Spotting the aberration spot in a speech with the aid of Fuzzy Inference system

C.R. Bharathi[1], Dr.V. Shanthi[2]

[1]*Research Scholar, Sathyabama University, Assistant Professor, Department of ECE,, Vel Tech University, Avadi, Chennai, India.*
[2]*Professor, Department of MCA, St. Joseph's College of Engineering, Chennai, India.*

Abstract

A wide range of researches are carried out in this field for denoising, enhancement and more. Besides the other, stress management is important to identify the spot in which the stress has to be made in speech. In this paper, in order to provide proper speech practice for the abnormal child (*mentally retarded (MR) child)*, their speech is analyzed. Initially, the normal and abnormal children speech is obtained with the same set of words. As an initial process, the Mel Frequency Cepstrum Coefficients (MFCC) is extracted from both words and the Principal Component Analysis (PCA) is applied to reduce the dimensionality of the words. From the dimensionality reduced words, the parameters are obtained and then these parameters are utilized to train using Support Vector Machines (SVM) for classification. After identifying the acute word (abnormal word), through the thresholding operation and then FFT is computed for the acute word and these parameters made use of the Fuzzy Inference system (FIS) for blemishing the acute spot in which the aberration is occurred in the world where the speech practice is required for the abnormal child which helps speech pathologist.

Keywords: Speech signal, Stress, Mel Frequency Cepstrum Coefficients (MFCC), Principal Component Analysis (PCA), Support Vector Machines (SVM), Fuzzy Inference system (FIS)

1. Introduction

Effective determination of quantitative speech disability remains one of the challenges in medical profession. Mental Retardation/Intellectual Disability (MR/ID) has been particularly challenging as the term *mentally retarded* carries significant social and emotional stigma. Developmental delay is often used inappropriately as synonymous with MR/ID. Speech synthesis among the disabled children is the first step for making speech corrective tool kit.

Generally, the speech recognition has two stages: feature extraction and classification (Bharathi and Shanthi, 2011).
This work is very useful for the speech practitioners in the way, in which position they have to improve the speech of the abnormal person. Initially, the samples of the normal as well as the abnormal person's speech have been obtained and with the aid of these samples, the further process has to be carried out. Initially, the MFCC of both the speeches are extracted and the PCA is applied to the MFCC to reduce the dimensionality of the speeches. After that the parameters are extracted from this MFCC of both speeches and then these speeches are inputted to generate the ANN.

The abnormal and normal features are used to train the network. Subsequently, the acute word is identified through the thresholding operation and then FFT of the acute word is identified. From the FFT of the acute word, number of maximum peaks and their amplitude is identified. The amplitude and number of peaks are inputted to the Fuzzy Inference System for the optimization process to identify the position in which the stress has to given.

1.1 MFCC

Feature extraction is a key issue for efficient speaker recognition. Additionally, a reduced feature set would allow more robust estimates of the model parameters, and less computational resources would be required. Best features are those that help to discriminate among speakers. A small amount of data is enough to estimate good models. State-of-the-art systems use the same short-term spectrum features (Mel-Frequency Cepstral Coefficients, MFCC) for speech and speaker recognition, because MFCC convey not only the frequency distribution identifying sounds, but also the glottal source and the vocal tract shape and length, which are speaker specific features. Additionally, it has been shown that dynamic information improves significantly the performance of recognizers, so MFCC are commonly used as features (Bharathi and Shanthi, 2011).

1.2 Principal Component Analysis (PCA)

PCA is an technique of multivariate statistical analysis (Bharathi and Shanthi, 2011), consists of computing the eigenvectors of the D*D covariance matrix X, then sorting them according to the corresponding eigenvalues, in descending order, and finally building the projection matrix A (called Karhunen-Loeve Transform, KLT) with the largest K eigenvectors (i.e. the K directions of greatest variance). Each feature vector X is then pre-processed according to the expression Y=A(X-U), where U represents the mean feature vector. KLT decorrelates the features and provides the smallest possible reconstruction error among all linear transforms, i.e. the smallest possible mean-square error between the data vectors in the original D-feature space and the data vectors in the projection K-feature space (Bharathi and Shanthi, 2011).

