Speech Based Anti Stuttering Algorithm using Matlab

R. Prabhu ,P. Praveen Kumar, J. Kavin kumar, Dhivyapriya E. L.

Abstract: *Stuttering is one of the most occurring speech disorders. Usual onset of stuttering is from 2 to 5 years which can be cured in childhood but for 20% of cases it prolongs to adulthood. Its severity increases when a stuttered person faces public and when he/she gets anxious. It is a long-term disorder which cannot be permanently cured. This drawback will be externally overcome with the help of designed Speech based algorithm. This algorithm is implemented in five stages namely magnitude filtering, silence ejection, speech to text, repetition removal and text to speech (TTS). The project aims at removing the repetition of words. This can be used in speech recognition systems. This system helps in encouraging the person who suffers from stuttering to give an open talk in public.*

Keywords: repetition removal, silence ejection, speech to text, text to speech, magnitude filtering.

I. INTRODUCTION

In today's world, there are different types of speech disorders like stuttering or stammering, apraxia of speech, dysarthria, lisping, cluttering, mutism, aphasia, spasmodic dysphonia etc., among which stuttering is considered here. Stammering is a genetic disorder. Some of the factors that influence stuttering are illness, traumatic life events, environmental conditions and reactions from listeners. Biologically, stuttered people have large amount of white matter in the right hemisphere and less in the left hemisphere of the brain. Statistical results show that stammering is more likely to male than female in the ratio of 3:1. Stuttering is mainly due to prolongations and repetitions. Prolongations are extension of the duration of a particular string in a stuttered speech sample. Repetitions are reoccurrence of a same word two or more times. Silent pause is an important parameter which is to be considered in stuttering if it crosses certain duration of time.

In day to day life speech plays an important role in operating equipment's/devices which are made using speech recognition systems. These systems are trained with input as speech samples using Artificial Neural Networks (ANN). These samples are taken from normal person's voice, when a stuttered person tries to access these systems, it is difficult for the system to recognize the instruction given by him/her. This paper discusses a new method for addressing these issues by converting stuttered samples to a normal speech sample which is helpful for speech recognition systems.

Revised Manuscript Received on February 05, 2020.
   R. Prabhu, Final year, Department of ECE, Sri Krishna College of Technology, Coimbatore.
   P.PraveenKumar, Final year,Department of ECE, Sri Krishna College of Technology, Coimbatore.
   J.Kavinkumar, Final year,Department of ECE, Sri Krishna College of Technology, Coimbatore.
   Dhivyapriya E.L., Assistant Professor, Department of ECE, Sri Krishna College of Technology, Coimbatore.

As mentioned above stammering is caused by prolongations, repetitions and silent pauses, different methodologies are implemented in order to overcome these issues. They are magnitude filtering, silence ejection, speech to text and text to speech.

Magnitude filtering is nothing but finding a threshold value for a particular speech sample and making the amplitude to zero when it is less than the threshold value. This method is used for removing prolongations. Silence ejection is a method to remove silent pauses in the output which are obtained from magnitude filtering. Magnitude filtering and silence ejection can be replaced by another way of approach called word splitting. This method is performed by splitting the speech sample into individual words and allocating them in different files. By finding average length of these words, the part of the string which exceeds the average word length is removed. This process is done for each and every word file.

Speech to text is done for the previous output for removing repetitions of words using python. Now the repetition of words are detected and removed. The output is converted back to speech ensuring that the voice samples are free from stuttering.

The following paper discusses about the database considered in section (II), the literature review of recent papers in section (III), algorithm for proposed model in section (IV), results in section (V) future work in section (VI), and conclusion in section (VII).

II. DATABASE

The audio samples used for the analysis is considered from a YouTube video. The video contains speech of few speakers (30) including male (19) and female (11) who are stutterers. The video is split according to the speakers. Now these videos are converted to audios (.mp3) and then they are converted to .wav form. Background music and noise are removed using audacity. Generally, there are two types of recordings based on the recorder we use. They are mono recording and dual recording, the audio sample considered here are dual recordings. Dual recordings consist of two set of amplitudes i.e., audio and stereo. The samples are extracted from the link given below
 https://www.youtube.com/watch?v=Po-WMo8vXRY

III. LITERATURE REVIEW

Many papers based on stuttered speech recognition and corrections have been proposed. The following section discusses about the proposed models briefly.

