International Journal of Latest Trends in Engineering and Technology
Vol.(8)Issue(4), pp.202-206
DOI: http://dx.doi.org/10.21172/1.84.33
e-ISSN:2278-621X

SPEAKER AND SPEECH RECOGNITION USING PATTERN RECOGNITION APPROACH IN CLEAN AND NOISY ENVIRONMENT

Gurpreet Kaur[1],[2] , Mohit Srivastava[3] & Amod Kumar[4]

Abstract: Speech signal consists of information regarding message, language, speaker and individual's emotions. Though speaker recognition and speech recognition are different fields but still both fields are overlapped. Combined speaker and speech recognition system is having different applications in human life from security related areas to aids for handicapped persons. This paper presents the speaker and speech recognition using pattern recognition approach. In this Mel Frequency Cepstral Coefficient (MFCC) are used as feature extraction method and for the classification, Dynamic Time Warping (DTW) algorithm is used. The recognition is done in clean as well as in noisy environment and accuracy is evaluated in terms of different parameters like false positive (FP), false negative (FN), true positive (TP) and true negative (TN).

Keywords: MFCC, DTW, speech recognition, speaker recognition

I. INTRODUCTION

The study of speech signals and their processing methods are called speech processing [1-2]. Speech processing is very vast area and there is lot of research going in the field for the last sixty years [3]. Important fields of speech processing are synthesis, recognition and coding of speech signals. Recognition itself is a wide topic consisting of three areas i.e. speech recognition, speaker recognition and language recognition. As the name tells, recognition of words is known as speech recognition, recognition of language is called language recognition and recognition of speaker is called speaker recognition [4-5]. Speech recognition may have two modes: speaker independent speech recognition and speaker dependent speech recognition [6-7]. In speaker dependent mode, the system is trained for only one speaker but in speaker independent mode, system is trained for multi speakers. Speaker recognition field is also divided into two categories i.e. text dependent and text independent. In text dependent speaker recognition mode, the speaker is to speak same words which is known to the system but in text independent mode, speaker can speak any set of words. Though speech recognition and speaker recognition are different fields, but the feature extraction methods in both the fields are overlapped [8-11]. These methods include predictive models based on Linear predictive coding coefficient (LPCC), Perceptual linear prediction (PLP), Mel Frequency Cepstral Coefficients (MFCC) and Relative Spectra Filtering (RASTA). These methods can be implemented in speech recognition as well as in speaker recognition. In this paper, we have implemented MFCC method on combined speech and speaker recognition. There are different applications in which combined speaker and speech recognition is used [12-13]. One application is voice operated wheelchair for handicapped persons with movement disability. In this paper, different parameters like TP, FP, TN and FN are calculated. True positive (TP) are those relevant features that are selected. False positive (FP) are those features that are selected but are not relevant. True negative (TN) are those features that are not selected and they are not relevant. False negative (FN) are those features that are not selected but the features are relevant. The figure 1 shows the understanding of the above mentioned parameters.

[1] *Assistant Professor, University Institute of Engineering & Technology, Panjab University, Chandigarh, India.*
[2] *Research Scholar, I.K Gujral Punjab Technical University, Kapurthala, India.*
[3] *Professor, Chandigarh Engineering College, Landran, Mohali, India.*
[4] *Scientist, Central Scientific Instruments Organisation, Chandigarh, India.*

