ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 5, Issue 4, April 2016

Detection and Analysis of Stuttered Speech

Vikhyath Narayan K N
MTECH MECS
Dept. of E & IE, DSCE
Bangalore

S P Meharunnisa
Assistant Professor
Dept. of E & IE, DSCE
Bangalore

Abstract*—Stuttering is a speech disorder normally faced by human being, it is also called as stammering. It involves dysfluencies or disruptions in speech. The observable signs of stuttering are repetition of syllable or word, interjection, prolongation, silent pauses, broken words and incomplete phrases. Repetition of syllable, interjection and prolongation are the main factor for the detection of stuttered speech. This paper presents a new approach for automatic detection of stuttered speech signal and classification of dysfluent and fluent speech using Mel-frequency cepstral coefficient (MFCC). The objective of paper is to classify the above mentioned dysfluency using Mel-frequency cepstral coefficients (MFCC) feature extraction and support vector machine (SVM) classification method. Classifier such as support vector machine applied on MFCC feature set to classify dysfluent and fluent speech. The SVM classifier yielded an accuracy of 90% and 96.67% for dysfluent and fluent speech respectively.

Keywords*—Mel-frequency cepstral coefficients (MFCC), Support vector machine (SVM)

I. INTRODUCTION

Stuttering is also known as stammering. It is a speech dysfluency that affects the continuity of speech. It is one of the major problems in speech disease. Approximately around 1% of the total population suffering from this dysfluency and has found to four times more in females as compared to males. Stuttering is the subject of interest to researchers from various domains like speech physiology, pathology, psychology, acoustics and signal analysis. Therefore, this area is a multidisciplinary research field of science. The speech fluency can be defined in terms of continuity, rate, co-articulation and effort. Continuity relates to the degree to which syllables and words are logically sequenced and also the Presence or absence of pauses. If semantic units follow one another in a continual and logical flow of information, the speech is interpreted as fluent. If there is a break in the smooth, meaningful flow of speech, then it is dysfluent speech.

In conventional stuttering detection process, the recorded speech is taken and dysfluencies like repetitions, prolongations and injections are identified. Then the frequency of each dyfluency is counted. This detection Processes are based on the knowledge and previous experience of speech pathologist. The main drawbacks of making such assessment are time consuming, subjective, inconsistent and also poor agreement when different judges make counts on same material

The objective of this paper is to develop a method capable of detecting the dysfluency in stuttered speech. This is one of the techniques for objective detection of stuttering. This helps Speech Language Pathologists (SLP) to assess stuttering patients, planning appropriate intervention program and monitoring the prognosis during the course of treatment. Also it improves interjudge agreements about stuttered events.

Table 1: Types of dysfluencies

Repetition	Syllable repetition (The baby ate the s-s-soup)
	Whole word repetition (The baby-baby ate the soup)
	Phrase or sentence repetition (The baby-the baby ate the soup)
Prolongation	Syllable prolongation (The baaaby ate the soup)
Interjection	Common interjections are “um” and “uh” (The baby um ate the um soup)
Pauses	The baby [pause] ate the [pause] soup

II. PROPOSED WORK

In proposed system Mel-frequency cepstral coefficients (MFCC) used for feature extraction of a signal. Decision logic is used for analysis of speech stuttering. Support vector machine is a classifier used for classification of stuttered speech signal.

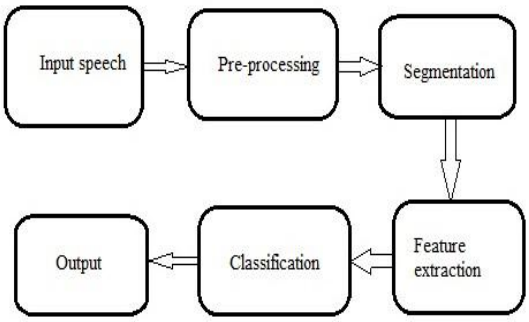

Figure 1: Schematic diagram of proposed system

1. INPUT SPEECH

The input speech was recorded speech samples of various stuttered persons which are in .wav format from UCLASS database. This .wav format samples are given as an input and analysis and classification of signal were done.

