be handled once as this may cause unsatisfactory results [13]. The length of every one of the frames ranges between 20 and 40 with an overlap equal to half or one third of the size of the frame for easy transition from one frame to another [18].

Step 3: **Hamming windowing**
At this stage each of the above frames is multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame of the signal and eliminate interruptions at the edges and distortion in the signal [7]. The Hamming window is represented as in Equation 2.

If the window is defined as \( W(n), 0 \leq n \leq N - 1 \), where \( N \) = number of samples in each frame, \( Y[n] \) = Output signal, \( X(n) \) = input signal, \( W(n) \) = Hamming window, then the result of windowing signal as in Equation (2) [3].

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

\[
Y(n) = X(n) \times W(n) \tag{3}
\]

**Step 4: Fast Fourier Transform (FFT)**

To convert each frame of \( N \) samples from time domain into frequency domain [3]. FFT is a fast approach of Discrete Fourier Transformation (DFT), on a specific set of \( N \) samples, the FFT Equation(4) is given as

Where \( k = 0, 1, 2 \ldots \ldots N-1 \)

\[
F(u) = \sum_{k=0}^{N} f(k) e^{-j2\pi uk} \tag{4}
\]

**Step 5: Mel scale filter bank Frequencies**

It is a set of trigonometric signals, to calculate the filter banks, triangular filters are used and the frequency in Hertz \( f \) each frequency filter is given by Equation(5) [7].

\[
Mel(f) = 2595 \times \log_{10} \left( 1 + \frac{f}{700} \right) \tag{5}
\]

These filters exhibit linear behavior for low frequencies and logarithmic behavior for high frequencies. The reason for the design of such filters is that the human ear is not sensitive to high frequencies [11].

**Step 6: Discrete Cosine Transform (DCT)**

DCT is the procedure of converting Mel scale spectra into time-based domains. This process’s result is referred to as the MFCC. The group of the obtained coefficients is referred to as the acoustic vector. Which means that until this step, audio inputs are converted to a stream of the acoustic vectors that will subsequently produce the group of inputs for algorithms of classification [17]. Given by the Equation(6) [19].

\[
C_n = \sum_{k=0}^{k-1} (logDk) \cos \left[ m(k-\frac{1}{2})\pi \right] \tag{6}
\]

where \( m = 0, 1 \ldots k-1 \), \( C_n \) represents the MFCC and \( m \) is the number of the coefficients here \( m=13 \) so, total number of coefficients extracted from each frame is 13.

2.3. **Vector quantization**

Vector quantization can be defined as a conventional method of quantization from signal processing and the data compress in spatial domain. Due to the fact that it’s one of the lossy techniques, therefore maintaining the quality of the image and the ratio of compression is a complicated task. Which is why, the code-book storing image data has to be designed in the best way [20]. K-means is utilized for the optimization of code-book [21].

2.4. **Classification:**

Classification is a method to extract the data and the division of this data into classes or specific groups in advance. It is a way to learn under the supervision of training requires disaggregated data. Classification is utilized for classifying every item in a dataset to one of the specified set of classes [22]. There are many classification algorithms in this thesis. Such as J48, KNN and Naive Bayes algorithm.
2.4.1. J-48 algorithm

J-48 is a simple of C4.5 classification decision tree. It produces a binary tree. The method of the decision tree is of highest usefulness in the tasks of classification. With this approach, a tree is created for the sake of modelling the procedure of the classification. As soon as the tree has been constructed, it’s implemented on every dataset tuple and produces the result of classification for the tuple. Throughout tree construction, J-48 overlooks missing values, which means that that item’s value may be projected according to the information available on attribute values for the rest of the records. The main concept is dividing data to range according to that item’s attribute values, which can be seen in training sample. J-48 can classify through decision trees or through the rules that are produced from those trees [22]. The J48 Algorithm as following:

- Basic algorithm, the tree is produced in a recursive top-down divide-and-conquer way. First, every one of the training samples is at root. Attributes are categorical (if continuous-valued, they’re previously discredited). The samples are recursively partitioned according to the chosen attributes. The test attributes are chosen based on statistical or heuristic measures (like the information gain)
- Conditions to stop partitioning: every sample for a certain node is part of the same class. There aren’t any examples left [13].

2.4.2. K-Nearest Neighbour classifier (K-NN)

K-NN is a very significant non-parameter algorithm in the area of pattern identification, and it is one of the supervised learning predictable classification algorithms [23]. Amongst different approaches of the supervised statistical pattern identification, the rule of the Nearest Neighbour has achieved constantly higher performance, with no prior assumption on distributions from where training samples are derived. A new sample is categorized through the calculation of distance to nearest training sample [24]. KNN classifier expands this concept through taking k nearest points and assigning the majority class. For the sake of simplifying the issue it’s usually fixed to odd a number (usually 1, 3 or 5) in order to leave no chance for ties. Bigger values of k are helpful in the reduction of the noisy points’ effects in the training dataset, and k selection is usually carried out via cross-validation [25]. The following steps give K-NN algorithm:

- Let k be a positive integer
- Compute distance \( d(x, x_i) \) for each \( i=1,2,3,...,n \) where ‘d’ is Euclidean distance.
- Sort the samples based on the computed distance values.
- Select the heuristically optimum KNN based according the value of the RMSE which is performed with the method of the cross validation.
- Compute the inverse distance weighted mean with the k-nearest multi-variate neighbours.

