Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Classification of Fluent and Dysfluent Speech Using KNN Classifier

[1]Monica Mundada, [2]Sangramsing Kayte, [3]Sumegh Tharewal, [4]Dr. Bharti Gawali
[1234] Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

Abstract*— Speech is a main source of human communication. Dysfluent speech is a problem with fluency, voice, and or how a person produces a speech sound. The main objective of the study is to identify the distinguish properties between fluent and dysfluent speech. The proposed work classifies the fluent and dysfluent speech. The KNN classifier is used to classify the fluent and dysfluent speech with classification rate of 93%. The experimental investigation elucidated MFCC and DTW with the accuracy rate of 88 % and 75% respectively.*

Keywords— MFCC, dysfluent speech, DTW, KNN and feature extraction.

I. INTRODUCTION

Speech signal is the most natural method of communication between humans. Normal speech sounds is produced through a series of precisely coordinated muscle movements involving breathing, phonation (voice production), and articulation (movement of the throat, palate, tongue, and lips). Muscle movements are controlled by the brain and monitored through our senses of hearing and touch. Lungs act as source of energy. Then, Vocal folds chop the airflow from lungs to produce quasi periodic excitation. The Shape of the vocal tract determines nature of the sound unit to be produced. Speech is an outcome of time-varying vocal tract filter driven by a time-varying excitation [5]. The state of the vocal cords, the positions, shapes and sizes of the various articulator changes slowly over the time which thereby results in producing the desired speech sounds [4]. However, there is 1% of the population have noticeable speech stuttering problem and it has been found to affect female to male with ratio 1: 3 or 4 times [1]. Dysfluent speech disorder is defined as a normal flow of which is disrupted by unintentionally of dysfluencies such as repetition, prolongation, interjection of syllables, sounds, words or phrases and involuntary silent pauses or blocks in communication. It is also defined as stammering or stuttering. Stuttering cannot be completely cured; it may go into remission for some time [1, 2]. Speech pathology treatments help stutterers to shape their speech into fluent speech. Therefore a stuttering assessment is needed to evaluate performance of stutterers before and after therapy. Traditionally, speech language pathologist (SLP) counts and classifies occurrence of dysfluencies such as repetition and prolongation in stuttered speech manually [3]. The evaluation of these treatment are subjective, inconsistent, time consuming and prone to error. Therefore, it might be good if stuttering assessment can be done automatically and thus having more time for the treatment session between stutterer and SLP. The work is beneficial to know the area of improvement in dysfluent speech at early stages. The variations in fluent and dysfluent speech are studied with the parameter of distance matrix. This manuscript studies the difference in dysfluent speech signal for Hindi and Marathi database. The experimental work in performed in Matlab2012a with inclusion of MFCC and DTW to calculate the distance matrix. The KNN classifier is studied to distinguish the fluent and dysfluent speech signals. The remaining part of paper is organized as follows: Methodology is described in section 1. The experimental setup and data acquisition are described in section 2. Feature extraction technique like MFCC and DTW are explained in section 3. The KNN classifier is elaborated in section4. Observation and results are followed in section5. Conclusion and future work in section6. References are mentioned in section7.

A. Methodology

The proposed work classifies the fluent and dysfluent speech samples with KNN classifier. The input speech signal is both fluent and dysfluent signal. Then the feature extraction is performed using MFCC. The energy entropy derived as MFCC coefficient is then considered as classifier derivative for KNN. Figure 1 shows the dataflow for proposed work.