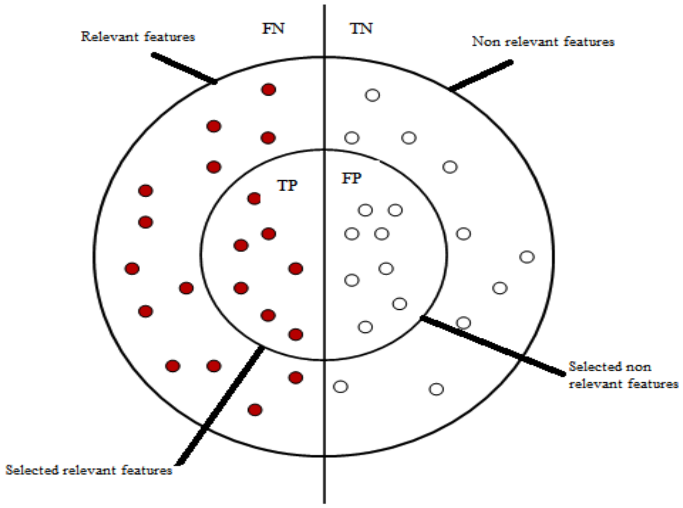


Figure 1. Understanding of feature selection parameter

Hence accuracy can be find out as

$$\text{Accuracy} = \frac{(TN+TP)}{(TN+FN+TP+FP)} \tag{1}$$

The evaluation is done for clean speech signal and by adding White Gaussian noise to the speech signal. Figure 2 shows the methodology used in speaker and speech recognition system. Speech is acquired by sound recorder with the help of headphone. Then preprocessing of each word is done to enhance the properties of the signal. For the classification, pattern matching is used. The pattern matching stage is for modeling. In this technique, the model consists of a template which is a feature vector from a fixed phrase. In this Dynamic Time Warping pattern recognition technique is used. This technique find out the distance between the input speech and stored patterns. The speaker and speech is recognized based on minimum distance.

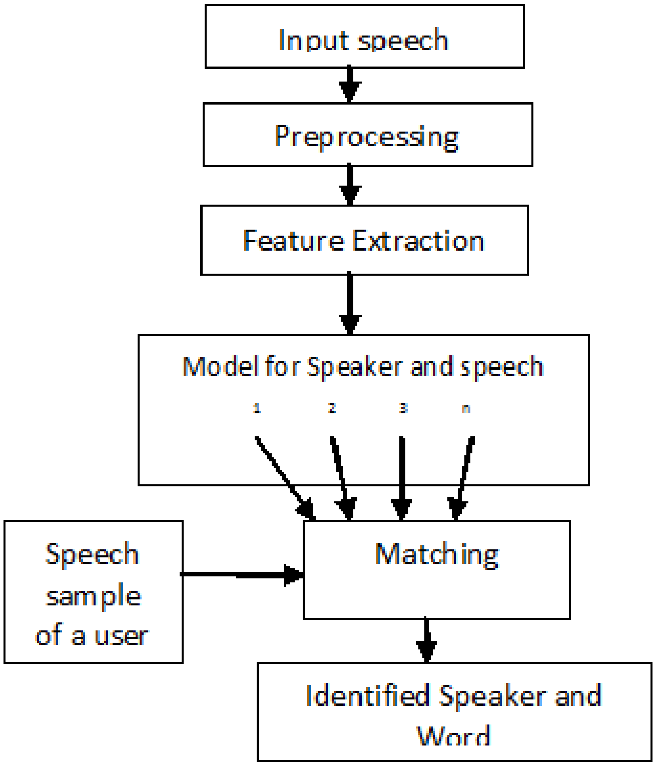


Figure 2. Block diagram of speaker and speech recognition

II. FEATURE EXTRACTION AND PATTERN RECOGNITION

The speech production mechanism can be modeled by a linear separable equivalent circuit [14-16]. This model is equivalent to a sound source G(ω) inputting into the articulation filter (vocal tract) to produce the speech. The sound source G(ω) can be categorized as a train of impulses (voiced) and random noises (unvoiced). Voiced sounds include /a/, /e/, /i/, /o/, /u/. On the other hand, unvoiced sounds are noise generated sounds such as /t/, /s/. The articulation H(ω) is a transfer function which models the vocal tract of the human speech organ. The output speech wave S(ω) is the combination of the sound source multiplied with the articulation given by the equation:

$$S(\omega) = G(\omega)H(\omega) \tag{2}$$

Feature extraction techniques like MFCC mentioned above, model the vocal tract articulation filter H(ω).

