Analysis of Speech and its Fluency Disorders

Monica Mundada
Department of Computer Science & IT, Dr.
Babasaheb Ambedkar Marathwada University,
Aurangabad, Maharashtra, India

Sangramsing N. Kayte
Department of Computer Science & IT, Dr.
Babasaheb Ambedkar Marathwada University,
Aurangabad, Maharashtra, India

ABSTRACT

The oldest way of communication and information exchange between human beings is speech. It is the most prominent and natural form of communication between humans. Speech has potential of being important mode of interaction with computer. The communication among human and computer is called human computer interaction. Man machine interface have always proved a challenging area in natural language processing and in speech recognition research. The primary aim of the study is to distinguish the difference between normal and abnormal speech. It aims to prepare a layout of factors for different types of speech and voices and its core production system. The speech disorders like misarticulation and stammering are studied with its features and associated existing therapies. The distinction is experimented with various techniques such as MFCC, DTW and K-means. The mean value of MFCC coefficients are explored for speech disorders and normal speech samples. The Euclidean distance measure is calculated between normal and abnormal speech and the variations lying in them are studied further. t-test is also experimented to justify the classification result.

Keywords
MFCC, DTW, K-Clustering

1. INTRODUCTION
Speech is a basic form of communication. Speech is the only bio metric area of research where the physical presence is not required. So thus, speech gave rise to new branch of research i.e machine learning. As speech is an integral part of communication, people suffering from various speech disorders are considered as barriers in growth of society. The proposed work will help to make the life of such people with the assertive devices and technology. This dissertation helps to automate the process of training session required in speech disorders. It also aims to prepare note of factors for different types of speech and voices and its production system. It formulates to studies the factors responsible for speech disorders and to design new observation for improvement of abnormal child speech at early stages. Current techniques, helps Speech–language pathologist to examine the subject manually. This work will benefit to build an expert system that will automate the process of speech and sound observations. The study also involves the similarities and differences associated with the normal and abnormal speech.

2. LITERATURE REVIEW
Modern speech production research [1] studies the interrelationships among various stages as the acoustic signal, the articulator shapes generating it and the muscle contraction controlling the movement of articulators. The differences [2] between speech perception, which includes the processes underlying comprehension of linguistic content of speech, and speaker perception, which includes effects of variability in speech that is not linguistically significant. In 1999, Fitch and Jay Giedd [3] studied the morphology and development of human vocal tract using MRI. It studied 129 normal human vocal tract, aged 2-25 years and observed that there is clear difference between male and female vocal tract and also the differences vary due to body i.e. height and weight. Masaaki Honda [4] presented the speech production mechanism and studied models for achieving speech functions that will enable a computer a mimic a voice that it hears. Pierre-Yves Oudeyer[5] [6], a technique which allows a control both age of synthetic voice and quantity of emotions that are expressed using mechanism on robot. The hierarchical linear modeling [7][8] the experiments are performed for both speech and vocabulary. The outcomes displayed that when children receive their implant before the age of 2.5 yrs result in the early bursts of growth in consonant-production accuracy and vocabulary and also had significantly stronger outcomes compared with age peers who had received their implants at later stages. PET (positron emission tomography) [9] and functional MRI (magnetic resonance imaging) tools are designed to investigate the brain activity of stuttering people. Researchers are studying the volunteer's patient's specific speech patterns and are instructed to avoid those patterns while speaking. The fuzzy C means clustering algorithm is also studied for fluency disorders [10]. The fuzzy clustering technique is based on minimization of objective function [11][12][13][14].

3. IMPLEMENTATION OF TECHNIQUES FOR PROPOSED WORK
The technical production of speech sound is explained in brief. The speech analysis techniques is divided into 3 types i.e Segmentation analysis, sub segmental analysis, supra segmental analysis on basis of their frame size. The feature extraction techniques like MFCC, DTW are implemented. Classification technique is experimented to classify the normal and abnormal speech. T-test is used to justify the classified the result [11][12][13][14].

4. DATABASE DESIGN
The database is designed with respect to speech disorder person suffering from stammering and misarticulation. For accuracy in the speech recognition, we need a collection of utterances, which are required for training and testing. The collection of utterances in proper manner is called the database. The list of 10 sentences are prepared with concerned for the stammering patient and the vocabulary of misarticulation includes 20 words. The database is specially designed with the help of speech therapist with the reference to stammering and misarticulation subject. Each sentence or word is uttered is for 3 times so this makes a total of 150

utterance for stammering patient and the database of same size with similar vocabulary is built up for normal subject.