2.4.3. Naïve Bayes classifier (NB)

The Naïve Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities. The algorithm tends to perform well and learn rapidly in various supervised classification problems, this property makes it suitable for datasets that are large [22]. The NB main idea is based on the so-called Bayesian theorem which is more suitable for inputs with high dimensionality. The model is called naive because it assumes that the attributes are conditionally independent of each other given the class. This assumption gives it the ability to compute Bayesian formula probabilities from a rather small training dataset [9]. Bayes classification is related to obtaining two probability types- Prior and posterior. The prior probability is not dependent on any data and is related only to the identification that a tuple is part of a specific class regardless any other data. Posterior probabilities are obtaining the
probability of the tuple that belongs to a particular class according to some data. It’s a conditional
probability[26]. Bayesian theorem can be represented as:

\[ P(H|T) = \frac{P(T|H)P(H)}{P(T)} \]  \tag{7}

“T” represents a tuple from a data-set D and H is the likelihood of a certain tuple T falling under a certain
class. \( P(T|H) \) and \( P(H|T) \) represent the posterior probabilities denoting the conditional evaluations of
T on H and H on T, respectively. \( P(H) \) and \( P(T) \) represent prior probabilities and are unconditional,
which means that they are not dependent on others. A tuple’s posterior probability ‘T’ is calculated
against every class ‘C’, which means that the classifier is going to indicate the tuple belonging to a class
of the maximum posterior likelihood. Then, the equation above is represented more specifically as:

\[ P(C|T) = \frac{P(T|C)P(C)}{P(T)} \]  \tag{8}

In the case where the classes C (C1, C2, and Cm) then the classifier finds:

\[ P(C_i|T) > P(C_j|T) \quad \text{for} \quad 1 \leq j \leq m, j \neq i \]

\( P(C_i|T) \) is referred to as the maximal posterior hypothesis. Due to the fact that \( P(T) \) is constant and
the prior probabilities \( P(C_i) \) are assume equal, in other words,

\[ P(C_1) = P(C_2) = \cdots = P(C_m). \]

As a result, \( P(T|C_i) \) is maximized.

- **Evaluation Measures**
  - Precision of the class represents the likelihood of an arbitrary sentence has to be categorized
    with this class, in this case it is the precise decision. For example positive class precision is
    obtained as:

\[ P = \frac{TP}{TP + FP} \]  \tag{9}

- Recall of a class represents the possibility that in the case where an arbitrary sentence is to
  be categorized with the class, then it’s the taken decision. A positive class recall is
  calculated based on the following Equation(10):

\[ R = \frac{TP}{TP + FN} \]  \tag{10}

- F-Measure of a class represents harmonic weighted average for each of the calculated recall
  and precision. F1-measure has been utilized, in order to evenly weigh each of recall and
  precision. F-Measure is computed according to the following equation [27]:

\[ F = \frac{2 \times P \times R}{P + R} \]  \tag{11}

3. **Results and Discussion**

As discussed in earlier sections, an application can be used to social communication between blind and
dumb the blind can speak through telephone, which would be recognized by our system. About, 5600
voices samples collected from different speakers are considered for training and testing by using cross-
validation. It was observed that the system works efficiently. The Table 1 and Figure 3 illustrate the
classification algorithms for the system under various Datasets.
Table 1. shows the classification algorithms

| Type  | Measures | J48 | Naïve byes | KNN  |
|-------|----------|-----|------------|------|
| Letters | Precision | 100.0 | 20.012 | 94.023 |
| Letters | Recall    | 100.0 | 20.007 | 94.164 |
| Letters | F-measure | 100.0 | 18.137 | 93.510 |

Figure 3. shows the classification algorithms

4. Conclusion
The aim of this paper is to suggest a system for social communication between the blind and the dumb by recording the blind voice and analysing it using MFCC, J48, KNN and Naïve byes techniques. Feature extraction was performed using Mel Frequency Cepstral (MFCC) coefficients and classification using J48, KNN, and Naïve byes methods. The extracted features are stored in a .WAV file using the MFCC algorithm. Experimental results were analysed with by using Java, and the results were proven to be effective. The experimental shows that J48 is the best nonlinear technology in the ranking with a performance rate of 100 % while KNN 94.023%, Naïve byes 20.012% for the letters (ﻱﺃ ﻟ) .
References

[1] Zahoor Mosaad Aydam 2018 Social Communication Between the Dumb and Blind (the College of Education for Pure Science, University of Thi-Qar).

[2] Zulfiquar Ibrahim Merhej and Ghadeer Ahmed Hamed 2016 Isolated Word Recognition (Tishreen University Faculty of Software and Information Systems).

