

Stameering Speech Signal Segmentation and Classification using Machine Learning

V. Naveen¹ and Dr. S. Nagasundaram²

PG Student, Department of Computer Applications¹

Assistant Professor, Department of Computer Applications²

Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India

22304129@vistas.ac.in and naga.smec@gmail.com

Abstract: *Stuttering or Stammering is a speech defect within which sounds, syllables, or words are rehashed or delayed, disrupting the traditional flow of speech. Stuttering can make it hard to speak with other individuals, which regularly have an effect on an individual's quality of life. Automatic Speech Recognition (ASR) system is a technology that converts audio speech signal into corresponding text. Presently ASR systems play a major role in controlling or providing inputs to the various applications. Such an ASR system and Machine Translation Application suffers a lot due to stuttering (speech dysfluency). Dysfluencies will affect the phrase consciousness accuracy of an ASR, with the aid of increasing word addition, substitution and dismissal rates. In this work we focused on detecting and removing the prolongation, silent pauses and repetition to generate proper text sequence for the given stuttered speech signal. The stuttered speech recognition consists of two stages namely classification using ANN and testing in ASR. The major phases of classification system are Re-sampling, Segmentation, Pre Emphasis, Epoch Extraction and Classification. The current work is carried out in UCLASS Stuttering dataset using MATLAB with 4% to 6% increase in accuracy by ANN.*

Keywords: Stuttering, Stammering

I. INTRODUCTION

Human uses Speech to communicate, express his thoughts and feelings to other human beings. Nowadays human controls the electronic devices with help of speech through the technology called Speech recognition. Speech recognition is an easiest, most natural way of human communication to give inputs to the devices or to control it. In recent years speech recognition popular way of controlling variety of application including medical application, industrial robotics, home automation, defence, machine translation etc. Initially speech recognition system used to recognize single word or number for which you have to maintain pause between each word and number. Since lot of research carried out in speech recognition right now we can able to recognize continuous speech like conversationally paced speech. Speech recognition systems digitize, separate speech from background noise, finds the phoneme from the audio frames, compare the phoneme to predict the word and finally based on the language properties next word will be predicted in a speech sequence.

Stuttering is a speech dysfluency disorder, such as sound/syllable reiteration or prolongations in the expression of short speech components and words. Speaker comprehends what to state however is unable to state it due to an automatic tedious prolongation or end of a sound. The speech dysfluency affects the ASR in several ways. Dysfluency like repetition generates longer utterance without any proper meaning. Usually well formatted dataset are used to train the ASR system but the dysfluency produces irrelevant content which will mismatch the training and testing data that leads to poor transcription. For example consider "I we-we-went to shopping uhmmm yesterday" in such a situation it is more complex for ASR to understand the sounds like "uhmmm" so needs to make some word addition, cancellation and dismissal to increase the quality of transcription.

II. PROBLEM STATEMENT

Machine Learning is defined as an application of artificial intelligence where available information is used through algorithms to process or assist the processing of statistical data. While Machine Learning involves concepts of automation, it requires human guidance. Machine Learning involves a high level of generalization in order to get a system that performs well on yet unseen data instances. Machine learning is a relatively new discipline within Computer Science that provides a collection of data analysis techniques. Some of these techniques are based on well-established statistical methods (e.g. logistic regression and principal component analysis) while many others are not.

2.1 PURPOSE OF PROJECT

1. Image Acquisition Image

Image Acquisition is the process of collection of images. These images are downloaded from the online dataset provider called Kaggle.com.

2. Image Pre-processing

Image pre-processing includes converting colour images to grey normal images into resize images. Grayscale images have the combination of black and white. Grayscale images help to reduce noise and also make the background neutral. It also helps to improve brightness of the image. Data augmentation is a way of creating new data which has benefits like the ability to generate more data from limited data and it prevents over fitting.

3. Image Segmentation

Image segmentation breaks the image down into meaningful regions. It divides digital image into multiple segments. The goal is to simplify or change the representation into more meaningful image. It differentiates between the objects we want to inspect further and the other objects or their background. It consists of segmenting the converted grayscale images using K means segmentation.

4. Feature Extraction

Feature extraction is extracting or showing of the segmented portion of the image so that classification becomes easy. Features are extracted in order to differentiate between the images. Features extraction is used in almost all machine vision algorithms. The common goal of feature extraction and representation techniques is to convert the segmented objects into representations that better describe their main features and attributes.