1.3 SVM

An SVM is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on (Bharathi and Shanthi, 2012).

1.4 Fuzzy Inference system (FIS)

Fuzzy knowledge -based systems are rules that built on fuzzy logic and fuzzy set theory. A fuzzy rule system is a rule system whose variables or part of its variables are linguistic variables (Mohammad Jalali Varnamkhasti, 2011). It is well known that Fuzzy Inference Systems (FIS) can be used to approximate closely any nonlinear input output mapping by means of a series of if–then rules. In the design of FIS, there are two major tasks, viz. the structure identification and the parameter adjustment (Hai-Jun Rong, 2006). A specific approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS), which has shown significant results in modelling nonlinear functions (Inan Guler and Elif Derya Ubeyli, 2005). ANFIS is a well known artificial intelligence technique that has been used currently in hydrological processes. The ability of neural network to learn fuzzy structure from the input output data sets in an interactive manner has encouraged many researchers to combine the ANN and the fuzzy logic effectively to organize network structure itself and to adapt the parameters of fuzzy system (Sven Nordholm et al., 2010). Generally, fuzzy control method is fit for systems with complex mathematical models. In addition, to determine the fuzzy rules, it is generally the same parameters, so the system does not have parameters or the parameters are relatively small (Mohammad Jalali Varnamkhasti, 2011). Fuzzy inference systems, despite its good performance in terms of accuracy and interpretability, have seen little application in the field of time series prediction as compared to other nonlinear modelling techniques such as neural networks and support vector machines (Dinesh et a., 2011).

The rest of the paper is organized as follows; the Feature Extraction, Classification using SVM and finding spot using FIS are described in Section 2. The experimental setup is described in Section 3, including the speech database used to train. Experimental Results are presented and discussed in Section 4: (1) Feature extraction using MFCC and PCA, (2) Training and Classification using SVM and (3) Finding acute spot using FIS. Finally, Section 5 summarizes our approach.

2. METHODOLOGY

2.1 Feature Extraction

Normal and pathological subject's speech dataset is obtained and then the MFCC is extracted from both the speeches. This speech dataset is developed through in which both the databases are same set of scripts. Subsequently, with the aid of the PCA the dimensionality is reduced for MFCC features extracted (Bharathi and Shanthi, 2012).

2.2 MFCC

The speech samples are extracted from the normal and pathological subjects with the aid of the audio synthesizer. Let AB D_a and N D_b are the abnormal children, normal children speech datasets respectively and from these datasets the MFCC feature is extracted. [4]

$$ABD_a = \{w_1, w_2, w_3 ... w_{N_w - 1}\} \quad (1)$$

$$ND_b = \{\omega_1, \omega_2, \omega_3 ... \omega_{N_w - 1}\} \quad (2)$$

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. The MFCC features are obtained from the normal as well as abnormal datasets which is referred as N M_a and AB M_b PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics - because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it. PCA was used to reduce the dimensionality. After this process completed, the following parameters are obtained from the MFCC featured vectors N M_a ABM_b The parameters are mean, standard deviation, maximum amplitude value and its id, minimum amplitude value and its id, MFCC length are extracted for the MFCC featured word and as well as for the original word also extracted and hence for each word we have 14 inputs (Bharathi and Shanthi, 2012).

2.3 Classification through Support Vector Machines (SVMs)

SVM is

An optimally defined surface

Linear or nonlinear in the input space

Linear in a higher dimensional feature space

Implicitly defined by a kernel function

K (A, B) → C

The very best algorithms today are typically quadratic and required multiple scans of the data (Bharathi and Shanthi, 2012).