5. EXPERIMENTAL ANALYSIS
5.1 MFCC (Mel-frequency Cepstral Coefficient)

MFCC is the most robust feature extraction technique as it is based on known variation of the human ear's critical bandwidth with frequency. The extraction of features in, this, phase, is, an important task as it affects its behavior in next phase. 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 .Mel frequency scale is used to study the phonetically characteristics of speech signal. MFCC is a spectral analysis method. The Mel-frequency Cepstral Coefficient (MFCC) is coefficient that is collectively made up the Mel Scale. The feature is derived from type of cepstral representation of speech signal. The effective difference between the cepstrum and Mel Frequency cepstrum is the MFC, the frequency band are equally spaced on Mel scale. The frequency warping of feature extraction is used for better representation of speech signal. The basic steps required for MFCC feature extraction is described in figure 5.1 [11][12][13][14].
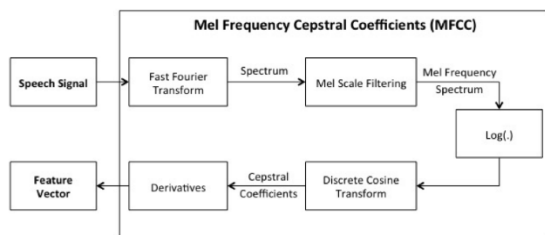

Figure 1: The basic steps of feature extraction using MFCC

5.2 Dynamic Time Warp

Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. For instance the similarities in walking patterns would be detected. DTW has been applied to video, audio, and graphics indeed, any data which can be turned into a linear representation can be analyzed with DTW. A well known application has been automatic speech recognition, which works with different speaking speeds. In general, DTW is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. This sequence alignment method is often used in the context of hidden Markov models. Dynamic time warping (DTW) is such a typical approach for a template based approach matching for speech recognition. Template with closest match defined in manner chosen as recognized word. To absorbed acoustic variations; statistical methods can be integrated into DTW approaches. DTW is quite efficient for isolated word recognition and can be adapted to connected word recognition [11][12][13][14].

5.3 K-Means

K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating globular clusters. The K-Means method is numerical, unsupervised, non-deterministic and iterative. There is always at least one item in each cluster. The clusters are non-hierarchical and they do not overlap. Every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the center' of clusters.

5.4 T-test

This is a test for comparing the means of two samples. It compares the actual difference between two means relation to the variation in the data

6. EXPERIMENTAL ANALYSIS
6.1 MFCC

Table 1: Distance matrix measure of Normal VS Abnormal Speech using MFCC

A\N	S1	S2	S3	S4	S5
S1	0.68	0.459	0.637	0.611	0.382
S2	0.871	0.65	0.828	0.802	0.573
S3	0.421	0.2	0.378	0.352	0.123
S4	0.999	0.778	0.956	0.93	0.701
S5	0.68	0.459	0.637	0.611	0.382

6.2 DTW

Table 2: Distance matrix measure of Normal VS Abnormal Speech using DTW

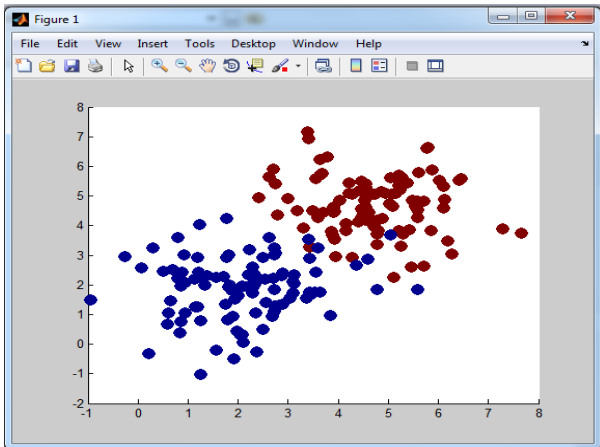| S1	S2	S3	S4	S5
197.71	387.96	160.88	377.58	116.88
180.10	200.77	146.05	346.99	94.12
184.79	325.87	132.14	216.19	98.67
207.32	413.09	206.10	394.47	142.60
184.69	325.22	265.03	307.21	144.16
279.95	373.17	173.79	303.98	183.05
251.30	327.69	206.33	311.45	155.63
137.45	174.91	129.12	304.45	107.06
190.30	228.77	196.17	302.26	104.23
227.29	377.84	193.18	301.45	103.34

