European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 202

A stuttered speech self-repair system using Mel-filter and Lasso regression machine learning technique

V.Subba Ramaiah, K.Raghu, Kasiprasad Mannepalli
Assistant Professor, Department of CSE,MAHATMA GANDHI INSTITUTE OF TECHNOLOGY,Hyderabad, Telangana, India.
Email: vsubbaramaiah_cse@mgit.ac.in
Assistant Professor, Department of ECE,MAHATMA GANDHI INSTITUTE OF TECHNOLOGY,Hyderabad, Telangana, India.
Email: raghukasula@mgit.ac.in
Associate Professor, Department of ECE, koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur(Dist), Vijayawada, A.P, India.
Email: mkasiprasad@gmail.com

Abstract

In this research proposes a listeners dis-fluencies prediction system, this implementation suggests that stuttered speech identification and classification model. When the speaker produces a stuttered speech, a listener can affect the dis-fluencies of speech. In this work addresses the visual participant's stuttered words and self-repairing classification model. The participants hear the sentences spoken by a speaker who stuttered and some heard sentences are spoken by the same speaker who produces the sentences without stuttering, in this scenario stuttered speech automatically repair by proposed MFLOR machine learning model. Results simulated has more accuracy compared to previous work; in this demonstration, the listeners can hear the processed speech. This work accomplished by applying the magnitude prediction technique and filtration on the select address. The accuracy 98.99%, sensitivity 92.34% and true positive rate 0.99 have increased, this work competes with existed methods.

Keywords: Stuttered speech, Mel-filter, lasso regression

1. Introduction:

Speech processing is a significant research area in every platform; speech signal has so many interruptions like stutters, repetitions and interjections and passes. The listeners hesitate these problems; therefore, in this research stuttered speech investigation had been performed. So, many methods are implemented but stuttered speech disorders do not recover completely. Different filters like Butterworth filter, Chebyshevfilter and adaptive Gaussian filters does not give the full accuracy and recall[1-3].

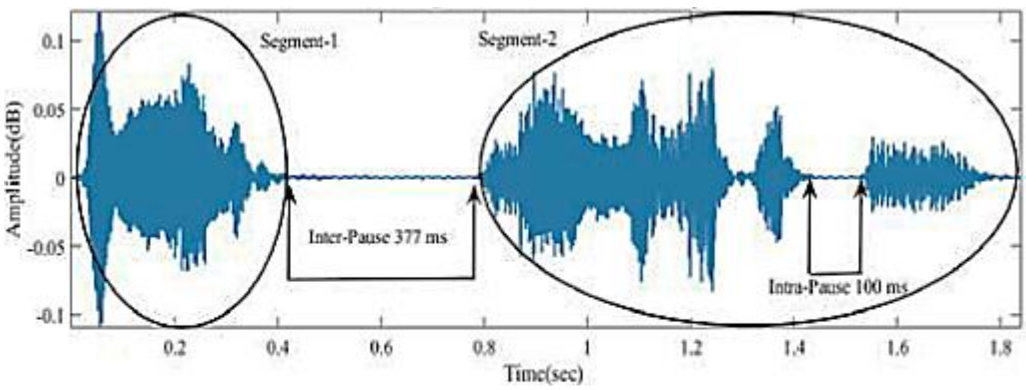

Figure 1: Speech sample

Various conventional methods with optimization techniques had developed, but improvement is needed for automation. Frequency bands and filtering are the essential key factors regarding speech analysis[3-6]. Finally stuttered speech is a significant parameter when understanding the speaker message. So, in this work develop an adaptive Mel filter based lasso-regression machine learning technique for self-repairing of a stuttered speech signal shown in fig 1.

