Journal of Image Processing and Artificial Intelligence
e-ISSN: 2581-3803
Volume 04, Issue 02
4th National Conference on Advancements in Information Technology

Vector Quantization for Stuttered Speech Recognition

[1]K.B.Drakshayini, [2]Dr.M.A.Anusuya
[1]Assistant Professor, Department of IS&E, VVIET, Mysuru
[2]Associate Professor, Department of CS&E, SJCE, Mysuru
Email: [1]drakshakb@gmail.com, [2]anusuya_ma@yahoo.co.in

Abstract

In current trends main challenges in stuttered speech recognition are detection of stuttering fluency disorders like repetitions, prolongations, and interjections of words, silent pauses, broken words, incomplete phrases and revisions between the words. In the proposed work, silent pauses and stopgaps of stuttered speech is considered for recognition. Mel Frequency Cepstral Coefficient (MFCC) and K-Means are the anticipated methods for classification of stuttered speech with salient pause and repetitions.

Keywords: *Stuttering; Feature Extraction; classification; 12MFCCs; Vector Quantization; K-means*

INTRODUCTION

One of the active research topics for the researchers is Automatic speech recognition. By the initiation of digital computing and signal processing, the exertion of speech recognition was noticeably proposed and systematically studied. With an increased wakefulness of the advantages of everyday systems these developments were encouraged. The series of the possible applications is open and includes voice controlled applications; command control applications, dialog system, fully equipped speech-to-text software, computerization of operator assisted services for the handicapped using voice recognition [6].

Stuttered speech recognition is an interdisciplinary research filed of science because in various domains like speech physiology, pathology, psychology, acoustic and signal laboratory analysis, stuttering is the main subject of interest. Very little research work has been done on the automatic SSR (Stuttering Speech Recognition) and classification by the methods of feature extraction, statistical and acoustic analysis [7]. Different steps involved in stuttered speech recognition in the proposed system are shown in Figure 1.

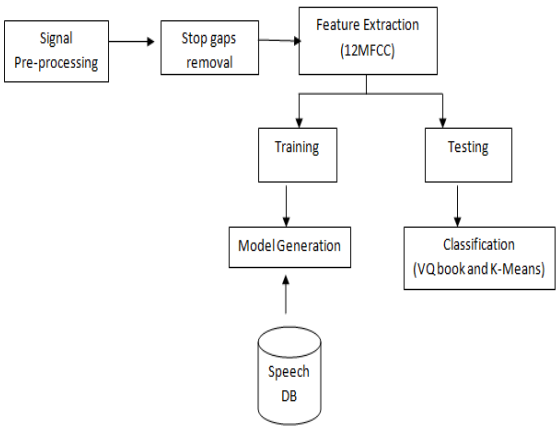

Fig 1: *Automatic speech Recognition system Architecture*

Linear Predictive Coefficients (LPC), Linear Predictive Cepstral Coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP) cepstra are experimented for the feature extraction algorithms. To convert an observed speech signal to some functional representation for supplementary investigation feature extraction method is processed. The MFCC feature extraction method is used in the proposed work. This technique is one of the top branded and most frequently processed features for stuttered and simple speech recognition applications. Usually Multidimensional feature vector for all frame of speech is produced from feature extraction package. The proposed work considered MFCC 12 coefficient [7] feature extraction method.

Organization of paper after Introduction starts with Section 2 which presents related work on stuttered speech recognition, followed by Section 3 presents experiments on Stuttered speech recognition system, and then Section 4 presents investigational Results and Discussions. Finally Future extension and Concluded concept of investigated work is presented in section 5.

RELATED WORKS

Authors of the [3], [4] practically worked on stammered speech using ANNs for classification of fluent and disfluent utterances. They analyzed 80 set of utterances each of 4-seconds lengths where half samples are with one or more disfluencies like stuttering starting with stop consonants followed by gaps and remaining were fluent. By reading a story and describing an illustration databases were recorded at different phases of patient's treatment. Each sample recording is with 22050 Hz of sample frequency, sequence of 171 vectors with Twenty one elements received from parameterization of 16 bits resolution. Each parameterization process includes FFT analysis, and then resultant is filtered using 21 digital 1/3-octave filters with 100 - 10000 Hz centre frequencies, and A-weighting filter used for weighing phase.