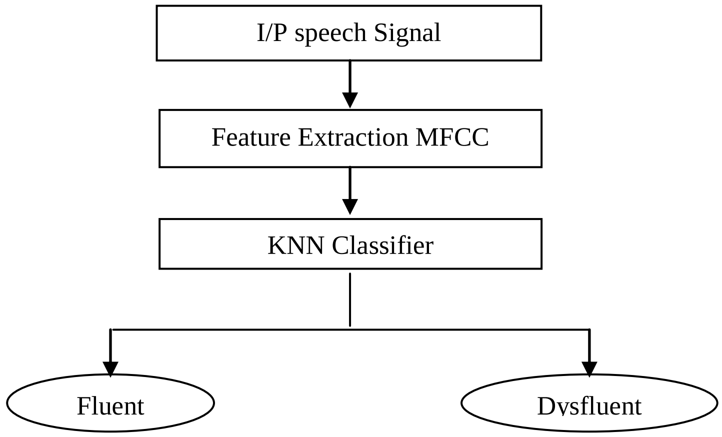


Figure1. Dataflow for proposed methodology

B. Database

The database is designed with the suggestion of speech therapist for the dysfluent speech. It consists of 10 sentences of both Hindi and Marathi language. The same database is used for fluent subjects to study the variations associated between them. The table 1 depicts the list of sentence used for creating the database.

TABLE 1
DATABASE FOR RECORDING OF FLUENT AND DYSFLUENT SPEECH SAMPLES

Sr.No	Sentences
1	मै आज क्लास गया था
2	आप कहाँ गये थे
3	मेरी आज परीक्षा था
4	काकूने काका साठी काकडी कापली
5	कबीर ने कल कपडे पहेना
6	गोपाल गाते गाते खेत में जा रहा था
7	घाट के उपर घोडा घास खा रहा था
8	मैने खाना खाया
9	खाते खाते मुझे खर्गोश दिखाई दिया
10	खाने में श्रीखंड खाया

C. Acquisition setup

The speech samples are collected for fluent and dysfluent speech subjects with the same age and gender. The quality of recording is maintained by capturing the signals using Visi Pitch. The use of Visi Pitch avoids the pre-processing of data signals. The frequency of 16000 kHz was maintained with the sound proof room and other parameters like room temperature and humidity are also controlled with normal conditions. The age of subject is ranging from 17 — 26 years. The database consist of 150 samples which includes 2 male and 1 female for fluency disorder and equivalent 150 samples for the normal subject with same criteria like age and gender are considered.

TABLE 2
SIZE OF DATABASE

Subject	Size of database
Fluent	150 utterance
Dysfluent	150 utterance

D. Feature Extraction technique

Feature extraction is the parameterization of the speech signal. This process helps to extract the necessary information from the signal and to discard the unwanted and redundant information. This is the most important step as the coefficients obtained are further processed for classification. The feature extraction helps to measure various speech derivatives like pitch, energy or frequency response. There are many feature extraction techniques like LPC, DTW and many more.

II. MEL FREQUENCY CEPSTRAL COEFFICIENTS MFCC

MFCC is the most robust feature extraction technique as it is based on known variation of the human ear's critical bandwidth with frequency [5]. The best values in the parametric representation of acoustic signals are an important task to produce a better recognition performance. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz [2]. Mel frequency scale is used to study the phonetically characteristics of speech signal. MFCC is a spectral analysis method. The overall process of the MFCC is shown in Figure 1.

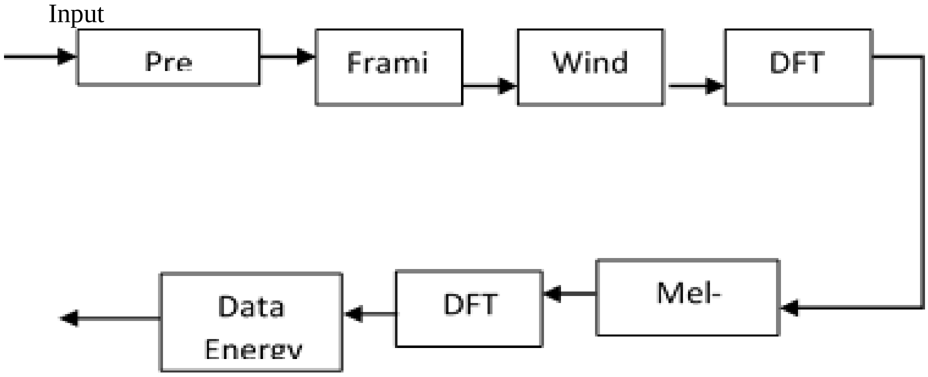


Figure 2. Block diagram of MFCC

A. *Dynamic time warp (DTW)*

DTW is a template based feature extraction approach. DTW is a technique that calculates the level of similarity between two time series in which any of them may be warped in a non linear fashion by shrinking and stretching the time axis [3]. Dynamic Time Warping (DTW) is an efficient method for finding this optimal nonlinear alignment. In this study minimum distance is calculated from test speech signal to each of the training speech signal in the training set [6]. This classifies test speech sample belonging to the same class as the most similar or nearest sample point in the training set of data. A Euclidean distance measure is used to find the closeness between each training set data and test data

