Grenze International Journal of Computer Theory and Engineering , Special issue

Study and Analysis of Speech Disfluency for Real Time Applications

Vinay N.A[1] and Dr.Bharathi S.H[2]
[1]Research Scholar, School of Electronics and Communication REVA University
vinayna07@gmail.com
[2]Professor& P.G Coordinator, School of Electronics and Communication REVA University
bharathish@reva.edu.in

***Abstract*—Speech is an important mode of communication for humans with them as well as with machines in real time applications. In many interactive applications machines or systems has to process and recognize the speech signal as a source of data and it has to respond accordingly. This speech recognition will be handled by Automatic Speech Recognition system. These ASR will recognize the fluent speech, but if it is a disfluent speech the ASR requires more time or it may not be able to recognize the disfluent speech. Disfluency, is a speech disorder includes blockages, prolongations, repetitions, apraxia & interjections, also known as stuttered voice or stammering voice. In order to recognize the stuttered voices different features are extracted using MFCC,LPC,LPCC,DTW etc techniques and it will classified using ANN,HMM,KNN,SVM techniques.**

***Index Terms*— Blockages, Prolongations, Repetitions, Apraxia & Interjections.**

I. INTRODUCTION

Speech recognition process includes: segmentation, feature extraction and classification. The speech recognition process will get disturbed due to noise, discontinuity in speech. This discontinuity is called as disfluency of speech. For recognition of such disfluency speech, Automatic Speech Recognition (ASR) requires more time. The reason for discontinuity in speech is due to noise, less capability in processing, sometimes if a person is in stress or due to excitement.
Recognition of disfluencies in the passage is a time consuming one and also results in poor assessment. Therefore automatic recognition involves feature extraction and classification techniques. In this paper various feature extraction and classification techniques are discussed.

A. Feature Extraction Techniques

Feature extraction is to change over watched discourse flag to parametric illustration for further examination and handling. A few element extraction calculations are utilized for this assignment, for example, [7]:

a. Linear predictive analysis (LPC)
b. Mel-frequency cepstral coefficients (MFCC)
c. Perceptual linear predictive coefficients (PLP)
d. Relative spectra filtering of log domain coefficients (RASTA)
e. Zero Crossings With Peak Amplitudes (ZCPA)
f. Dynamic Time Wrapping (DTW).

Grenze ID: 01.GIJCTE.3.4.53

B. Classification Techniques

Classification techniques includes concentrate the prominent components and making them in various gatherings relying upon their list of capabilities. Order is the issue of recognizing to which of an arrangement of classes another perception has a place, on the premise of a preparation set of information containing perceptions whose classification participation is known.

The most widely used classification techniques are:

a. Artificial Neural Networks (ANN).
b. Hidden Markov Model (HMM).
c. KNN
d. Support Vector Machine (SVM)