An additional concept was offered by Noth and collegues [1], [5]. By Considering many factors like stuttered fragments ratio, speaking rate, stammered and fluent words average length, silent pauses and filled pauses average length , they used HMMs mainly for evaluation of speech fluency. On 37 individuals of different divisions like age and gender suffering on different instability test was performed. Required database was collected by asking people to read the story and under the distinct constraints.

Automatic recognition of stuttering using both ANNs and rough set was proposed in Paper [2], and this method works on stop-gaps, vowel prolongations, detection of syllable repetitions. 6 samples of smooth speech and 6 samples with stop-gaps speech employed for experiment. Better accuracy score was obtained and it was implied in the results. More than 90% accuracy obtained using Rough set which gives the Minimum and the Maximum approximation of the original set is better than the ANNs average accuracy equal to 73.25%. Paper [8] Improvements in identification and categorizing the non-fluent and smooth speech record using Neural Network.

Paper [10] uses MFCC for Feature extraction, K-means and Vector Quantization for Classification; existing work can be further processed for quantitative analysis of stuttered speech recognition system.

DATABASES

The speech signals are collected from ordinary environment where background noise is present. These background noise signals are pre-processed using audacity software. Speech signals are acquired using Monochannel, 8000 KHz samples with 8 bit quantisation. Totally 60 signals

are considered for the tainting and testing purpose. Manual Removal of stop gaps has been done for all 50 samples (each numbers 10 times). Stuttering is eliminated by considering the fact that voiced speech has more energy than the unvoiced speech. An Adult male was asked to utter numbers (1 through 5) 10 times each for recording the microphone is used for better quality recording. Among 50 speech signals only 40 signals are selected for training. For testing 10 signals (from each number) are taken which are independent from trained set. These signals are considered for both male and female signals.

RESULTS AND DISCUSSION

Stuttered speech recognition works as follows

Pre-processing

Speech signals are pre-processed in order to augment the correctness and effectiveness of the extraction process

Stutter Removal

Speech signals are categorized into two as voiced speech and unvoiced speech. Energy and zero crossing rates are used to differentiate voiced and unvoiced signals.

Framing

Analyzing a stationary signal is simple and easy compared to continuously varying signal. The speech signal is continuously varying but from a short time point of view it is stationary, this is from the fact that glottal system cannot change immediately and research states that speech is typically stationary in the window of 20ms.

Feature Extraction

To identify a speech signal features should be matched with the previous signal or upcoming signal. To convert acoustic signal to some types of numeric representations for further analysis feature extraction is performed. Windowing, Fast Fourier Transform, and Mel Frequency Wrapping: Discrete Cosine Transform is performed as a part of feature extraction.

Vector Quantization

To translate huge amount of vector from a multidimensional space to a predefined number of clusters VQ is one of the methods. Each of clusters is defined by its central vector or centroid. According to the Euclidian distance function, K-Means algorithm clusters the data in to K groups then assigns objects to their closest cluster.

Table 1, 2 shows the KMeans Clusters before and After Removing Stopgaps (C1-C5 are 5 K clusters formed for k=5), Table 3 shows the number of signals trained and tested in experiment. Figure 2 shows the accuracy percentage for speaker dependent and independent systems.

Table1: *K-Means Clusters after removing stop gaps*

K-Means (k=5)	C1	C2	C3	C4	C5
K-Means Clusters After Removing Stopgaps	1	2	4	3	1
	1	2	4	2	1
	3	2	4	4	1
	2	2	4	2	1
	2	2	4	2	1
	5	2	4	3	1
	2	2	4	4	1
	2	2	4	1	1
	2	4	4	3	1

Table 2: *K-Means Clusters Before removing stop gaps*

K-Means (k=5)	C1	C2	C3	C4	C5
K-Means Clusters Before Removing Stopgaps	1	1	1	3	2
	3	2	5	1	1
	4	2	3	3	1
	4	2	5	2	1
	4	3	3	3	4
	1	2	5	2	1
	4	4	1	1	1
	4	1	1	1	2
	4	5	1	2	4

Table 3: *Percentage of Accuracy of SSR for different types of speech Signals.*

TYPES OF SPEECH SIGNALS	Speaker Dependent (MALE)	Speaker Independent (MALE)	Speaker Independent (FEMALE)
No of samples used for training	40	40	40
No of Samples used for testing	10	10	10
Accuracy (%)	70	20	20