2. Literature survey

Reference No	Technique	keypoints
[1]	Easy location disfluency elimination method	This method easily extract stuttered speech from the original speech
[2]	Partner localization method	The filtration of original speech for extracting the stutters is possible by using localization method.
[3]	Information maximization technique	The deconvolution and neural networks can separate the disfluency from speech.
[4]	Structured treatment complimentary method	This approach can enhance the stuttered speech and generate the original samples.
[5]	Speech sample filtration method	The sequence of samples are filtered and repaired, therefore generating the original speech
[6]	Role assignment approach	This mechanism can solves the issues of disfluency speech.
[7]	ERP modelling	This is an accurate segmentation and filtration process from original speech for separation of stutters
[8]	Guassianfilteration model	The combination of pre-processing and filtration gives an accurate output of speech samples.
[9]	Listener localization method	It is a combination of pre-processor and classified.
[10]	Statistical evaluation method	In this method, machine learning models are feature extracted and classify the speech samples.
[11]	PCA method	In this method, principal component analysis is used and extract stuttered free speech.
[12]	Language comprehension method	In this method, language comprehension and stuttered free samples are differentiated.

The above all literature survey briefly explains the speech processing techniques and stuttered samples analysis [15-18]. These methods do not solve the major issues of spear dependent models. So, in this research work, unsupervised machine learning models are used to design a stuttered free recognition system[19-24].

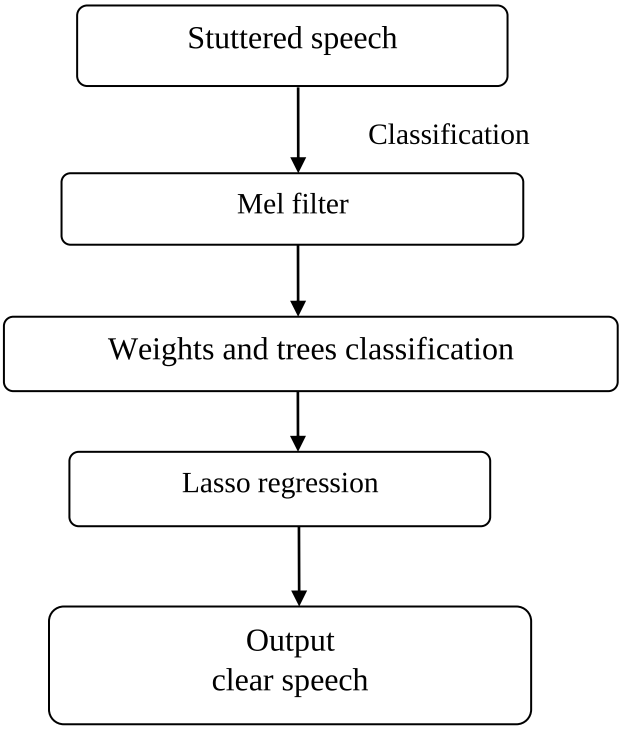

Figure 2: Block diagram of Mel-LRML

Fig 2 explains about pre-processing and classification of stuttered speech preparing system. In this, Mel-filter plays an essential role in any level of speech reorganization task. Any user can able to talk with their own voice using audacity soft computing manner. Because of this signal spectral noise is emphasized; therefore, the noise has been compensating the original speech signal. At this time, various filters are used to train the speech signal, but existed methods have more limitations. Multiple filters such as FIR, IIR and butter worth filters do not handle this type of stuttered speech noise signal. So, improvement is required; in this work mel-filter with sharp coefficients are used to train the stuttered speech and gives the error-free signal [25-28].

Mel-filter

The Mel-filter is an advanced speech processing filter,which can compute the parallel filter banks.The accurate coefficients of Mel filter can adjustits filter value by autonomus manner, and this facility is not present in any filter design.

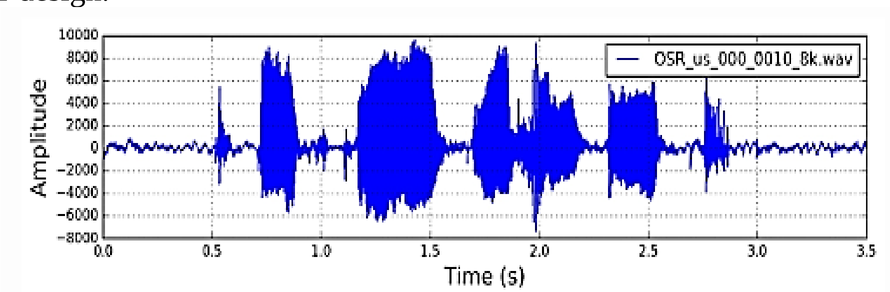

Figure 3 Time-domain signal

Fig 3 explains that raw signal in the format of .wav file, which is processed for filtering stage. The maximum amplitude of this signal is 1600mw with the time period of 3.5sec.