2. PRE-PROCESSING

Pre-emphasis is a very simple signal processing method which increases the amplitude of high frequency bands and decreases the amplitudes of lower bands. Pre-emphasis is not really required, it was introduced many years ago when limited computing resources forced developers to create tricky methods. It was noted that higher frequencies are more important for signal disambiguation than lower frequencies. In those days it would be easier to apply analog pre emphasis filter to get slightly better results so the pre-emphasis become popular. Another good property of pre emphasis is that it helps to deal with DC offset which is often present in recordings and thus it can improve energy-based voice activity detection. Modern speech recognition do not require pre-emphasis. Pre emphasis is compensated on later stages with channel normalization like cepstral mean normalization so it should have no effect at all. It is an artifact from a older system design. Pre-processing of speech signals is considered a crucial step in the development of a robust and efficient speech or speaker recognition system. Pre-emphasis filter is implemented in this paper to reduce noise.

3. SEGMENTATION

Syllable segmentation refers to the ability to identify the components of a word, phrase, or sentence. Identifying how many syllables are in a word or phrase (again using auditory, visual, and/or numerical representations) is a very important step in developing phonological awareness. Very young children may have a difficult time understanding the concept of a syllable, and may want to split one-syllable words like bees into two: "bee-zzz". To help these children understand the concept, I'll say a two syllable word like baseball using a "sing-song" voice with each syllable on a different note; then I'll do the same thing with bees, putting "bee" and "zzz" on different notes, which sounds kind of silly. Syllable and phoneme segmentation refers to the ability to identify the components of a word, phrase, or sentence. Identifying how many syllables are in a word or phrase (again using auditory, visual, and/or numerical representations) is a very important step in developing phonological awareness. Very young children may have a difficult time understanding the concept of a syllable, and may want to split one-syllable words like bees into two: "bee-zzz". To help these children understand the concept, I'll say a two syllable word like baseball using a "sing-song" voice with each syllable on a different note; then I'll do the same thing with bees, putting "bee" and "zzz" on different notes, which sounds kind of silly. A new algorithm to automatically segment a continuous speech signal into syllable-like segments The algorithm for segmentation is based on processing the short-term energy function of the continuous speech signal. The short-term energy function is a positive function and can therefore be processed in a manner similar to that of the magnitude spectrum. In this paper, we employ an algorithm, based on group delay processing of the magnitude spectrum to determine segment boundaries in the speech signal. Speech segmentation is the process of dividing the continuous speech into basic units having finest boundaries. It is an important step in speech recognition. It also plays an important role in certain applications. Speech segmentation can also be used for speech recognition systems.