Figure . 1: Data Flow for Feature Extraction

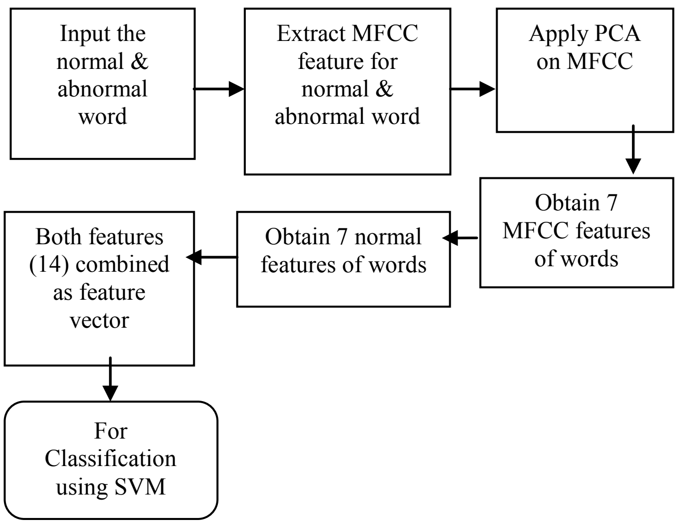
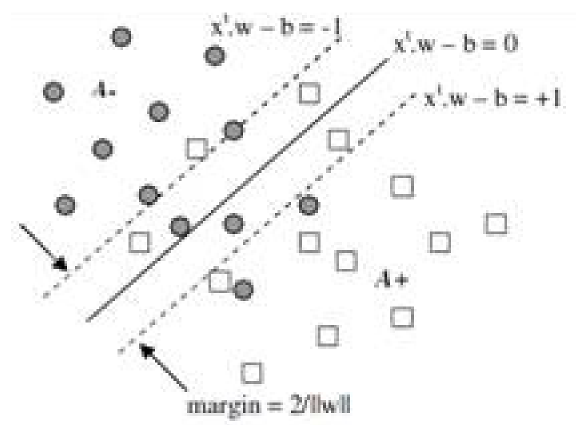


Figure. 2: Linear separation of the datapoints into two classes



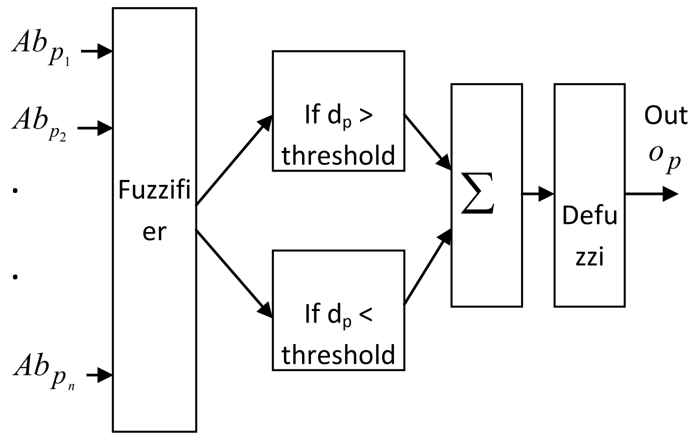
Let us consider a linear binary classification task, as depicted in figure 1 & 2, with m data points in the n dimensional input space Rn, represented by the mxn matrix A, having corresponding labels ±1, denoted by the mxm diagonal matrix D of ±1. For this problem, the SVMs try to find the best separating plane, i.e. furthest from both class +1 and class -1. It can simply maximize the distance or margin between the support planes for each class (xtw = b+1 for class +1, xtw = b-1 for class -1). The margin between these supporting planes is 2/||w||. Any point falling on the wrong side of its supporting plane is considered to be an error. Therefore, the SVMs have to simultaneously maximize the margin and minimize the error Both the words of normal and pathological subjects are inputted to the SVM to be trained and to identify the abnormal word. The 14 features extracted for each word is given as input to SVM. Here we combine these 14 inputs into a single vector for training SVM i.e. IP . The SVM is utilized to train this IP with the intention of recognizing the abnormal words.