“importnumpy import scipy.io.wavfile from scipy.fftpack import dctsample_rate, signal = scipy.io.wavfile.read('OSR_us_000_0010_8k.wav') # File assumed to be in the same directory signal = signal[0:int(3.5 * sample_rate)] # Keep the first 3.5 seconds”

To apply the emphasis process on a selected signal for amplifying low frequencies. The emphasis process consists of following significant points. Which are illustrated below clearly.

- Frequency balancing for low magnitude frequencies.
- Fourier transforms numerical problem evaluation.
- Improve the signal to noise ratio.

“which can be easily implemented using the following line, where typical values for the filter coefficient (αα) are 0.95 or 0.97, pre_emphasis = 0.97: emphasized_signal = numpy.append(signal[0], signal[1:] - pre_emphasis * signal[:-1])”

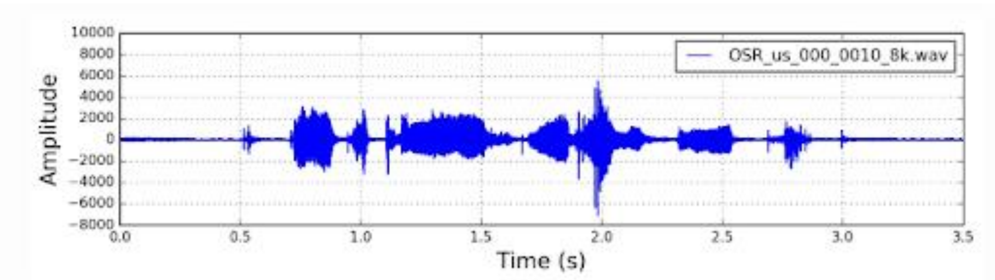

Figure 4: After free emphasis process.

Fig 4 demonstrates that Free emphasis process on real-time selected speech. This can remove the noise in the original speech and repairs the stuttered speech.

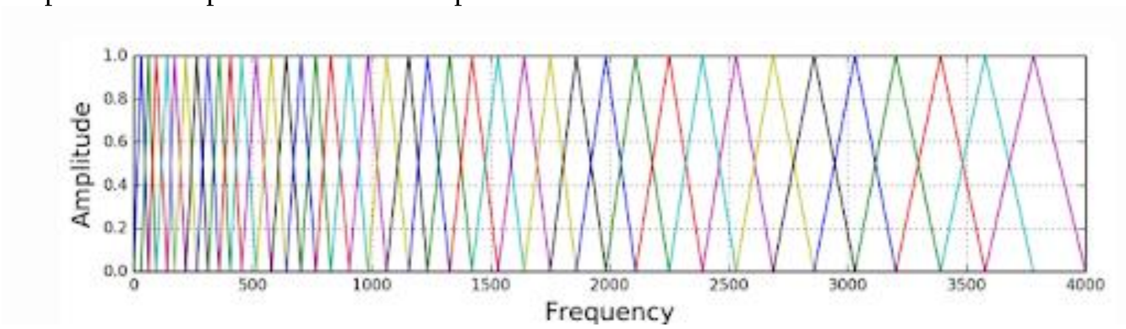

Figure 5: Mel-filter scale

Fig 5 is the explanation of mel-filter output from the extracted structure speech sample, it is clearly eliminated the stuttered speech in the signal.

