**Speech recognition for people who stutter**

I.INTRODUCTION:

Speech is an essential mode of communication between the people in day to day life and expressing their feelings and emotions too. Stuttered/Stammering is a speech disorder that basically affects the normal flow of speech. The people who have speech disorders facing many problems in society while communicating with others .Correction of stuttered speech recognition systems have developed in recent years.These systems are helpful for people who have speech disfluencies .Billions of people are affected by the stuttering of speech and other speech related disorders across the world.Stuttering doesn’t have a cure,but there are ways to make it easier to speak.People called speech pathologists or communication experts can figure out if you stutter and help you to talk smoothly.They can teach you ways to reduce or even stop stuttering,but it’s not something that can be completely fixed. By Medical research,stuttering of speech is not hundred percent curable.These models can be helpful for the people while communicating with others. According to recent searches, stuttering is caused by many factors like genetics. Some people have deficits in the connection between the basal ganglia,thalamus and cerebral cortex.As Machine Learning models can be used for continuously monitoring the speech.These speech impairments can manifest different forms which include apraxia, cluttering, dysarthria, stuttering, etc... [10].

There are many challengings to solve this stuttering problem based on how the person speaks and how environment affects the speech.The purpose of using deep learning techniques to improve the accuracy and robustness of the system the people who stutter.There are different types of stuttering behaviors as shown in table 1.Which contains the repetitions, prolongations,Blocks or stops,Interjections[6,10,11]. Blocks/Pause refers to involuntary pause while conversation. Interjections refer to the extra words like “um,uh”. Repetitions refers to repeating of the syllable,sound,word,or phrase.Prolongations holding onto a sound for an extended period of time.

TABEL I

THE DIFFERENT CATEGORY OF STUTTERING BEHAVIOR

|  |  |
| --- | --- |
| **Stuttering Events** | **Example** |
| **B**locks/Pause | I want (involuntary pause) Apple |
| **R**epetitions | I want want Apple |
| **P**rolongations | I want Appppppplllle |
| **I**nterjections | I want um ah Apple |
| **N**ormal | I want an Apple |

Stuttering/stammering impacts approximately 1% of the world’s population(i.e 70 million), although estimated incidence ranges from 2% to 5% for sub-populations such as children and males[19].Some people stutter only during some specific situations like,while communicating within the group of people, while talking with teacher or infront of higher officers, difficulty in addressing large crowd,conversation in phone ect.... [6,4].

Previous research on stuttering prediction has used different methods like Bi-Directional LSTM(a type of neural network), Support Vector Machine(SVMs),Hidden Markov Models(HMMs)[9]. Main goal of this paper is to detect and correct the stuttered speech sample using MFCC for feature extraction and CNN for developing models. Deep Learning is the efficient way to detect easily while specialists may take longer time to detect which leads to delay in treatment. People who stutter(PWS) frequently experience bias and have difficulty while speaking with their peers.They also face difficulty in handling services and accessing speech related applications which are simple for others.[17].

II.LITERATURE SURVEY

Many of the papers published have worked on speech recognition systems using different

models, feature extraction techniques and classification methods. A brief information about existing techniques and approaches to stuttered speech recognition and correction is provided here:

In[2]Neural Network based Speech Assistance tool to enhance the fluency of adults who stutter proposed by Sharan Narasimhan,Rohini Rao.Here the authors designed a tool which resembles like mobile phone.this tool uses the technique called Fluency Enhancing Technique[FET],To speak too smoothly FET includes two keywords they are Gentle onset and slower speech.This tool is portable..In[3] “Detecting multiple speech disfluencies using a deep residual network with bidirectional long short-term memory,” proposed by Tedd Kourkounakis , Amirhossein Hajavi, and Ali Etemad .the authors proposed a deep neural network called FluentNet for stutter detection,which does not rely on language models but directly classifies a speech signals.In[4].Arjun, K. N., S. Karthik, D. Kamalnath, Pranavi Chanda, and Shikha Tripathi have discussed about automatic correction of stutter in disfluent speech.It involves repetitions, prolongations and long pauses in disfluent speech using signal processing techniques. Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coefficient (LPC) are used to extract the features.