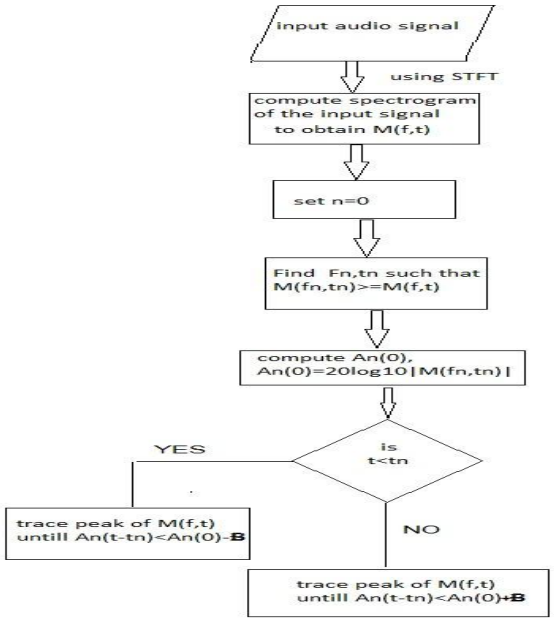

Figure 2: Flowchart for segmentation

4. MFCC FEATURE EXTRACTION

The time domain waveform of a speech signal carries all of the auditory information. From the phonological point of view, very little can be said on the basis of the waveform itself. However, past research in mathematics, acoustics, and speech technology have provided many methods for converting data, which can be considered as information if interpreted correctly. In order to find some statistically relevant information from incoming data, it is important to have mechanisms for reducing the information of each segment in the audio signal into a relatively small number of parameters, or features. These features should describe each segment in such a characteristic way that other similar segments can be grouped together by comparing their features. Coefficients are common in ASR, although 10-12 coefficients are often considered to be sufficient for coding speech (Hagen at al., 2003). The most notable downside of using MFCC is its sensitivity to noise due to its dependence on the spectral form. Methods that utilize information in the periodicity of speech signals could be used to overcome this problem, although speech also contains a periodic content. The non-linear frequency scale used an approximation to the Mel-frequency scale which is approximately linear for frequencies below 1 kHz and logarithmic for frequencies above 1 kHz. This is motivated by the fact that the human auditory system becomes less frequency-selective as frequency increases above 1 kHz. The MFCC features correspond to the cepstrum of the log filter bank energies. The extraction and selection of the best

parametric representation of acoustic signals is an important task in the design of any speech recognition system; it significantly affects the recognition performance. A compact representation would be provided by a set of mel-frequency cepstrum coefficients (MFCC), which are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale.

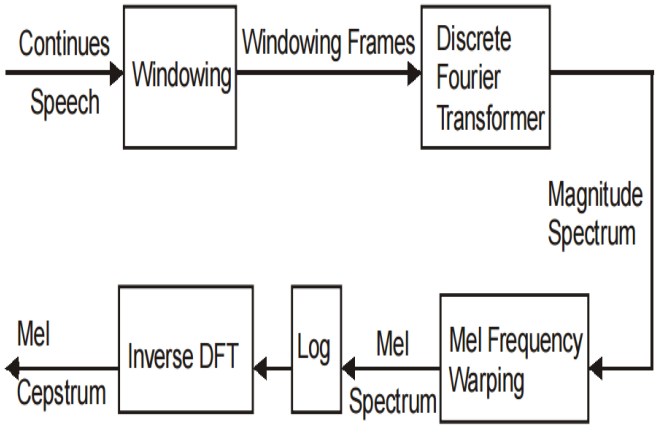

Figure 3: MFCC block diagram

5. SUPPORT VECTOR MACHINE

In machine learning, support vector machines (SVMs, also called support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it 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 falls on. In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. Kernel methods have received major attention, particularly due to the increased popularity of the Support Vector Machines. Kernel functions can be used in many applications as they provide a simple bridge from linearity to non-linearity for algorithms which can be expressed in terms of dot products. Some common Kernel functions include the linear kernel, the polynomial kernel and the Gaussian kernel.