Fathima Afroz et al. have discussed about a method to

recognize and classify the pauses which occurs in stuttered speech. Identification of pauses has been done using automatic blind segmentation approach based on energy and zero crossing rate. The Intra-morphic pauses and Inter-morphic pauses are compared for stuttered and normal speech. An accuracy of 98% is obtained by this method [1].

Gunjan Jhawar et al. described an approach for recognition of speech disorder using the Mel frequency cepstral coefficient (MFCC), which is extracted from the stammered speech sample. Windowing, Pre-emphasis, DFT, Framing, Mel filter bank, DCT and Delta and energy spectrum are performed on input signal and the obtained parameters are compared with that of normal speech signal. Before training and testing of speech samples using Maximum Likelihood Classifier (ML) and Support Vector Machine (SVM), Principal Component Analysis is done. Compared to ML, SVM gives out the results more efficiently with respect to MFCC [3].

Naveen Kumar et al. proposed a method for distinguishing three events which occur in stuttered speech sample i.e., repetitions, prolongations and articulatory fixations. They have been done by developing an Adaptive Optimization based Artificial Neural Network (AOANN) approach. In this method efficiency is tested using Mel Frequency Cepstral Coefficient (MFCC) [6].

Lim Sin Chee et al. reviewed the overall recognition and classification techniques that are employed for stuttered samples. Several recognition techniques are already been proposed, this paper discusses about them briefly. Some of the Classification techniques such as Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Hidden Markov Model (HMM) and are also discussed [7].

Ankit Dash et al. developed an algorithm for detection and correction of speech disfluency. By dividing the sample into short overlapping frames, Amplitude thresholding is done using neural networks to remove prolongations. Repetitions of string are removed using existing Text to Speech (TTS) method. They had tested 50 samples out of which an accuracy of 86% is obtained [9].

Arya A Surya et al. developed a method for removing prolongation and repetitions from a stuttered speech sample which are used in automatic speech recognition system (ASR) because stammered speech is not so efficient for ASR applications. This can be done using three methods i.e., Supervised model, stuttering pruning and Automated text-to-speech [10].

Junbo Zhang et al. developed an algorithm for detecting Chinese repetitive stuttered speech sample. This is done in three phases, where multi-syllable repetitions are detected using decoding networks, errors are reduced using recursive search by including a branch penalty factor to the network. In the last phase improvement of the reliability is done for the detected stuttered speech. This improves system performance by detecting the error rate more appropriately [11].

Some of the bottlenecks of the above papers are speech recognition systems that are confined to regional languages and some of the papers fails to discuss about correction of stuttered speech. The Amplitude thresholding is done using neural networks [9] but the process is complex. Some of these issues are addressed in this paper which discusses about different methods for removal of prolongations and string repetitions.

IV. ALGORITHM FOR PROPOSED MODEL

The followings instructions explain the procedure to remove stuttering from a stuttered speech sample. The Figure 1 defines the flow mechanism of the proposed algorithm.

1. Database is extracted from the resource mentioned in section (II) and it is given as input.

2. As we have considered dual recording which consists of two set of amplitudes (audio and stereo). The input can be only audio or only stereo or both audio and stereo. The results obtained for these three are almost same, so any form of input can be considered. The further approach is based on considering audio set of amplitudes as input.

3. Magnitude filtering is making the amplitude to zero when it is less than the threshold value. This can be done in three ways i.e., taking threshold value as zero, as average of all amplitudes, third method is by considering two thresholds individually for positive and negative amplitudes.

4. Threshold values can be found with the help of MATLAB functions and magnitude filtering is done using MATLAB.

5. Silence ejection is done by removing the longer durations of zeros and replacing them with nominal duration of zeros. The duration of zeros (silent pause) between a word and its successive word in audio sample of a normal person is considered as the nominal duration of zeros.