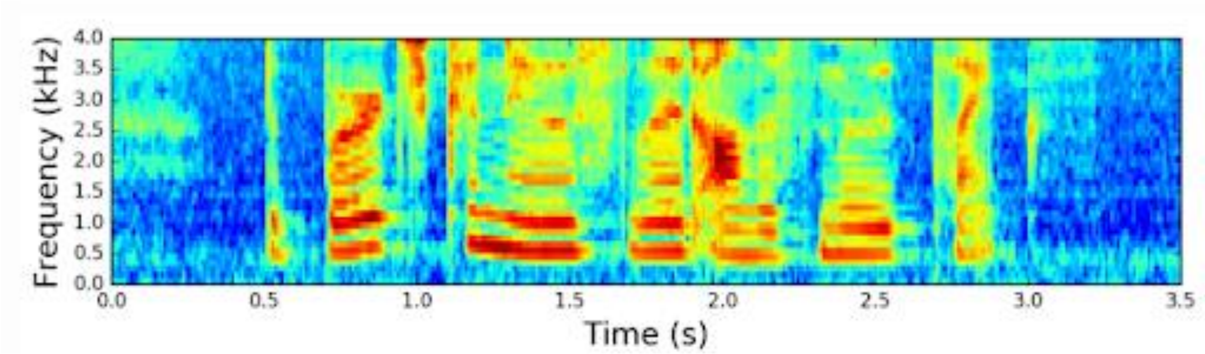

Figure 6: Normalized filter bank

After the filtration process, we can extract the refined signal with a normalized filter bank. It can easily eliminate the remaining noise factors and stuttered in the signal. Fig 6 clearly analyze the normalized filter bank with various instances of time [29-30].

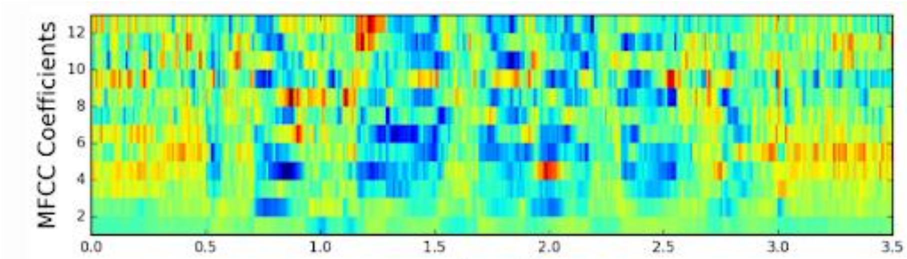

Figure 7: Normalized Mel-filter coefficients.

Fig 7 describes after the sharpening of Mel-filter coefficients with efficient utilization. The clear picture of coefficient normalization is to make the system accurate.

$$\hat{y} = X_{new}\beta_{PLS}\text{---(1)}$$

$$\beta_{PLS} = W(P^T W)^{-1}Q^T\text{---(2)}$$

$$\hat{y} = \hat{\beta}_0 + x_1\hat{\beta}_1 + \ \dots + x_p\hat{\beta}_p\text{---(3)}$$

$$\hat{\beta}_{LASSO} = arg_\beta min\left\| y - \sum_{j=1}^{p} x_j\beta_j \right\|^2\text{---(4)}$$

$$\|\beta\|_1 \leq t\text{---(5)}$$

$$\hat{\beta}_{enet} = \left(1 + \frac{\lambda_2}{n}\right)\left\{arg_\beta \min\left\|\lambda - \sum_{j=1}^{p} x_i\beta_j\right\|^2 + \lambda_1\|\beta\|_1 + \lambda_1\|\beta\|_2{}^2\right\}\text{---(6)}$$

From equation 1 to 6 are represents the lasso mathematical regression computations; these equations are used to analyze the input speech signal.

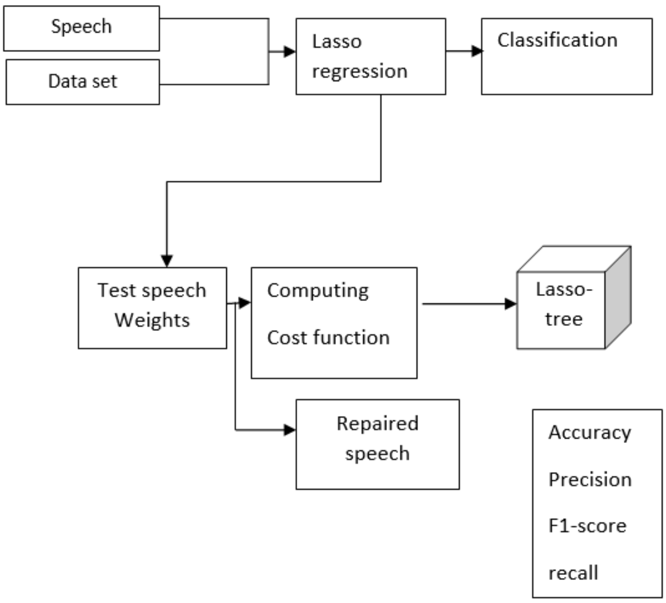

Figure 8: Lasso regression classification.