5. Classification

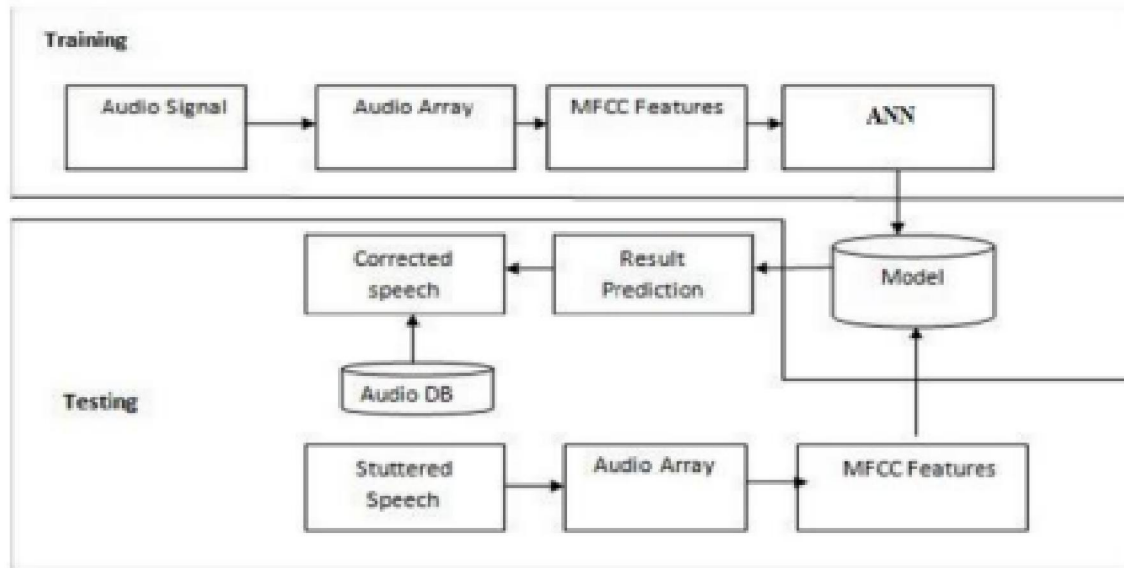
Here we use the concept of ANN for classification method. The last module includes the classification in which Tensor Flow and Machine Learning algorithm will be used. Tensor Flow is a mat lab-friendly open source library for numerical computation that makes machine learning faster and easier. Tensor Flow allows developers to create dataflow graphs - structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor.

III. PROPOSED SYSTEM

1. Read the source image into input.
2. For pre-processing step, the input image is converted to resize format from normal format.
3. GREY format dataset is analysed into red, green and blue plane which helps analysing each pixel individually.
4. Histogram is generated which helps in differentiating red, green and yellow plane from which net deterministic value for each pixel differentiation.
5. Training parameters are obtained from histogram differentiations which are integrated with intelligence.
6. Various models for various planes are generated. Then parameters of pixels obtained from histogram technique are compared with results.
7. A 3D matrix is obtained from results where each dimension refers to a particular boundary with pixel differentiation.
8. Non pixel data generated from convolution is removed then it is integrated with K means.
9. Median filter decides the maxima and minima in convolution models supplied to it as input.
10. Median filter technique integrates the similar

colour pattern on a particular pixel boundary of our convolution model. Similarly it does for other dimension and our colour pattern on that particular image is generated which is our final output

IV. SYSTEM ARCHITECTURE:



BIBLIOGRAPHY

- [1]. R.Klevansand R.Rodman, "Voice Recognition, Artech House, Boston, London 1997.
- [2]. M.A.Anusuya , S.K.Katti "Speech Recognition by Machine: A Review" International journal of computer science and Information Security 2009.
- [3]. M.A.Anusuya and S.K.Katti, "Speech Recognition by Machine: A Review", (IJCSIS) International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 181-205, 2009
- [4]. Kuldeep Kumar R. K. Aggarwal, "Hindi speech recognition system using HTK", International Journal of Computing and Business Research, vol. 2, issue 2, May 2011.
- [5]. Mohit Dua, R.K.Aggarwal, Virender Kadyan and Shelza Dua, "Punjabi Automatic Speech Recognition Using HTK", IJCSI International Journal of Computer Science Issues, vol. 9, issue 4, no. 1, July 2012.
- [6]. D. Yu and L. Deng, Automatic Speech Recognition—A Deep Learning Approach. New York, NY, USA: Springer, Oct. 2014.
- [7]. M. Hariharan, V. Vijean, C. Y. Fook, and S. Yaacob, "Speech stuttering assessment using sample entropy and Least Square Support Vector Machine," in IEEE 8th International Colloquium on Signal Processing and its Applications (CSPA), 2012, pp.240-245, 23-25 March, 2012.
- [8]. Kaushik, M., Trinkle, M., Hashemi-Sakhtsari, A. 2010. Automatic detection and removal of disfluencies from spontaneous speech. Proc. 13th Australasian Int. Conf. on Speech Science and Technology Melbourne, 98-101.
- [9]. M. Gales, S. Young, "The application of hidden Markov models in speech recognition", Found. Trends Signal Process. 1 (3) (2007), pp. 195-304
- [10]. Nöth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., et al. (2000). Automatic stuttering recognition using hidden Markov models.
- [11]. L. Helbin T. Tian-Swee and S. H. Salleh. "Application of Malay speech technology in Malay Speech Therapy Assistance Tools". In: Intelligent and Advanced Systems (2007), pp. 330-334.

- [12]. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining Knowl. Discov., vol.2, pp. 121-167, 1998.
- [13]. A. Reda, El-Khoribi, "Support Vector Machine Training of HMT Models for Land Cover Image Classification," ICGST-GVIP, vol.8, issue 4, pp. 7-11, December 2008.
- [14]. K. M. Ravikumar, R.Rajagopal, and H.C.Nagaraj, "An Approach for Objective Assessment of Stuttered Speech Using MFCC Features," ICGST International Journal on Digital Signal Processing, DSP, vol. 9, pp. 19-24, 2009.
- [15]. P. S. Savin, P. B. Ramteke and S. G. Koolagudi, "Recognition of Repetition and Prolongation in Stuttered Speech Using ANN," Proc. 3rd International Conference on Advanced Computing, Networking and Informatics, pp. 65–71, 2016.