6. After silence ejection, the output obtained in form of speech is transformed to text using a python code. It produces a set of outputs as alternatives of the given input. Out of which a best alternative is chosen and it is stored in a text file. This process is done for three methods of magnitude filtering. Comparatively, third way of approach is efficient because in second method the threshold value is near to zero due to which in first and second methods, the negative amplitudes is completely removed which causes more noise and inappropriate results in the form of text.

7. Repetitions of a word in the text are detected and removed using repetition removal code in python.

8. The output in form of text is converted back to speech using existing Text to Speech (TTS) system in MATLAB.

Fig.1.Flow chart of the model

V. RESULTS

The results obtained from each and every step is discussed here by considering one of the input speech samples. The input signal is given below.

Fig.2 Input of a stuttered speech signal

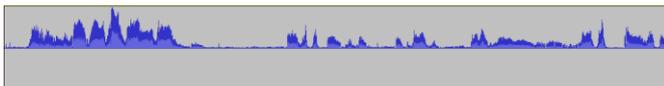The output obtained from Magnitude filtering when the threshold considered as zero is shown below

Fig.3(a) Magnitude filtered output when threshold is zero

The output obtained from Magnitude filtering when the threshold considered as average of amplitudes is shown below
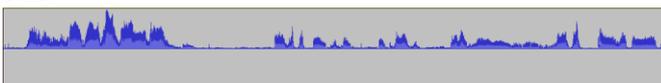

Fig.3(b) Magnitude filtered output when threshold is average of amplitudes

The output obtained from Magnitude filtering when the threshold considered as average of positive and negative amplitudes is shown below

Fig.3(c) Magnitude filtered output when threshold is average of positive and negative amplitude.

The figure below shows the output after removing silent pauses from Fig.3(a)
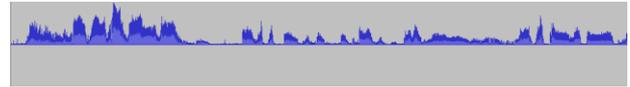

Fig. 4(a) Silence ejection output from 3(a)

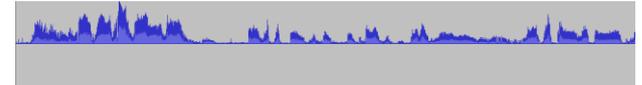The figure below shows the output after removing silent pauses from Fig.3(b)

Fig. 4(b) Silence ejection output from 3(b)

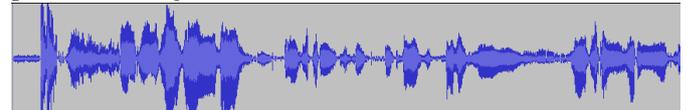The figure below shows the output after removing silent pauses from Fig.3(c)

Fig. 4(c) Silence ejection output from 3(c)

Silence ejection outputs are converted to text using Speech to Text method.

The output produced from 4(a) is "I don't know what happened".

The output produced from 4(b) is "I don't know what's happening ".

The output produced from 4(c) is "sometimes I worry about my my speech I don't know what's gonna happen next ".

From the above three outputs, it is observed that the third way of approach gives more appropriate result because in first and second outputs the text obtained is not complete because of noise due to elimination of negative amplitudes during magnitude filtering process. So, the third output is considered for the further process.

This text is modified by removing the repeated words and the generated output is "sometimes I worry about my speech I don't know what's gonna happen next ".

This modified text is converted to speech using MATLAB based on trained neural networks and the output Fig.5 waveform is given below

Fig.5 Output devoid of stuttering

By Analysing the outputs the following results are obtained,

No. of Samples considered = 30
No. of Samples corrected efficiently = 26

$$\text{Accuracy} = \frac{\text{No. of samples corrected efficiently}}{\text{No. of samples considered}} \times 100$$

Thus an accuracy of 86% is obtained from the outputs of the algorithm proposed.

VI. FUTURE WORK

The algorithm to remove prolongations and repetitions can be extended to implement in languages other than English.