Fig 8 is a detailed explanation of the lasso regression model for classification. The input of this block diagram is dataset and real-time samples. In this work at the pre-processing stage, Mel-filter is applied, and for classification, we are applying the lasso net regression machine learning technique. At the output stage, we attain the stuttered free speech samples.

3. Information

Tool: Matlab 2015b
Methods: adaptive Mel filter
Logistic regression with the runway scheduling algorithm
Input: stuttered speech samples
Output: repaired speech
Comparison with 3- existed methods

4. Results:

The proposed method with mellfiltering gives the stuttered free speech samples from the original signal. It is an accurate method and solves the many issues of real-time applications.

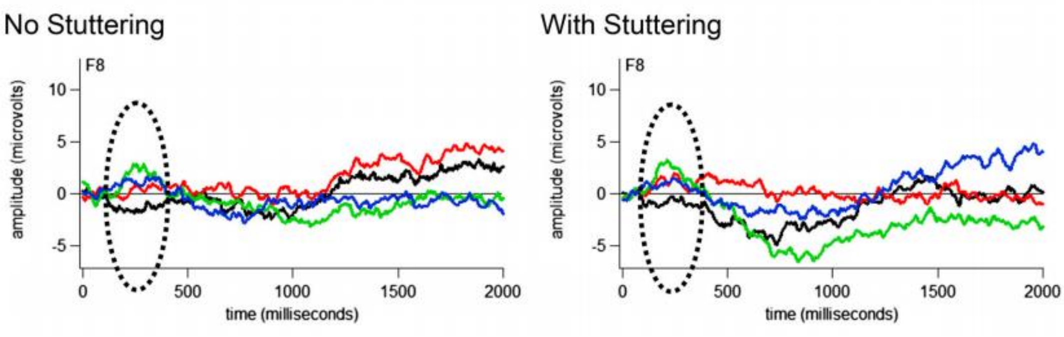

Figure 9: Stuttered free speech signal.

As of fig 9 shows that the stuttered in the speech signal entirely eliminated by the proposed Mel-filter lasso net regression model. This model simplifies the speaker-dependent applications.

Table 1: Comparison of methods

Dataset	**Threshold Value**	**SR**	**Repaired BPP**	**F1 score**	**precision**	**efficiency**
Dataset 1	0.5	3.94	2.22	49.78	0.9997	0.93
	0.1	3.89	2.53	50.12	0.972	0.912
	1.5	3.76	2.42	50.13	0.9123	0.932
	2.0	3.24	2.62	49.99	0.9712	0.911
	2.5	3.47	2.13	53.12	0.9831	0.934
Dataset 2	0.5	3.94	2.22	49.78	0.9997	0.93
	0.1	3.89	2.53	50.12	0.972	0.912
	1.5	3.76	2.42	50.13	0.9123	0.932
	2.0	3.24	2.62	49.99	0.9712	0.911
	2.5	3.47	2.13	53.12	0.9831	0.934
Dataset 3	0.5	3.94	2.22	49.78	0.9997	0.93
	0.1	3.89	2.53	50.12	0.972	0.912
	1.5	3.76	2.42	50.13	0.9123	0.932
	2.0	3.24	2.62	49.99	0.9712	0.911
	2.5	3.47	2.13	53.12	0.9831	0.934

Table 1 is the dataset which is collected from the MNIST database and real-time samples. Using these parameters calculate the accuracy, efficiency and throughput.