2.4 Proposed Technique for identifying abnormal spot

2.4.1 Fuzzy Inference system (FIS) to spot the aberration spot

The fuzzy inference system is comprised of 3 phases that are Fuzzification, Rules Evaluation and Defuzzification. Fuzzy inference is the process of creating a mapping from a given input to an output by means of a fuzzy logic. Then, the mapping provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves Membership Functions, Logical Operations, and If-Then Rules. In this FIS, we utilize the peaks of the abnormal words in order to identify the aberration spot. These peak values are utilized for the fuzzy system (Figure 3).

Figure 3: Fuzzy Inference System Structure



2.4.2 Fuzzification

In the fuzzification process, the crispy quantities are changed into fuzzy in order to accomplish the tasks. In our proposed method, the peaks of each abnormal word are utilized for identifying the apt aberration spot.

2.4.3 Fuzzy Rules generation

Subsequent to the fuzzification, the fuzzy rules are generated based on the fuzzy values of each word. Generally, fuzzy rules are in the form of "IF A THEN B". The IF-part of the rule is called as antecedent, and the THEN-part of the rule is called as the conclusion. The generated fuzzy rules will train the FIS with the output values based on the threshold.

2.4.4 Defuzzification

In the defuzzification process, the fuzzy set is utilized as an input and output is obtained as a single number. To the extent that fuzziness supports the rule evaluation during the intermediate steps, and the final output for every variable is usually a single number.

3. Experimental Setup

3.1 The speech database

In this with the aid of the **Free Audio Editor** we generate the dataset with the normal and abnormal female children within the age limit 6-10 yrs. For normal data 2 female children are utilized and for abnormal data a female child is untilized for the system and their normal frequency range is from 20 - 4 kHz. The proposed technique is tested with the database of 100 words with two normal children and an abnormal child each of same scripts (Bharathi and Shanthi, 2011).

3.2 Classification using SVM

Here we detail the SVM classification for the abnormality of words. The following equation is the SVMs' objective function which may identifies the support vector for the classification.

$$R\ s = \sum_i \omega_i * K\ (s_i, Ipr) + b_i \qquad (3)$$

Here

s_i = support vectors

ωi = weight

b = bias

Ipr = vectors for classification

kr = kernel function

The equation (3) is the objective function utilizes an optimization method to identify the support vectors, weights and bias for classifying the vector IP where kr is a kernel function. In the case of a linear kernel, kr is the dot product (Figure 4).

Figure. 4: Normal speech 1 "wild animals"

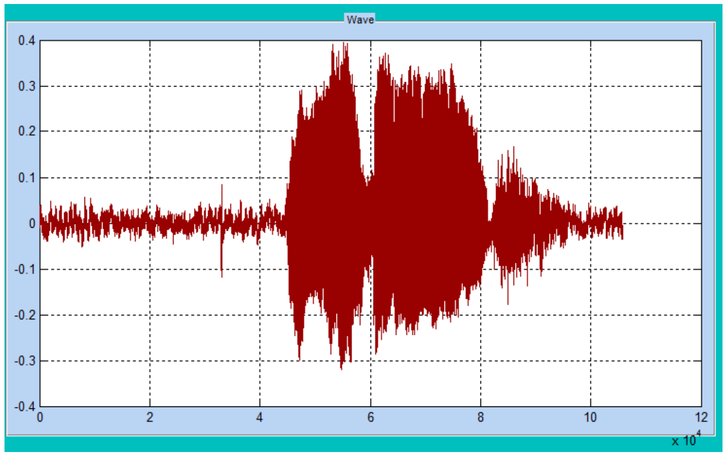


Figure. 5: Normal speech 2-"wild animals"