6.3 K-means

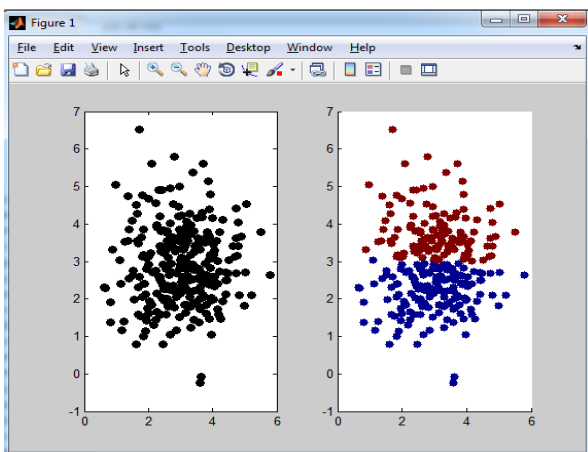**Figure 2. K-means for Normal and Abnormal speech**

Figure 3. K-means result for classification rate of 93%

6.4 T-test

Table3. T-test result of all subjects with their respective recognition rate

Subject	Gender	t-test	Recognition rate
S1	M	0.06	93%
S2	F	0.03	92%
S3	M	0.07	95%
S4	M	0.05	94%
S5	M	0.02	92%

T-test is carried out to justify the result which shows that the recognition rate is above 90% for the developed system [11][12][13][14].

7. RESULTS AND CONCLUSION

The distance matrix measure of normal and abnormal speech using MFCC and DTW with the recognition rate of 88 % and 75 % respectively. K-means classifier is experimented for normal and abnormal speech samples with classification rate of 93%. T-test is even carried out justify the result.

8. REFERENCES

[1] Van Riper, C." *Speech correction: Principles and methods".* 6th ed. Englewood Cliffs, N.J.Prentice-Hall.1978

[2] Van Riper, C. & Irwin, J.V. *"Voice and articulation".* London: Pitman Medical Publishing Company.1959.

[3] Katherine S Harris, "Physiological aspects of speech production", May 1976.

[4] Robert M. Krauss and Jennifer S. Pardo," Speaker Perception and Social Behavior: Bridging Social Psychology and Speech Science".

[5] W. Tecumseh Fitch and Jay Giedd. "Morphology and development of the human vocal tract: A study using magnetic resonance imaging". Volume 106, Issue 3. 1999.

[6] Masaaki Honda." Human Speech Production Mechanisms ".Vol. 1 No. 2 May 2003.

[7] Pierre-Yves Oudeyer. "The production and recognition of emotions in speech: Features and algorithms". Int.J Human-Computer Studies. 2003.

[8] Connor, Carol McDonald; Craig, Holly K.; Raudenbush, Stephen W.; Heavner, Krista; Zwolan, Teresa A.." The Age at Which Young Deaf Children Receive Cochlear Implants and Their Vocabulary and Speech-Production Growth: Is There an Added Value for Early Implantation?" Ear & Hearing: - Volume 27 - Issue 6 - pp 628-644 December 2006.

[9] Carolyn Asbury." Brain Imaging Technologies and Their Applications in Neuroscience". http://www.dana.org/uploadedFiles/News_and_Publicati ons/Special_Publications/brainimagingtechnologies.pdf.

[10] C.R.Bharathi , Dr.V.Shanthi." Disorder speech clustering for clinical data using Fuzzy C-Means clustering and Comparison with SVM Classification". Indian Journal of Computer Science and Engineering. Vol. 3 No.5 Oct-Nov-2012. ISSN: 0976-5166.

[11] Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014.

[12] Monica Mundada, Sangramsing Kayte "Classification of speech and its related fluency disorders Using KNN" ISSN2231-0096 Volume-4 Number-3 Sept 2014

[13] Monica Mundada, Sangramsing Kayte, Dr. Bharti Gawali "Classification of Fluent and Dysfluent Speech Using KNN Classifier" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 9, September 2014

[14] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" International Journal of Computer Science and Information Technologies (IJCSIT)