III. RESULTS

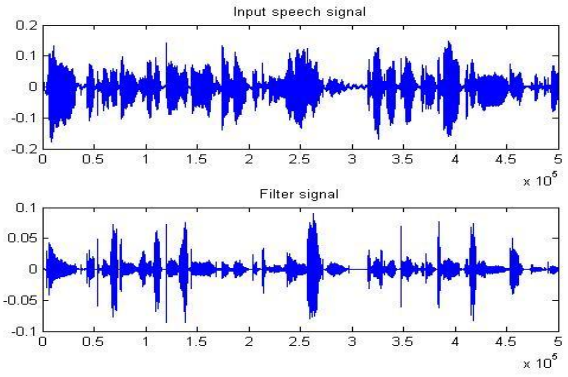

Figure 4: Pre-emphasis filter output

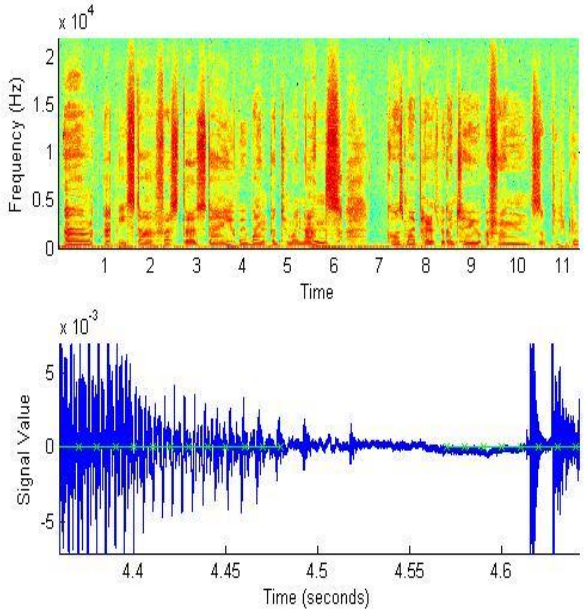

Figure 5: Segmented signal output

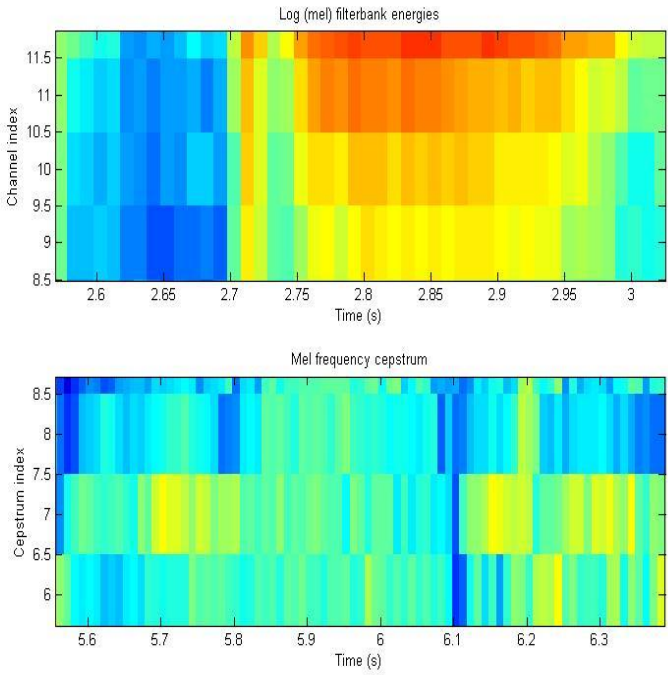

Fig. 6: MFCC output

Table 2: Stuttering classification

	Repetition	Prolongation	Interjection
UCLASS Recorded samples	24	26	15

IV. CONCLUSION

The speech signal can be used as a reliable indicator of speech abnormalities. We have proposed an approach to discriminate dysfluent and fluent speech based on MFCC feature analysis. Classifier such as support vector machine applied on MFCC feature set to classify dysfluent and fluent speech. The SVM classifier yielded an accuracy of 90% and 96.67% for dysfluent and fluent speech respectively. In this work we have considered combination of three types of dysfluencies which are important in classification of dysfluent speech.

REFERENCES

[1] Adel Belouchrani, Karim Abed-Meraim, Boualem Bosash, "Time Frequncy and Array Processsing of Non-stationary Signals", EURASIP journal 2012, 2012:230.
[2] Alfredo Maesa, Febio Garzia, Michele Scarpiniti, Roberto Cusani, "Text Independent Automatic Speech Recognition System using Mel-Frequency cepstrum Coefficient and Gaussian Mixture Models" Journal of Information security 2012, 3, 335-340.
[3] M. Adams, "Voice onsets and segment duration of normal Speakers and beginning stutters," Journal of Fluency Disorders, vol. 6, pp. 133-140, 1987.
[4] E. Yairi and B. Lewis, "Disfluencies at the onset stuttering Journal of speech & hearing research, vol. 27, pp. 154-159 1984.
[5] W. Johnson et al. "The onset of stuttering, minneapolies, University of minnesata press," 1959.
[6] K. M. Ravikumar and R. Rajagopal, "Altered Auditory Feedback Systems for Adult Stutter", Proceedings of the Sonata International Conference on Computer and control, pp. 193-196, November 2006.
[7] Neeta Awasthy, J.P.Saini and D.S. Chauhan, "Spectral Analysis of speech: A new Technique," International Journal of signal processing, vol. 2, no. 1, pp. 19-29, 2006.
[8] L. Rabiner and B.H. Juang, "Fundamental of speech Recognition,' PTR prentice Hall, 1993.
[9] Manish P. Kesarkar, "Feature extraction speech" seminar Report, Electronic systems Group, EE. Dept., IIT Bombay, November 2003.
[10] Elizabeth E Shriberg, "Phonetic Consequences of Speech Dysfluency" 1999, pp. 619-622.