III. CLASSIFICATION TECHNIQUES

Classification technique involves studying the distinguished features and creating them in different groups depending on their feature set. Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations whose category membership is known. In this work our focus is to perform two class classification i.e fluent and dysfluent speech samples using KNN classifier.

A. *KNN classifier*

KNN classifies new instance query based on closest training examples in the feature space [7]. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is delayed until the classification is done. Each query object (test speech signal) is compared with each of training object (training speech signal). Then the object is classified by a majority vote of its neighbors with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor [8] [10]. Our aim is to perform two class classification (fluent vs. dysfluent) using the MFCC features. We have considered, training data set; one for dysfluent speech samples that includes 3 types of dysfluencies such as repetitions, prolongations and interjections, second training data set is for fluent speech. For each test samples the training data set is found with k nearest members. Further, for this k nearest members, suitable class label is identified based on majority voting. Class labels can be fluent speech or dysfluent speech.

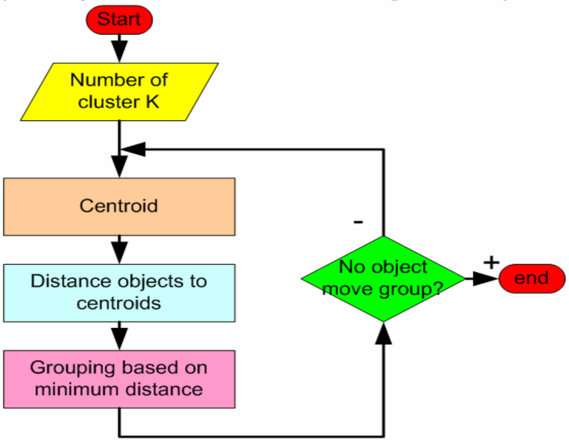


Figure 2. Flowchart of KNN classifier

The flowchart of KNN depicts the fluent and dysfluent speech signals with respect to their energy entropy value. Here the distinction is made by considering the energy aspect ratio for both the signals. The figure 3 plots initially all the data set of both fluent and dysfluent values. The energy entropy value differentiates the both the speech samples.

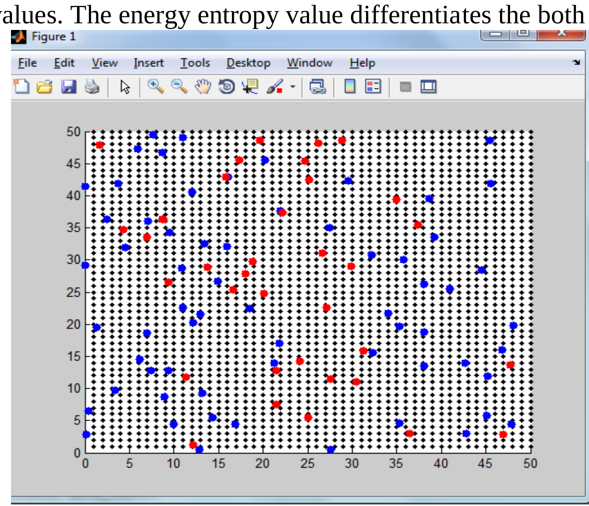


Figure 3. Plots the energy entropy value of both fluent and dysfluent speech samples

Initially here k= 1, plots the testing data and make the clusters with the nearby centriod value. Figure 4.shows the values plotted for k=1.Here the testing data of size 50 utterances of dysfluent value are passed for making group centroids. KNN now classifies the train data with the testing data and make groups depending on the centriod value accepted. Figure5. Plots the testing data with k=2. The value of k= 3 and k=4 is plotted in figure 6 and 7 respectively.