A. Mel frequency Cepstral Coefficient (MFCC)

Vocal tract system of a human can be modeled using MFCC method [17-18]. The block diagram of MFCC is as below:

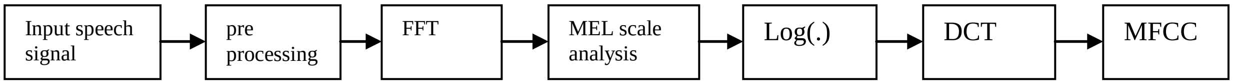


Figure 3. Mel Frequency Cepstral Coefficients

Equivalent filter of the vocal tract articulation is shown as:

$$S(\omega) = G(\omega)H(\omega) \tag{3}$$

Logarithm of $S(\omega)$ is

$$\mathrm{Log}|S(\omega)| = \mathrm{Log}|G(\omega)| + \mathrm{Log}|H(\omega)| \tag{4}$$

Cepstral coefficients are calculated by taking inverse Fourier transform of $\mathrm{Log}|S(\omega)|$.

$$C(\tau) = F^{-1}\mathrm{Log}|S(\omega)| = F^{-1}\mathrm{Log}|G(\omega)| + F^{-1}\mathrm{Log}|H(\omega)| \tag{5}$$

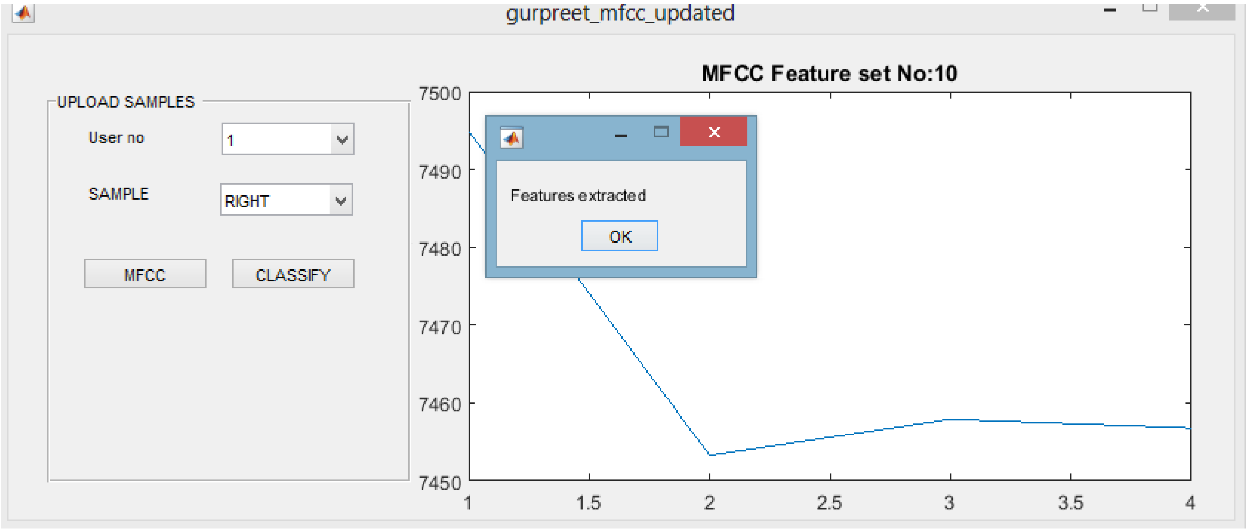
From this equation spectral envelope and fundamental frequency can be measured.

B. Pattern Recognition

Pattern recognition approach deals with pattern matching. Feature extraction gives the unique features for every word. Then reference patterns are made using the sound samples. Classification is done to match the reference pattern and test pattern and accordingly decision is made. In our system, we have used Dynamic Time Warping algorithm to compute the distance between the input speech and stored reference patterns. The speaker and speech is recognized based on minimum distance.