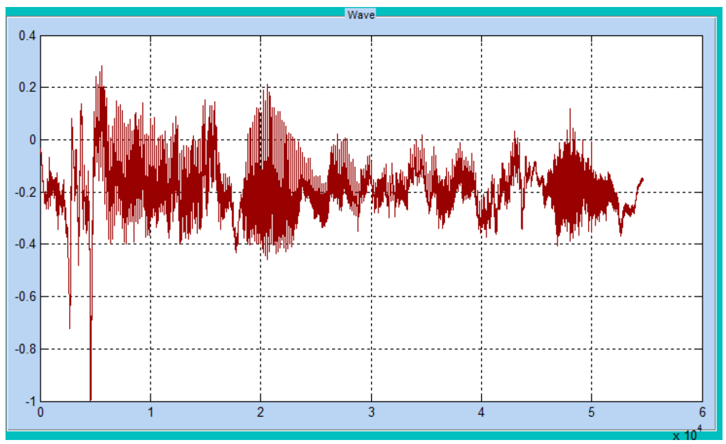
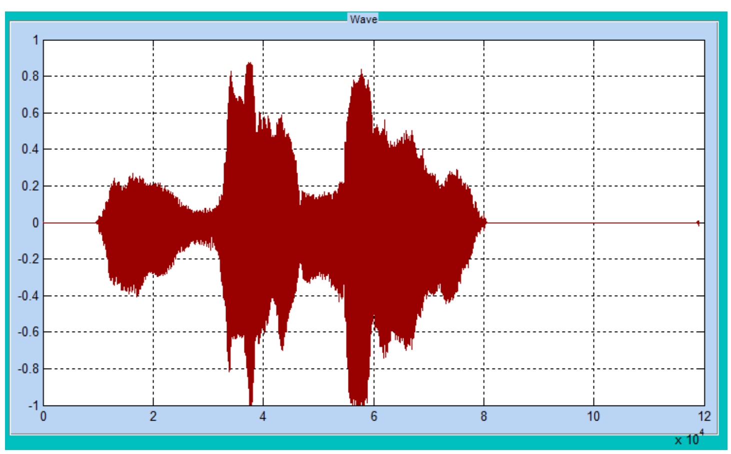


Figure. 6: Abnormal speech – "wild animals"



3.2.1 Algorithm:

Step 1: Check whether the "Res" is greater than Threshold value

Step 2: If "Res" is greater than or equal to the Threshold value then:
The word is considered as "Normal"

Step 3: Else:
The word is considered as "Abnormal"

The above algorithm indicates that if the value of the variable "Res" greater than the threshold value then the class belongs to the normal category else it may belong to the abnormal category. SVM contains error also the error minimization function is as follows

$$\arg\min P \sum_{x=0}^{n_s'-1} \nu_x + 0.5\lambda^T.\lambda \quad (4)$$

with the following constraints,

$$\ell_x(\lambda^T k\ (\hat{Ipr}_x) + c1) \geq 1 - \nu_x \quad (5)$$

and

$$\nu_x \geq 0 \quad (6)$$

In Eq. (4), p is the penalty constant, ν - handles the data and λ - matrix of coefficients. According to the Eq. (5) and (6), ℓ_x is the class label of the x^{th} dataset, $c1$ is a constant and k is the kernel which transforms the input data to the feature space. Hence, by minimizing the error function, the SVM learns the training dataset $\hat{Ipr}$ well and so that it can classify the vector that is similar to the training set. After the errors are minimized we obtain the abnormal words separately. [Excite] (Figure 7)

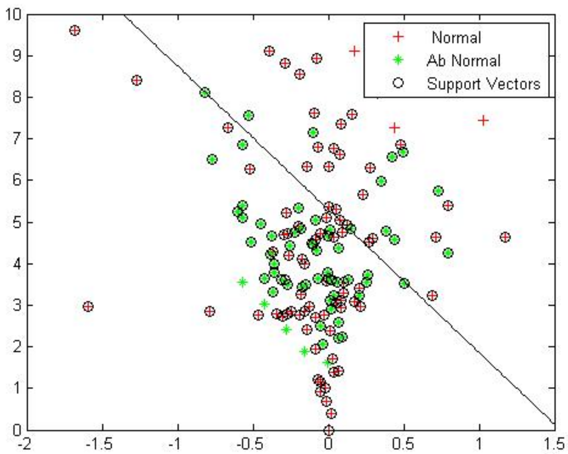


Figure 7: Regression Result for the SVM Classification to Identify the Abnormal Word

TABLE 1: Few output for THE SVM Classifier for the Identification of Abnormal Word