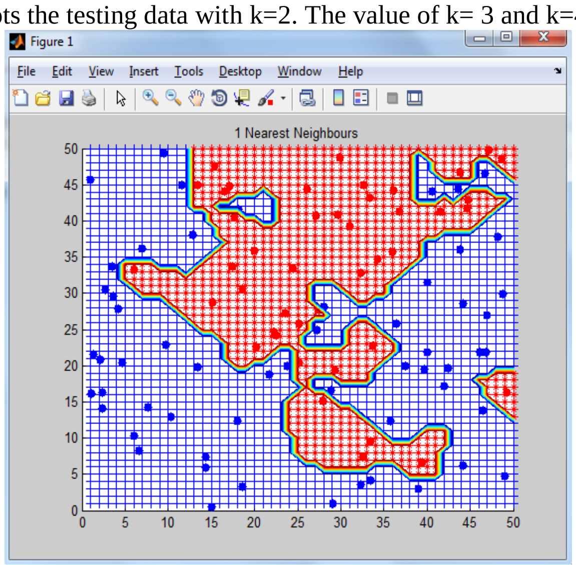


Figure 4.Plots the values of testing data with value k=1

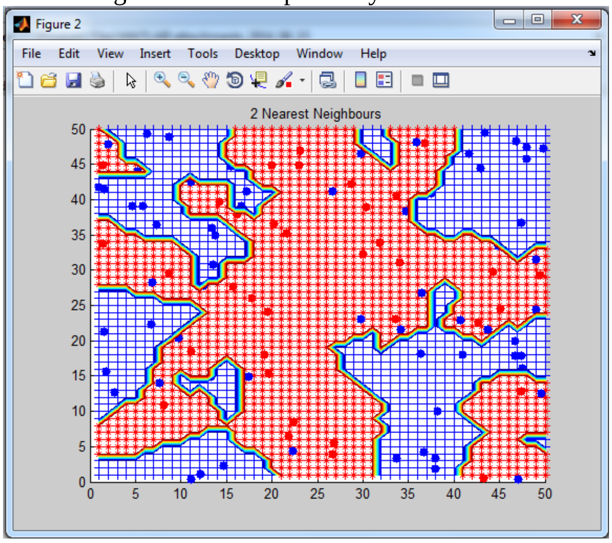


Figure 5.Plots the values of testing data with value k=2

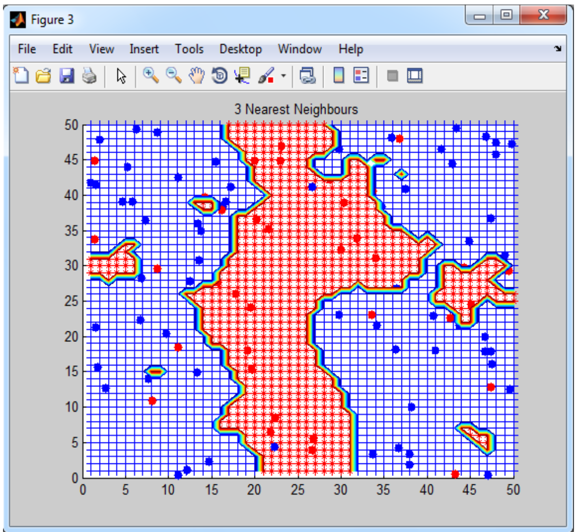


Figure 6.Plots the values of testing data with value k=3

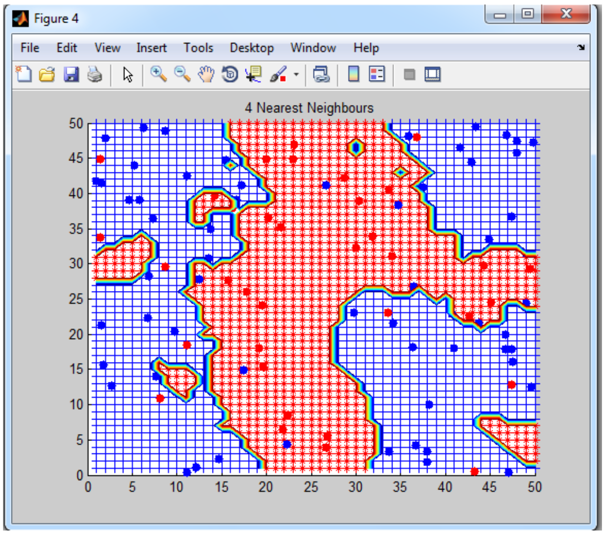


Figure 7.Plots the values of testing data with value k=4

k=5, finally the clear classification is produced with more number of energy values with respect to other iterations of k. Thus KNN helps to classify the fluent and dysfluent speech on basis of their energy entropy value. As the number of iterations of k