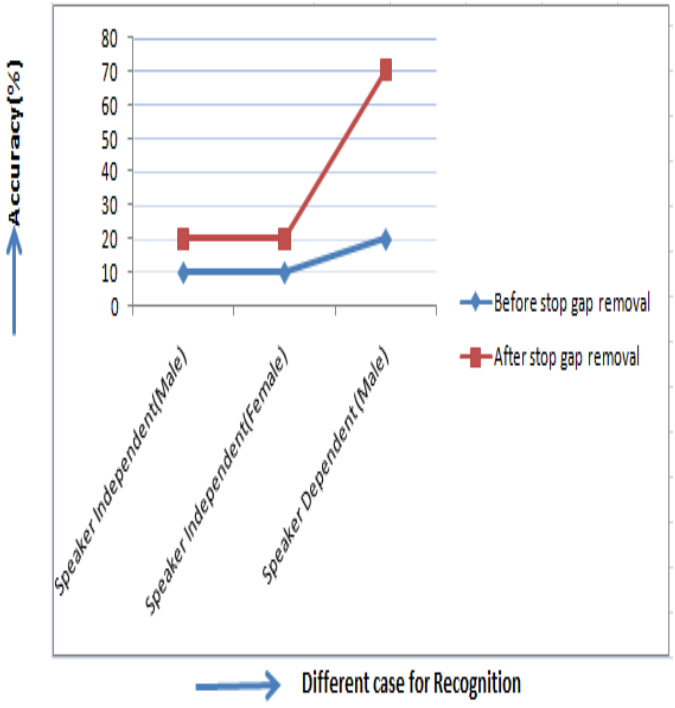

Fig 2: *Accuracy percentage of SSR for different types of speech Signals.*

CONCLUSION AND FUTURE WORK

This paper is concludes that experiment is successful in achieving 70% accuracy for speaker dependent stuttered speech recognition and 20% accuracy for speaker independent stuttered speech recognition after removal of stop gaps, which is great improvement compared to SSR without removing stop gaps. On future direction for developing technique in stuttered

speech recognition system is to achieve 100% accuracy even for speaker independent stuttered speech recognition. The work can be extended to apply both wavelet and fuzzy models for feature extraction and classification of stuttered speech recognition to get an improved version of speech feature vector which has reduced dimension with better features having more information.

REFERENCES

1. Noth E., Niemann H., Haderlein T., Decher M., Eysholdt U., Rosanowski F., Wittenberg T., "Automatic stuttering recognition using Hidden Markov Models" , Proc. Int. Conf. on Spoken Language Processing, vol. 4, Bejing, China, 2000, pp 65-68.
2. Czyzewski A, Kaczmarek A, Kostek B." Intelligent processing of stuttered speech. Journal of Intelligent Information Systems", 2003; 21(2):143–71.
3. Szczurowska I., Kuniszyk-JÓ´Zkowikak W., Smołka E., "The Application of Kohonen and Multilayer Perceptron Networks in the Speech Nonfluecy Analysis", Archives of Acoustics, vol. 31, 2006, pp 205.
4. Swietlicka I., Kuniszyk-Jó´ZKOWIKAK Zkowikak W., Smołka E., "Artificial Neural Networks in the Disabled Speech Analysis, in Computer Recognition System", 3. vol. 57/2009, Springer Berlin / Heidelberg, May 12, 2009, pp. 347-354.
5. Maier A., Haderlein T., Eysholdt U., Rosanowski F., Batliner A., Schuster M., Noth E., Peaks," A system for the automatic evaluation of voice and speech disorders, Speech Communication", 51, 2009, pp. 425-437.
6. M.A.Anusuya, S.K.Katti, "Speaker Independent Kannada Speech Recognition using Vector quantization ",2012
7. P. Mahesha, D.S. Vinod, Automatic Classification of Dysfluencies in Stuttered Speech using MFCC ", 2012
8. Szczurowska I, Jozkowiak WK, Smolka E., The application of Kohonen and Multilayer Perceptron Networks in the speech nonfluency analysis. Archives of Acoustics", 2014; 31(4)
9. Marek WI´Sniewski, Wiesława Kuniszyk-JÓ´Zkowiak,"Automatic Detection of Stuttering in a Speech", 2015
10. V.Naveen Kumar1, Y Padma Sai, C Om Prakash,"Design and Implementation of Silent Pause Stuttered Speech Recognition System", 2015.