In[6] The Author of this paper introduced Stuttering Events in Podcast which contains over 28k datasets which contains blocks,repetitions,interjection and prolongation.As they have divided SEP-28k clips into three splits containing 25k samples for training, 2k for validation and 1k for testing. As they introduced SEP-28,which consists of more annotations than the available datasets, they added new annotations to FluencyBank. In[7].S. A. Sheikh, M. Sahidullah, F. Hirsch and S. Ouni,has discussed about Stuttering Detection Using Time Delay Neural Network, Most of the existing work in this domain uses ASR models for stuttering detection but this introduces stutternet: A deep learning based stuttering detection which is capable of detecting and classifying various types of disfluencies.They have used UCLASS stuttering dataset consisting of more than 100 speakers.In[8] The author of this paper used Recurrent Neural Network Transducers(RNN-T).For evaluation they used clean data , data with disfluencies and a separate dataset with stuttered speech.A model was tested to see how well it could improve stuttering .Compared to basic configuration of this model reduced the errors in the speech by 22% for disfluencies and 16% for stuttered speech.

In[9] The author of this paper used CNN-based Automated Identification System(ASIS) which helps speech pathologist independently diagnose,classify and log fluency disorders like blocks,prolongations,word repetitions,interjection.They have used SEP-28k as datasets as build individual models for each disfluency label and measured accuracy ,precision,recall,and F1 measure.As they have build for five times, and averages of all models are taken.The main achievement of this paper is the use of a specific model based on tensorflow/keras,which is a popular framework for deep learning,to create a system for predicting stuttering. They stated that this methodology improves the quality of life of stutters. Performance accuracy for this model is 95%.In[10]The author of this paper investigated the impact of multi-task(MTL) and adversarial learning(ADV) to learn robust stutter features.This paper also used SEP-28k dataset to evaluate their system.The main contributions of this paper is to investigate the impact of applying MTL and ADV .AS they observed that ADV framework learns robust stutter features. As to address class imbalance, they used a multi branch training scheme.They have achieved 10%,6.78% and 2% improvement in repetitions, blocks and interjections respectively.

In[11]The author of this paper proposed a model based on log mel spectrogram and 2D atrous convolutional networks designed to learn spectral and temporal features.They have used two stuttering datasets i.e UCLASS and FluencyBank.As they have extracted using log mel-spectrogram feature from the signal and then its is passed to atrous convolutional network,designed to learn spectral and temporal features. In[12] A NLP-based Approach to Improve Speech Recognition Services for People with Speech Disorders proposed by Antonio Celesti, Maria Fazio, Lorenzo Carnevale, Massimo Villari they approach the solution in two phase methodology i.e using deep learning(DL) and Natural language processing (NLP).In first methodology ,the automatic speech recognition system identifies the syllables and words,this method not worked efficiently due to limited set of data set,therefore they are mainly focus on the second methodology to solve this problem.They implement the second phase in 3 steps 1] correction of mangled words 2]Semantic control 3]Nonsense word displacement.In [13].Shonibare, O., Tong, X., & Ravichandran has explained about Enhancing ASR for stuttered speech with limited data using detect and pass.They have worked on recent improvements in Automatic Speech Recognition(ASR).They have proposed a simple but effective method called “Detect and Pass” to make modern ASR systems accessible for people who stutter. They have got an accuracy in Word Error Rate(WER) from 12.8% to 71.24%.

In[16]Payal Mohapatra, Bashima Islam, Md Tamzeed Islam, Ruochen Jiao, Qi Zhu, "Efficient Stuttering Event Detection Using Siamese Networks",the authors proposed a method called DisfluentSiam this method is based on the Siamese neural network with 10M trainable parameters.In [17] The authors Krishna Basak, Vineet Sharma, Sarangh Ramesh Kv, Nilamadhab Mishra have discussed An Integrated Usage of Bidirectional LSTM Computer-based Cognitive Attention to Categorize Speech Stutters. This paper proposes a deep learning model,Bi-Directional Long-short term Memory with attention module. The mentioned model is able to classify the stuttering event by using only Mel-Spectrogram of raw audio clips.In[20]The Author of this paper proposed method is called SSR, which is based on Weighted Mel Frequency Cepstral Coefficient feature extraction algorithm and Convolutional Neural Network for classification of stuttered events.This paper mainly focused on detecting and removing the prolongation,silent pauses,and repetition to generate proper text sequence for the given stuttered speech.They have used MFCC as feature extraction and CNN as classifier.By using this technique they had got the accuracy of 89%.In[18]Liu Y, Li Y, Deng G,Et al. developed a comprehensive technique called ‘ASTER’: Automatic Speech Recognition System Accessibility Testing for Stutterers.This model will generate stuttering audio samples, applying mutators designed to mimic common stuttering patterns.It was used to evaluate the performance of commercial ASR systems on the generated audio samples. They have used four open-source ASR and three commercial ASR for evaluating the technique they developed.