increases, the clusters are placed more nearly to the centroid and produce the classified output. The observation shows that the classification rate of KNN classifier is 93%.

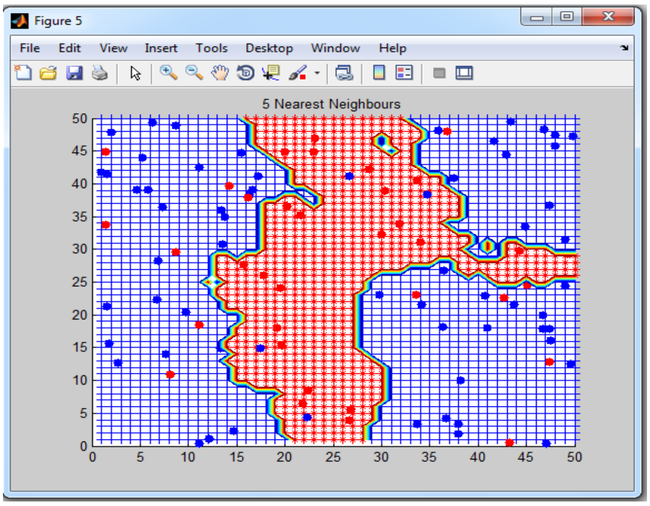
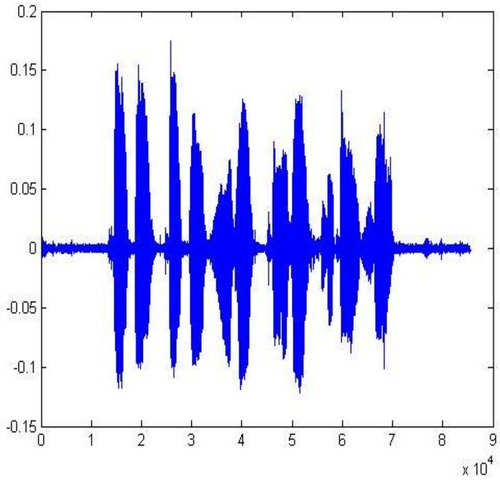


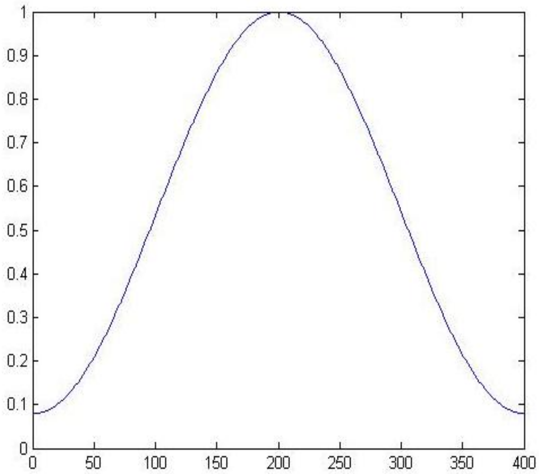
Figure 8.Plots the values of testing data with value k=5

IV. Observation and Results

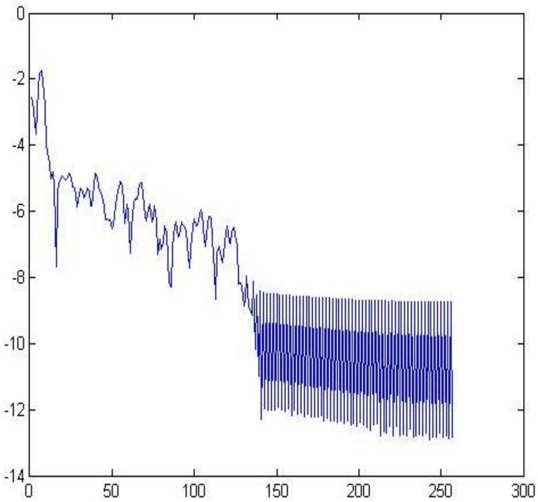
The simulation work is performed in MatlabR2012a. The experimental analysis included MFCC and DTW calculations for both fluent and dysfluent subjects. MFCC for Fluent speech for sentence 1 from database table is displayed in figure 4.