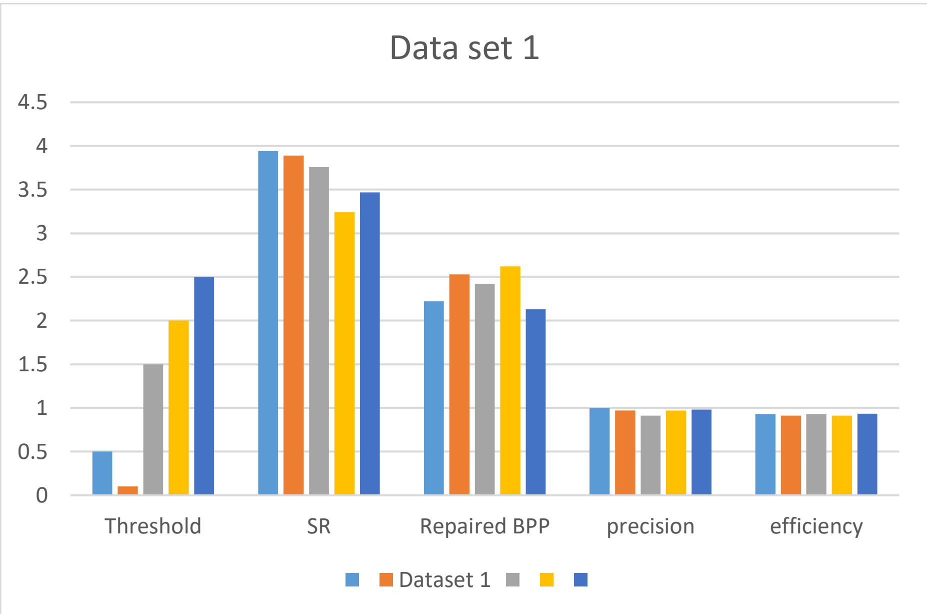

Figure 10: Dataset

Fig 10 clearly explains about the available dataset and its performance measures. In this work the proposed method achieves more improvement compared to existed methods.