References

[1]A Dash, N Subramani, T Manjunath, V Yaragarla, and S Tripathi, “Speech recognition and correction of a stuttered speech,” in ICACCI, 2018

[2].Narasimhan, Sharan, and Rohini R. Rao. "Neural Network based speech assistance tool to enhance the fluency of adults who stutter." In 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), pp. 1-5. IEEE, 2019.

[3]T Kourkounakis, A Hajavi, and A Etemad, “Detecting multiple speech disfluencies using a deep residual network with bidirectional long short-term memory,” in ICASSP. IEEE, 2020.

[4].Arjun, K. N., S. Karthik, D. Kamalnath, Pranavi Chanda, and Shikha Tripathi. "Automatic correction of stutter in disfluent speech." Procedia Computer Science 171 (2020): 1363-1370.

[5]. Tedd Kourkounakis, Amirhossein Hajavi, Ali Etemad, "FluentNet: End-to-End Detection of Stuttered Speech Disfluencies With Deep Learning", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.29, pp.2986-2999, 2021

[6].C. Lea, V. Mitra, A. Joshi, S. Kajarekar and J. P. Bigham, "SEP-28k: A Dataset for Stuttering Event Detection from Podcasts with People Who Stutter," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 6798-6802, doi: 10.1109/ICASSP39728.2021.9413520.

[7].S. A. Sheikh, M. Sahidullah, F. Hirsch and S. Ouni, "StutterNet: Stuttering Detection Using Time Delay Neural Network," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 426-430, doi: 10.23919/EUSIPCO 54536.2021.9616063.

[8].Mendelev, Valentin, Tina Raissi, Guglielmo Camporese, and Manuel Giollo. "Improved robustness to disfluencies in rnn-transducer based speech recognition." In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6878-6882. IEEE, 2021.

[9]Yash Prabhu, Naeem Seliya, "A CNN-Based Automated Stuttering Identification System", 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), pp.1601-1605, 2022.

[10]Shakeel A. Sheikh, Md Sahidullah, Fabrice Hirsch, Slim Ouni, "Robust Stuttering Detection via Multi-task and Adversarial Learning", 2022 30th European Signal Processing Conference (EUSIPCO), pp.190-194, 2022.

[11]Abedal-Kareem Al-Banna, Eran Edirisinghe, Hui Fang, "Stuttering Detection Using Atrous Convolutional Neural Networks", 2022 13th International Conference on Information and Communication Systems (ICICS), pp.252-256, 2022.

[12].A. Celesti, M. Fazio, L. Carnevale and M. Villari, "A NLP-based Approach to Improve Speech Recognition Services for People with Speech Disorders," 2022 IEEE Symposium on Computers and Communications (ISCC), Rhodes, Greece, 2022, pp. 1-6, doi: 10.1109/ISCC55528.2022.9912940.

[13].Shonibare, O., Tong, X., & Ravichandran, V. (2022). Enhancing ASR for stuttered speech with limited data using detect and pass. arXiv preprint arXiv:2202.05396.

[14].L. Barrett, J. Hu and P. Howell, "Systematic Review of Machine Learning Approaches for Detecting Developmental Stuttering," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1160-1172, 2022, doi: 10.1109/TASLP.2022.3155295.

[15].K. Murugan, N. K. Cherukuri and S. S. Donthu, "Efficient Recognition and Classification of Stuttered Word from Speech Signal using Deep Learning Technique," 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 2022, pp. 774-781, doi: 10.1109/AIC55036.2022.9848868.

[16]Payal Mohapatra, Bashima Islam, Md Tamzeed Islam, Ruochen Jiao, Qi Zhu, "Efficient Stuttering Event Detection Using Siamese Networks", ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.1-5, 2023.

[17]Krishna Basak, Vineet Sharma, Sarangh Ramesh Kv, Nilamadhab Mishra, "An Integrated Usage of Bidirectional LSTM and Computer-based Cognitive Attention to Categorize Speech Stutters", 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp.981-985, 2023.

[18].Liu Y, Li Y, Deng G, Juefei-Xu F, Du Y, Zhang C, Liu C, Li Y, Ma L, Liu Y. ASTER: Automatic Speech Recognition System Accessibility Testing for Stutterers. arXiv preprint arXiv:2308.15742. 2023 Aug 30.

[19].Lea, C., Huang, Z., Narain, J., Tooley, L., Yee, D., Tran, D. T., ... & Findlater, L. (2023, April). From User Perceptions to Technical Improvement: Enabling People Who Stutter to Better Use Speech Recognition. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[20] Pani Bhushan , Vani H Y ,D K Shivkumar, Sreeraksha M R: Stuttered Speech Recognition using Convolutional Neural Networks.NCCDS - 2021 Conference Proceedings.