[3] Anjali Bala, Abhijeet Kumar And Nidhika Birla 2010 Voice Command Recognition System Based on MFCC And DWT (International Journal of Engineering Science and Technology Vol. 2 No 12).

[4] Eng. Rama hasan 2017 Improve the Results of the voice recognition based on the result of the integration of different systems (Tishreen University Electrical and Mechanical Engineering Department of Computer and Automatic control Engineering).

[5] Santosh K Gaikwad, Bharti W Gawali and Pravin Yannawar 2010 A Review on Speech Recognition Technique (International Journal of Computer Applications Vol 10, No.3).

[6] Nidhi Desai, Prof. Kinnal Dhameliya and Prof. Vijayendra Desai 2013 Feature Extraction and Classification Techniques for Speech Recognition (International Journal of Emerging Technology and Advanced Engineering Vol 3, Issue 12).

[7] K. M. Shiva Prasad, G. N. Kodanda Ramaiah and M. B. Manjunatha 2017 Speech Features Extraction Techniques for Robust Emotional Speech Analysis/Recognition (Indian Journal of Science and Technology Vol 10, No 3).

[8] Swathy M S and Mahesh K R 2017 Review on Feature Extraction and Classification Techniques in Speaker Recognition (International Journal of Engineering Research and General Science Vol 5, Issue 2).

[9] Hernan Faustino Chacca Chuctaya, Rolfy Nixon Montufar Mercado and Jeyson Jesus Gonzales Gaona 2018 Isolated Automatic Speech Recognition of Quechua Numbers using MFCC, DTW and KNN (International Journal of Advanced Computer Science and Applications Vol 9, No 10).

[10] Ashutosh Bhardwaj and Navneet Kumar Sahu 2013 A Survey On: Voice Command Recognition Technique (International Journal of Advanced Research in Computer Science and Software Engineering, Vol 3, Issue 5).

[11] Dr. Jaffar Alkhier 2017 Improvement of Speech Recognition by Merging Two Features Extraction Algorithms (Tishreen University Journal for Research and scientific- Engineering Sciences Vol 39, No 1).

[12] Sheya Narang and Ms. Divya Gupta 2015 Speech Feature Extraction Techniques (International Journal of Computer Science and Mobile Computing Vol 4, Issue 3).

[13] Shivanker Dev Dhangra, Geeta Nijhawan and Poonam Pandit 2013 ISOLATED SPEECH RECOGNITION USING MFCC AND DTW (International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol 2, Issue 8).

[14] Namrata Dave 2013 Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition (International Journal for Advance Research In Engineering and Technology, Vol 1, Issue V1).

[15] Parwinder Pal Singh and Pushpa Rani 2014 An Approach to Extract Feature using MFCC (IOSR Journal of Engineering Vol 04, Issue 08).

[16] Koustav Chakraborty, Asmita Talele and Prof. Savitha Upadhya 2014 Voice Recognition Using MFCC Algorithm (International Journal of Innovative Research in Advanced Engineering, Vol 1, Issue 10).

[17] Ms. Rupali S Chavan and Dr. Ganesh Sable 2013 An Overview of Speech Recognition Using HMM (International Journal of Computer Science and Mobile Computing Vol 2, Issue 6).

[18] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi 2010 Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques (Journal Of Computing Vol 2, Issue 3).

[19] Shikha Gupta, Jafreezal Jaafar, Wan Fatimah wan Ahmad and Arpit Bansal 2013 Feature Extraction using MFCC (International Journal (SIPIJ) Vol 4, No 4).
[20] Tina R. Patil and Mrs. S. S. Sherekar 2013 Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification( International Journal Of Computer Science And Applications Vol 6, No 2).

[21] Nisha 2017 Voice Recognition Technique(International Journal for Research in Applied Science & Engineering Technology (IJRASET) Vol 5, Issue V)

[22] Anshul Goyal and Rajni Mehta 2012 Performance Comparison of Naive Bayes and J48 Classification Algorithms(International Journal of Applied Engineering Research Vol 7 No 11),

[23] Uzair Bashir and Manzoor Chachoo 2017 Performance Evaluation of J48 and Bayes Algorithms for Intrusion Detection System( International Journal of Network Security & Its Applications (IJNSA) Vol 9, No 4).

[24] Susan Dumais, John Platt and David Heckerman 1997 Inductive Learning Algorithms and Representations for Text Categorization( Mehran Sahami Computer Science Department Stanford University).

[25] S B Harisha, S Amarappa and S V Sathyanarayana 2017 Kannada Speech Recognition Using MFCC and KNN Classifier for Banking Applications( International Journal of Innovative Research in Computer and Communication Engineering Vol. 5, Issue 1).

[26] Tb. Ai Munandar and Edi Winarko 2015 Regional Development Classification Model using Decision Tree Approach ( International Journal of Computer Applications Vol 114 , No 8)

[27] Vincent Labatut and Hocine Cherifi 2002 Accuracy Measures for the Comparison of Classifiers( Remote Sens Environ vol 80)