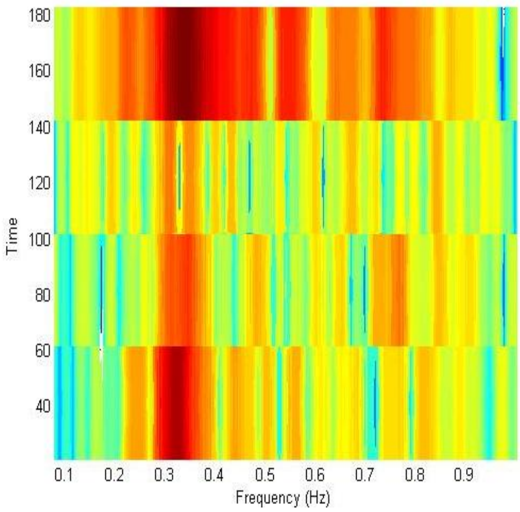
5.3.a. Original signal of fluent Subject



5.3.b Hamming window of fluent Subject

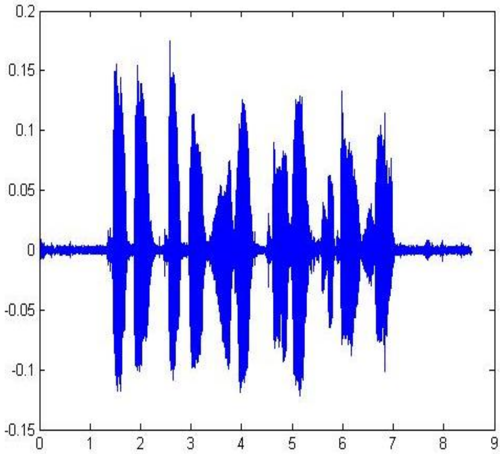


5.3.c DCT of fluent subject

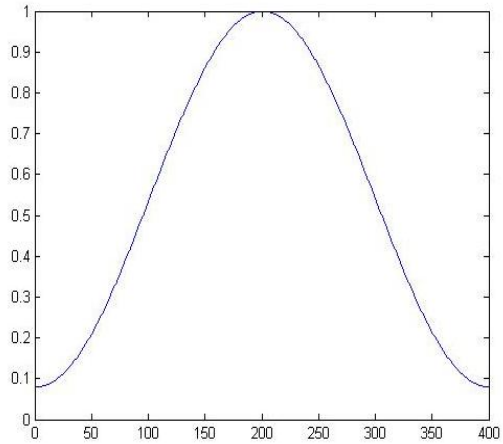


5.3.d MFCC of fluent Subject

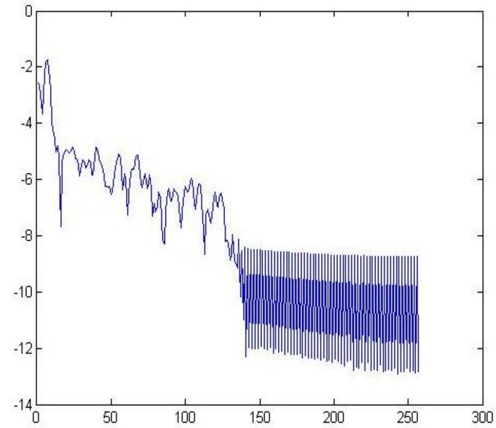
Figure 5.4 shows the various stages of MFCC for dysfluent subject for sentence 1 from database table.