Original	Proposed (SVM)
grapes.wav	abnormal word
kumudha.wav	abnormal word
Sugar	abnormal word
Fish	normal word
Dinosaur	abnormal word
ink pot	abnormal word
Umbrella	abnormal word
Rabit	abnormal word
Wild animals	normal word
Aeroplane	abnormal word

3.3 Identifying the acute spot of abnormality

In this section, the spot in which the speech practice is required is found. Prior to this process, the identified abnormal word is stored in the vector A_v comprised of abnormal words. With the aid of the thresholding operation the acute word is extracted and stored. In this module, the identified abnormal speech sample is utilized to identify the position in abnormal speech, to be trained by the speech practitioners. For this, the abnormal speech is identified and the FFT is computed for the abnormal speech. The sudden changes in frequency can be seen in FFT by maximum peaks. The numbers of maximum peaks are analyzed with its amplitude using FFT. This each peak is considered as a gene. The result of FFT is analyzed and number of peaks are approximately chosen as 20. The amplitude and its value in the FFT obtained are the parameters utilized by the FIS to spot the acute location in the speech for improvisation by speech pathologist.

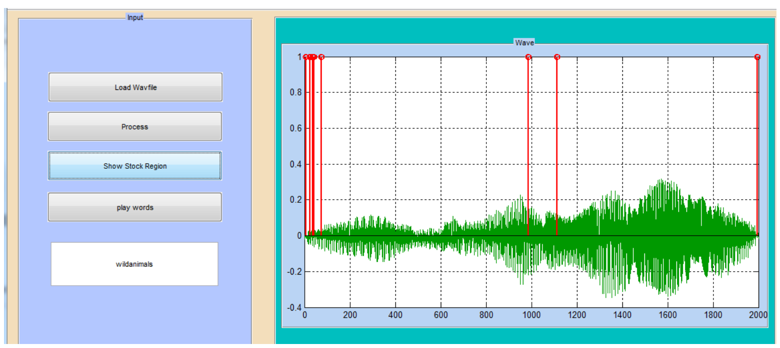
3.3.1 Fuzzy Inference system (FIS) to spot the aberration spot

The fuzzy inference system is comprised of 3 phases that are Fuzzification, Rules Evaluation and Defuzzification. The process of fuzzy inference involves Membership Functions, Logical Operations, and If-Then Rules. In this FIS, we utilize the peaks of the abnormal words in order to identify the aberration spot. These peak values are utilized for the fuzzy system (Figure 8).

3.3.2 Fuzzification

In the fuzzification process, the crispy quantities are changed into fuzzy in order to accomplish the tasks. In our proposed method, the peaks of each abnormal word are utilized for identifying the apt aberration spot. Here the following equation indicates

Figure 8: Signal for the Identified acute spot of aberration for the word 'Wildanimals' [Abnormal child]



that the general peaks of an abnormal word.

$$A = \{p_1, p_2, p_3 \cdots p_{N_p - 1}\} \quad (7)$$

For the fuzzification process we utilize the normal words' peak value in order to identify the divergence of both peaks if the peak may have high divergence when compared to the threshold then it may marked in the signal of that abnormal word.

Compare the peaks of Normal words' and
  abnormal words' peak values i.e. p_d
If p_d value is greater than the threshold then
  The amp value is selected
Else the amplitude value of the abnormal
  Word is deselected.

3.3.3 Defuzzification

In the defuzzification process, the fuzzy set is utilized as an input and output is obtained as a single number. To the extent that fuzziness supports the rule evaluation during the intermediate steps, and the final output for every variable is usually a single number. Here the single number value indicates either the amplitude is selected to mark or unmark it to indicate the aberration spot. The FIS is trained with the fuzzy rules and the testing process is performed by the number of testing words.

4. Experimental Results

4.1 Feature Extraction

Initially, the words are extracted from the both normal and abnormal children and then the MFCC feature has been extracted from it. Subsequently, the PCA is applied to reduce the dimensionality of the words. Few speech samples which are sent as input to MFCC Feature Extraction (Bharathi and Shanthi, 2012).