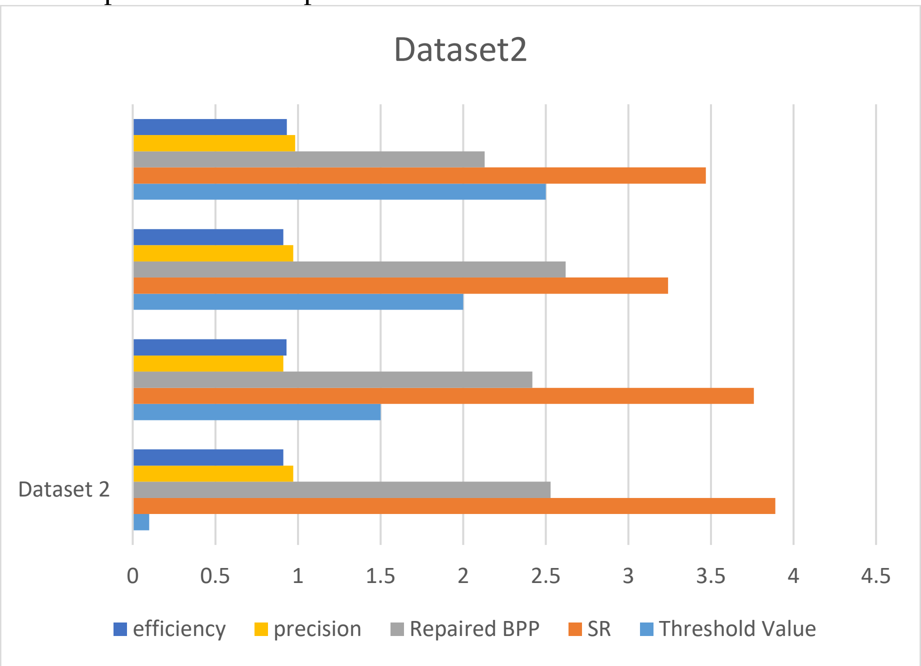

Figure 11: Dataset 2

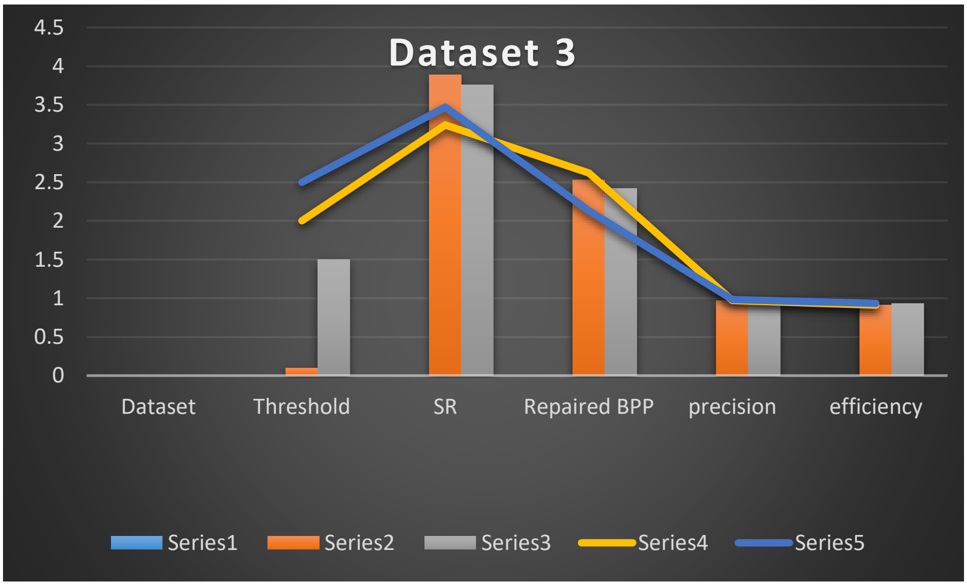

Figure 12: Dataset 3

The fig 11&12 briefly describes the performance measures of stuttered free speech signal. It is identified that the proposed method achieves and outperforms the operations effectively.

5. Conclusion:

Stuttering speech disorders like repetition prolongation, interjection and pauses are the problems of listeners who are faced by all age group people. In this research work presents a new approach for detection of the stuttered speech signal and gives the solution for the same stuttered message. The previous work with various filters provides the answer to the above problem. But accuracy and throughput are very less. Therefore in this investigation, Mel filter with spectral coefficients as a pre-processing unit and runway scheduling logistic regression algorithm as a classifier. Because of pre-processing and classification, the accuracy yields to 97.6% & 98.76% for dysfluent& fluent speech, respectively.

6. References

[1] Bailey, K. G., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. Journal of Memory and Language, 49(2), 183–200.
[2] Beilby, J. M., Byrnes, M. L., Meagher, E. L., &Yaruss, J. S. (2013). The impact of stuttering on adults who stutter and their partners. Journal of Fluency Disorders, 38(1), 14–29.
[3] Bell, A. J., &Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Computation, 7(6), 1129–1159.
[4] Blomgren, M. (2010). Stuttering treatment for adults: An update on contemporary approaches. Seminars in Speech and Language, 31(4), 272–282.
[5] Brown, S. F. (1945). The loci of stutterings in the speech sequence. Journal of Speech Disorders, 10(3), 181–192.
[6] Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. Cognitive Psychology, 42(4), 368–407.
[7] Coulson, S., & Van Petten, C. (2007).
[8] A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. Brain Research, 1146, 128– 145.
[9] Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. Journal of Fluency Disorders, 34(2), 61–71.
[10] Cyprus, S., Hezel, R. T., Rossi, D., & Adams, M. R. (1984). Effects of simulated stuttering on listener recall. Journal of Fluency Disorders, 9(3), 191–197. Dien, J. (2010).
[11] The ERP PCA toolkit: An open-source program for advanced statistical analysis of event-related potential data. Journal of Neuroscience Methods, 187(1), 138–145.
[12] Dien, J., &Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), Event-related potentials: A methods handbook (pp. 189–208).
[13] Cambridge: MIT Press. Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. Current Directions in Psychological Science, 11(1), 11–15.
[14] Friederici, A. D. (1998). Diagnosis and reanalysis: Two processing aspects the brain may differentiate. In J. Fodor& F. Ferreira (Eds.), Reanalysis in sentence processing (pp. 177– 200). Dordrecht: Springer. Friederici, A. D., Hahne, A., &Mecklinger, A. (1996).
[15] Temporal structure of syntactic parsing: Early and late event-related brain potential effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1219.
[16] arma, A. V. S. S., &Mannapalli, K. (2019). Design of a fault tolerant razor flip flop sklansky adder for delay reduction in FIR filter. International Journal of Innovative Technology and Exploring Engineering, 8(9), 2647–2651.https://doi.org/10.35940/ijitee.i8986.078919
[17] BenniloFernandes, J., Mannepalli, K. P., Saravanan, R. A., & Kumar, K. T. P. S. (2019).Fuzzy utilization in speech recognition and its different application. International Journal of Engineering and Advanced Technology, 8(5 Special Issue 3), 261–266. https://doi.org/10.35940/ijeat.E1058.0785S319
[18] Nagaram, S. K., Maloji, S., &Mannepalli, K. (2019). Misarticulated /r/ - speech corpus and automatic recognition technique. International Journal of Recent Technology and Engineering, 7(6), 172–177.
[19] Mannepalli, K., Sastry, P. N., &Suman, M. (2018). Analysis of emotion recognition system for Telugu using prosodic and formant features. In Advances in Intelligent Systems and Computing (Vol. 664, pp. 137–144). Springer Verlag. https://doi.org/10.1007/978-981-10-6626-9_15
[20] Danthala, S., Rao, S., Mannepalli, K., &Shilpa, D. (2018). Robotic manipulator control by using machine learning algorithms: A review. International Journal of Mechanical and Production