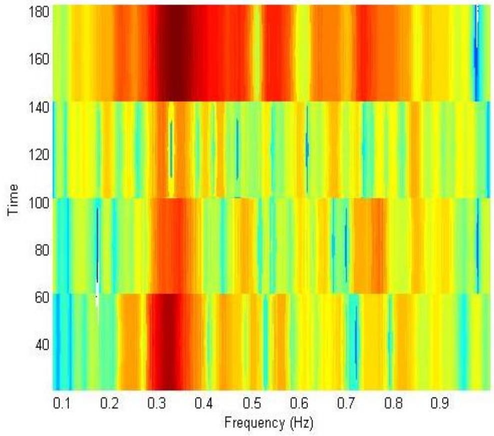
5.4.a Original Signal of dysfluent subject



5.4.b Hamming window of dysfluent subject



5.4.c DCT of dysfluent subject



5.4.d MFCC of dysfluent subject

The statistical parameters like mean and standard deviation of fluent and dysfluent subject is calculated to study the variations associated with them. Table 2 depicts the statistical values for fluent and dysfluent speech samples. Table 3 shows the comparative recognition rate of MFCC and DTW values.

TABLE 3

MEAN AND STANDARD DEVIATION VALUES FOR FLUENT AND DYSFLUENT SPEECH

Parameter		**Dysfluent Speech**		**Fluent Speech**	
Age	**Gender**	**Mean**	**Std.Dev**	**Mean**	**Std.Dev**
26	M	3.704	9.583	4.384	6.698
17	M	3.925	9.388	4.575	8.512
21	F	3.747	10.884	4.125	8.767
21	M	3.773	10.848	4.703	8.658
26	M	4.002	10.514	4.384	6.698

TABLE 4

COMPARATIVE RECOGNITION RATE OF MFCC AND DTW

	Vector size	**Recognition rate**
MFCC	150	88%
DTW	150	75%

V. CONCLUSION

The proposed study classifies the fluent and the dysfluent speech with KNN classifier with classification rate of 93%. The study also measures the distance matrix using MFCC and DTW in dysfluent and fluent speech samples with accuracy rate of 88% and 75%. In future, the work will be extended to study the accurate position of production system which produces the dysfluent speech. The expert system will be designed to classify the speech and outlines the factors responsible for various speech disorders.

REFERENCES

[1] M. Hariharan & Lim Sin Chee & Ooi Chia Ai & Sazali Yaacob." Classification of Speech Dysfluencies Using LPC Based Parameterization Techniques". Springer, LLC 2011. DOI 10.1007/s10916-010-9641-6.

[2] Tian-Swee, T., Helbin, L., Ariff, A. K., Chee-Ming, T., and Salleh, S. H., Application of Malay speech technology in Malay SpeechTherapy Assistance Tools. Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on, 2007, pp. 330–334.

[3] Akanksha Singh Thakur, NamrataSahayarn " Speech Recognition Using Euclidean Distance" International Journal of Emerging Technology and Advanced ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3, March 2013.

[4] ANJALI BALA, ABHIJEET KUMAR, NIDHIKA BIRLA." VOICE COMMAND RECOGNITION SYSTEM BASED ON MFCC AND DTW". International Journal of Engineering Science and Technology. Vol. 2 (12), 2010, 7335- 7342.

[5] Rabiner, Lawrence R., and Ronald W. Schafer. Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[6] 6. Kashyap Patel, R.K. Prasad." Speech Recognition and Verification Using MFCC & VQ". International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319-6378, Volume-1, Issue-7, May 2013.

[7] 7. P. Mahesha S. J. College of Engineering, Mysore D. S. Vinod S. J. College of Engineering, Mysore Feature based classification of dysfluent and normal speech.

[8] 8. S.K.Mangal." Statistics in psychology iarlc4 education". PHI Learning Private Limited 2010.

[9] 9. Najafian, Maryam (2013). 'Modeling accents for automatic speech recognition'. University of Birmingham Graduate School Research Poster Conference 2013, 12th June 2013, University of Birmingham.

[10] 10. Monica Mundada, Bharti Gawali ." A review and study of influence factors associated with speech and its disorders". International Conference on Emerging Research in Computing, Information, Communication and Applications. (ERCICA-14).