4.2 Classification using SVM

Subsequently, for MFCC features the PCA is applied to reduce the dimensionality of the words and then they are inputted to the SVM to identify the abnormal and the normal word. The result is shown as,

4.3 Finding spot using FIS

In this work, the proposed technique is tested with the database of 100 words with normal children 2 and an abnormal child data. Then the acute spot of aberration for the abnormal words is identified using FIS.

5. Conclusions and Future work

The proposed system was implemented in the working platform of **MATLAB (version 7.11).** In this with the aid of the **Free Audio Editor,** the dataset with the normal and abnormal female children within the age limit 6-10 yrs are generated. For 100 normal data (words) 2 female children were utilized each and for 100 abnormal data, a female child was utilized for the system and their normal frequency range is from 20 – 4 kHz. The speech input is recorded at a sampling rate of 44.1 kHz. This work is an effective system to identify the acute spot in abnormal word (MR Child's speech), where the speech has to be improved. Such that this information helps the Speech pathologist to give more practice in that region of word for that child directly. For this initially the MFCC is obtained from both the normal and abnormal words and PCA dimensionality reduction is done. Then classification of normal and abnormal words is done using SVM. After that, acute Spotting aberration in speech of pathological subject is done using FFT and FIS. The result of this work was discussed with outputs.

6. References

1• Bharathi CR., Shanthi V, (2011). Classification of speech for Clinical Data using Artificial Neural Network. IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1, November 2011 ISSN (Online): 1694-0814

2• Bharathi CR., Shanthi V, (2012). Disorder Speech Clustering For Clinical Data Using Fuzzy C-Means Clustering and Comparison With SVM classification. Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3, No.5

3• Bharathi, CR., Shanthi V, (2012). Discriminant Analysis of Disorder Speech For Clinical Data. European Journal of Scientific Research, Volume 90.

4• Bharathi CR., Shanthi V, (2012). Disorder Speech Classification for Clinical Data using SVM", National Conference, IETE.

5• Bharathi CR., Shanthi V, (2011). Finding acute Peaks and Amplitudes of Speech for Clinical Data using FFT. ICACM, International conference,

6• Elseiver – 2011

7• Bharathi CR., Shanthi, V, (2011). MFCC Feature Extraction Algorithm for Clinical Data. NCCCES'11, pp. 103-106.

8• Bharathi CR., Shanthi V, (2011). Feature Extraction using MFCC and Survey on Classification Algorithms for Clinical Data. International Conference on Computer Science Engineering CSE – 2011

9• Bharathi CR., Shanthi V, (2012). Survey on Objective Assessment of Stuttered Speech Signal for Disabled Children. International Conference on Cloud Computing and eGovernance.

10• Bharathi CR., Shanthi V, (2012). An Effective System for Acute Spotting Aberration in the Speech of Abnormal Children Via Artificial Neural Network and Genetic Algorithm. American Journal of Applied Sciences 9 (10): 1561-1570.

11• Sven Nordholm, Thushara Abhayapala, Simon Doclo, Sharon Gannot, Patrick Naylor and Ivan Tashev, (2010). Microphone Array Speech Processing. EURASIP Journal on Advances in Signal Processing. pp. 1-3, 2010

12• Marius Crisan, (2007). Chaos and Natural Language Processing. Acta Polytechnica Hungarica, Vol. 4, No. 3, pp. 61-74.

13• Rashad, Hazem M. El-Bakry and Islam R. Ismail (2010). Diphone Speech Synthesis System for Arabic Using MARY TTS. International journal of computer science & information Technology (IJCSIT), Vol. 2, No. 4, pp. 18-26.

14• Stelzle, Ugrinovic, Knipfer, Bocklet, Noth, Schuster, Eitner, Seiss and Nkenke, (2010). Automatic, computer-based speech assessment on edentulous patients with and without complete dentures - preliminary results. Journal of Oral Rehabilitation, Vol.37, No. 3, pp. 209-216.