Engineering Research and Development.Transstellar Journal Publications and Research Consultancy Private Limited (TJPRC).https://doi.org/10.24247/ijmperdoct201834

[21] Mannepalli, K., Sastry, P. N., &Suman, M. (2018). Emotion recognition in speech signals using optimization based multi-SVNN classifier. Journal of King Saud University - Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2018.11.012

[22] Mannepalli, K., Maloji, S., NarahariSastry, P., Danthala, S., &Mannepalli, D. (2018). Text independent emotion recognition for Telugu speech by using prosodic features. International Journal of Engineering and Technology(UAE), 7, 594–596. https://doi.org/10.14419/ijet.v7i2.7.10887

[23] Mannepalli, K., Maloji, S., NarahariSastry, P., Danthala, S., &Mannepalli, D. (2018). Text independent emotion recognition for Telugu speech by using prosodic features. International Journal of Engineering and Technology(UAE), 7, 594–596. https://doi.org/10.14419/ijet.v7i2.7.1088

[24] Mannepalli, K., Sastry, P. N., &Suman, M. (2017).A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition. Alexandria Engineering Journal, 56(4), 485–497. https://doi.org/10.1016/j.aej.2016.09.002

[25] Mannepalli, K., Sastry, P. N., &Suman, M. (2017). Accent recognition system using deep belief networks for telugu speech signals. In Advances in Intelligent Systems and Computing (Vol. 515, pp. 99–105). Springer Verlag. https://doi.org/10.1007/978-981-10-3153-3_10

[26] Mannepalli, K., Sastry, P. N., &Suman, M. (2016). FDBN: Design and development of Fractional Deep Belief Networks for speaker emotion recognition. International Journal of Speech Technology, 19(4), 779–790. https://doi.org/10.1007/s10772-016-9368-y

[27] Mannepalli, K., Sastry, P. N., &Suman, M. (2016). MFCC-GMM based accent recognition system for Telugu speech signals. International Journal of Speech Technology, 19(1), 87–93. https://doi.org/10.1007/s10772-015-9328-y

[28] Mannepalli, K., Sastry, P. N., & Rajesh, V. (2015). Accent detection of Telugu speech using prosodic and formant features. In International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE (pp. 318–322). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SPACES.2015.7058274

[29] Mannepalli, K., NarahariSastry, P., & Rajesh, V. (2015).Accent detection of Telugu speech using supra-segmental features. International Journal of Soft Computing, 10(5), 287–292. https://doi.org/10.3923/ijscomp.2015.287.292

[30] Mannepalli, K., Sastry, P. N., & Rajesh, V. (2014).Modelling and analysis of accent based recognition and speaker identification system. ARPN Journal of Engineering and Applied Sciences, 9(12), 2807–2815.