The existing speech to text and text to systems which are based on neural networks are confined to limited number of datasets. By training neural networks with a greater number of datasets of different languages and by using these systems in the algorithm, it will be helpful to produce unstuttered output in various languages. Text to speech can be implemented by developing speech synthesis to produce output in original voice of the stuttered person. If the output produced in the form of original voice of the speaker, it will be helpful for them in public speaking. An application can be developed which internally implements this method and produce unstuttered output in his/her original voice within a short duration of time.

VII. CONCLUSION

The most common speech disorder is stuttering. The algorithm discussed in this paper helps to correct stuttered speech sample into a normal speech sample with an accuracy of 86%. Repetitions and prolongations are the main factors of stuttering which can be removed by implementing magnitude filtering, silence ejection, text to speech and speech to text method in this algorithm. The output obtained can be improvised in order to get the output in the originals speaker's voice which is discussed in section(V) of the paper. This can be used in speech recognition system which fails in recognizing stuttered speech. By implementing this, stutterers feel free to communicate and can concentrate more on their work without worrying about their disability.

REFERENCES

1. F. Afroz and S. G. Koolagudi, "Recognition and Classification of Pauses in Stuttered Speech Using Acoustic Features," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2019, pp. 921-926.
2. Selim S. Awad, Louis Przebienda, Richard Merson, "The Application of Digital Speech Processing to Stuttering Therapy", IEEE Instrumentation and Measurement Technology Conference, Ottawa, Canada, pp. 1361–1367, May 1997.
3. Gunjan Jhawar, Prajacta Nagraj, P. Mahalakshmi, "Speech Disorder Recognition using MFCC", International Conference on Communication and Signal Processing April 6-8, 2016, India.
4. P. Howell, S. Sackin, and K. Glenn, "Development of a Two-Stage Procedure for the Automatic Recognition of Dysfluencies in the Speech of Children Who Stutter: I. Psychometric Procedures Appropriate for Selection of Training Material for Lexical Dysfluency Classifiers," J Speech Lang Hear Res, vol. 40, pp. 1073-1084, October 1, 1997 1997.
5. L. S. Chee, O. C. Ai, M. Hariharan, and S. Yaacob, "Automatic detection of prolongations and repetitions using LPCC," 2009, pp. 1-4.
6. V. Naveen Kumar, Y. Padma Sai, C. Om Prakash, "Design and Implementation of Silent Pause Stuttered Speech Recognition System", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.4 Issue 3, March 2015.
7. Lim Sin Chee, Ooi Chia Ai, Sazali Yaacob, "Overview of Automatic Stuttering Recognition System", Proceedings of the International Conference on Man-Machine Systems (ICoMMS), 11-13 October 2009, Batu Ferringhi, Penang, Malaysia.
8. M. Hariharan, V. Vijean, C. Y. Fook and S. Yaacob, "Speech stuttering assessment using sample entropy and Least Square Support Vector Machine," 2012 IEEE 8th International Colloquium on Signal Processing and its Applications, Melaka, 2012, pp. 240-245.
9. A. Dash, N. Subramani, T. Manjunath, V. Yaragarala and S. Tripathi, "Speech Recognition and Correction of a Stuttered Speech," *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, 2018, pp. 1757-1760.
10. Arya A Surya, Surekha Mariam Varghese, "Automatic Speech Recognition System for Stuttering Disabled Persons", International Journal of Control Theory and Applications, Volume 9 Number 43, 2016.
11. J. Zhang, B. Dong and Y. Yan, "A Computer-Assist Algorithm to Detect Repetitive Stuttering Automatically," 2013 International Conference on Asian Language Processing, Urumqi, 2013, pp. 249-252.
12. Qingli Zang. The evaluation and treatment of speech-language disorders. Shijiazhuang, 1991: 185-199.
13. Bergl P, Lusty T, Cmejla R, et al. Assessment of Dysfluency In Stuttered Speech, 2010.
14. J. Au-Yeung, P. Howell, "Lexical and syntactic context and stuttering", Clinical Linguistics & Phonetics, vol. 12, no. 1, pp. 67-78, 1998.
15. O. Chia Ai and J. Yunus, "Overview of a computer-based stuttering therapy," 2006, pp. 207-211.

2155