III. EXPERIMENT AND IMPLEMENTATION OF THE SYSTEM

A database is created with the help of sound recorder using headphone at 16 KHz frequency at room environment. Database consist of two hundred words, recorded by four speakers of age group 27-34 years, two males and two females. Five words are recorded: LEFT, RIGHT, BACKWARD, FORWARD, STOP. Ten samples are taken for each word. Using MFCC technique, features are extracted and pattern matching is done with the help of DTW algorithm. Both these techniques are implemented on recorded samples. speaker as well as speech features are extracted and MATLAB tool is used for the evaluation of the results. Graphic User Interface (GUI) is designed to make it easy to use.


gurpreet_mfcc_updated

UPLOAD SAMPLES

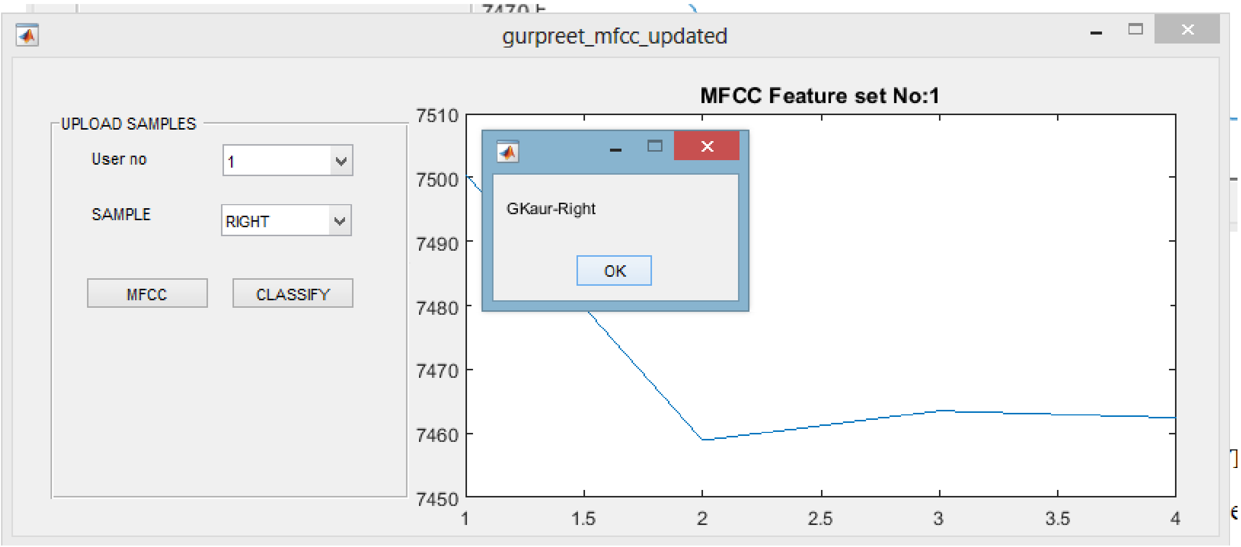
User no 1

SAMPLE RIGHT

Features extracted

Figure 4. Snapshot of speaker and speech recognition system with extracted features

Figure 4 shows one of the examples of training stage of system with extracted features. Training is done by the speaker (Female 3) for the word RIGHT and MFCC feature extraction method is used.



gurpreet_mfcc_updated

UPLOAD SAMPLES

User no: 1

SAMPLE: RIGHT

GKaur-Right

Figure 5. Snapshot of speaker and speech recognition system with recognized word

Figure 5 shows the testing stage of system with recognized word with speaker. (GKaur (Female1) is speaker and the recognized word is RIGHT). Accuracy is calculated with two types of samples. One type of samples that are recorded in room and other type of samples by adding White Gaussian Noise (WGN) in speech samples. The results are shown in table 1:

Table 1. % Accuracy for speaker and speech recognition

	Male1	**Male 2**	**Female 1**	**Female 2**
Words	**Accuracy**	**Accuracy**	**Accuracy**	**Accuracy**
Backward	**94.19**	**94.42**	**94.66**	**92.51**
Backward*	**82.14**	**84.12**	**83.95**	**84.21**
Forward	**94.69**	**93.58**	**94.31**	**95.55**
Forward*	**83.38**	**84.95**	**81.57**	**85.29**
Left	**94.73**	**94.61**	**94.22**	**93.66**
Left*	**83.86**	**82.16**	**83.12**	**85.22**
Right	**94.63**	**94.45**	**94.41**	**93.96**
Right*	**82.89**	**84.81**	**84.70**	**84.38**
Stop	**93.47**	**94.19**	**94.33**	**94.66**
Stop*	**83.29**	**85.40**	**85.23**	**83.57**

*With WGN

Table 2 shows the average recognition rate for the system is 94.25% in clean environment and when WGN is considered with the speech samples, then recognition rate decreases to 83.98%.

IV. CONCLUSION

In this paper, we have recognized the speaker as well as speech using MFCC and DTW technique for isolated words. The result shows that average accuracy for the system is 94.25% in clean environment and 83.98% in noisy environment. Using these results, voice operated patient vehicle can be modeled. There can be an improvement in results if neural network is used for classification.

V. REFERENCES

[1] D.R. Reddy, "Speech Recognition by machine: A review", proceedings of the IEEE, vol.64,issue-4, pp.501-531, 1976.

[2] P. Sehgal, and R. K. Jain, “Speech Processing,” vol. 5, no. 2, pp. 83–87, 2013

[3] S. Furui, “50 Years of Progress in Speech and Speaker Recognition Research,” vol. 1, no. 2, pp. 64–74, 2005

[4] J. Campbell, "Speaker Recognition, A Tutorial", proceedings of the IEEE, vol.85, number 9, , 1997

[5] L. Mary, and B. Yegnanarayana, “Extraction and representation of prosodic features for language and speaker recognition,” *Speech Communication.*, vol. 50, no. 10, pp. 782–796, 2008

[6] I. Bhardwaj, “Speaker Dependent and Independent Isolated Hindi Word Recognizer using Hidden Markov Model (HMM),” vol. 52, no. 7, pp. 34–40, 2012

[7] N. Carolina, “The Recognition of Isolated Words On a Speaker Dependent System,” pp. 5–8, 1989

[8] S. Squartini, E. Principi, R. Rotili, and F. Piazza, “Neurocomputing Environmental robust speech and speaker recognition through multi-channel histogram equalization,” *Neurocomputing*, vol. 78, no. 1, pp. 111–120, 2012

[9] N. S. Dey, R. Mohanty and K. L. Chugh, “Speech and Speaker Recognition System Using Artificial Neural Networks and Hidden Markov Model,” 2012

[10] D. A. Reynolds and L. P. Heck, “Recognition Systems,” pp. 869–872, 1991

[11] T. Gaafar, H. Bakr, M. Abdalla, " An Improved Method for Speech/Speaker Recognition", proceedings of IEEE, 2014

[12] G Kaur., M Srivastava., A. Kumar, "Analysis of Feature Extraction Methods for Speaker Dependent Speech Recognition," International Journal of Engineering and Technology Innovation, vol. 7, No. 2, pp. 78 - 88", ISSN 2223-5329(P), 2226-809X(O), 2017

[13] V. Fontaine and H. Bourlard “Speaker Dependent Speech Recognition Based on Phone-Like Units Models Application To Voice Dialing,” pp. 2–5

[14] S. King, J. Frankel, K. Livescu and E. Mcdermott, “Speech production knowledge in automatic speech recognition,” no. February, pp. 723–742, 2007

[15] A. B Scassellati, “Gender Recognition in Adult Male and Female Speech,” 2004

[16] H. Hoge, “Basic Parameters in Speech Processing The Need for Evaluation.”

[17] G. Kaur, R. Khanna and A. Kumar, "Automatic Speech and Speaker Recognition using MFCC: Review", International Journal of Advances in Science and Technology, Vol. 2, Issue 3, 2014

[18] G. Kaur, R. Khanna and A. Kumar, "Implementation of Text Dependent Speaker Verification on MATLAB", proceedings of IEEE, 2015