II. RELATED WORK

In 1990s, Andrzej Czyzewski et al [5] have said that faltered discourse evaluation can be helped out through time space examination by measuring the length of dysfluencies and contrasting it and the span of the whole entry. Howell et al [6] have done research on acoustic examination of faltered discourse to know whether disfulency happens just at particular "minutes" as well as affected by the encompassing discourse. In 1991, Zebrowski et al [7] have done examination of faltered discourse utilizing (a) mean term of sound/syllable reiteration and sound prolongation, (b) mean number of rehashed units per example of sound/syllable and entire word redundancy, and (c) different related measures of the recurrence of all between and inside word discourse dysfluencies.In 2004, Michael D. Mc Clean et al [8] accomplished great outcomes utilizing engine measures. The kinematic measures that are most as often as possible analyzed are developments of lips, jaw and tongue. From their exploration, it was found that speed of development of lip, jaw, tongue was more in the stammered discourse when contrasted with that of familiar discourse.In 2005, Waldemar Suszyński et all in their study main aim stands to elaborate a database algorithm for aof different sorts of discourse disfluencies in stammering individuals' expressions, on the premise of their parameterised attributes in the abundancy recurrence space [9].In 2009 K.M Ravikumar et all[10],provide an objective assessment for the detection of syllable repetitions in an stuttered speech. The prevailing components of repetitions are: 1. Word Repetitions, however not part word Repetition is a predominant element of early stammering [11]. 2. In early stammering, there is a high extent of Repetition as a rule, rather than different sorts of disfluency like prolongation .In 2011, Aini Hafizah Mohd Saod and Dzati Athiar Ramli worked on Grouping of Apraxia Speech utilizing Support Vector Machine. Apraxia of discourse is one of discourse issue issues concentrated on discourse pathology inquire about Apraxia is a neurogenic discourse issue coming about because of incapacity of the sensor engine charges to program the development of muscles for discourse creation. It can be isolated into three principle classes, which are Apraxia among kids, stroke-related Apraxia and stress-prompted Apraxia [12] .In 2014, Monica Mundada et all focused on the identification of the distinguish properties between fluent and dysfluent speech,the proposed work classifies the fluent and dysfluent speech[13] .In 2014, G. Manjula and Dr. M. Shiva Kumar worked on the recognition of stuttered speech for robot control applications. In this work both transmitter and receiver part comprises of a Voice Recognition (VR) module. The voice order are sent to the chip utilizing a receiver and after that the module changes over the voice summon to bearing charge that is predefined and unmistakable by the robot. The recipient is the mechanized robot. It comprises of two DC engines and will make the robot move in forward or in the retrogressive heading. For the acknowledgment of faltered discourse, the test information is changed over to layouts. The acknowledgment procedure then comprises of coordinating the approaching discourse with put away layouts [13]. Inquire about work of Jing Jiang et al [15] was identified with the grouping of stammered discourse on: 1. More Typical (MT) of stammering speakers (faltering brought on because of wavering, redundancy, monosyllabic word reiteration , Part-word syllable reiteration with 2 or less without strain and 2. Less Typical (LT) of stammering speakers (Monosyllabic word reiteration part-word syllable redundancy with at least 3 reiterations prolongation, squares). The outcomes demonstrated that the LT and the MT can be ordered with high precision in view of the cerebrum movement. The cerebrum districts that made most commitment to the detachment of MT and LT stammering speakers were: the left sub-par frontal cortex and reciprocal precuneus, both of which demonstrated higher action in the MT than in the LT; and the left putamen and right cerebellum which demonstrated the inverse movement design. The outcomes likewise demonstrated that the mind movement for Whole word reiteration (WWR) was more like that of the Less Typical (LT) of stammering speakers and familiar discourse than to that of the MT. These discoveries give a neurological premise to isolating the MT and the LT sorts, and support the generally utilized MT/LT side effect gathering plan.

III. Feature Extraction techniques

A. Linear Predictive Coding (LPC)

The steps involved in LPC technique is as shown in the figure, LPC technique is frame based analysis, i.e. input speech signal is splitted into frames, which gives observation vectors. After splitting into N number of frames, these samples undergoes windowing. In windowing, finite number of digitized samples is generated and in this step speech samples is auto correlated to generate the coefficients. The order of the LPC analysis will be taken from the highest auto-correlated value. By using Durbin's method in the LP analysis, highest value of auto-correlated value is converted into LPC parameter.

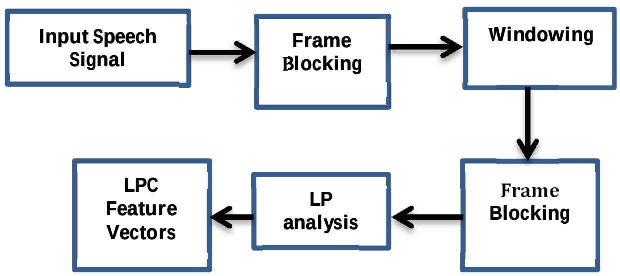

Figure 1:LPC Feature Extraction Process

Direct forecast is a numerical operation which gives an estimation of the present example of a discrete flag as a straight mix of a few past specimens. The expectation blunder i.e. the distinction between the anticipated and genuine esteem is known as the lingering. If the current sample Y_i of the audio signal be predicted by the past p samples and

$Y_i{}'$is the predicted value then we have:

$$Y_i = -a_2 Y_{i-1} - a_3 Y_{i-2} - \cdots - a_p + Y_{i-p} \text{ ---(1)}$$

Here {1,2,...,ap+1 } are the (p+1) filter coefficients. In this case the signal is passed through an LPC filter which generates a element feature vector and a scalar which represents the variance of the predicted signal

B. Mel-Frequency Cepstrum Co-Efficient (MFCC)

The Mel-Frequency Cepstrum Coefficient (MFCC) method is regularly used to make the unique finger impression of the sound documents. It depends on audible frequency range of the sound signal. To generate MFC coefficient, speech signal is splitted into frames. Let each frames comprise of N tests and let neighboring edges be isolated by M tests where M<N. Each casing is duplicated by a Hamming window where the Hamming window condition is given by:

$$W_n = 0.54 - 0.46 \cos\,(2\pi n/\mathrm{M} - 1) \text{ ------ (2)}$$

Then the signal is converted from time domain to frequency domain by using Fast Fourier transform (FFT). In FFT, number of computation will be reduced. So, the FFT of a signal is given by:

$$Y_K = \sum_{n=0}^{M-1} y_n\, e^{\frac{-j2\pi kn}{M-1}} ----------- (3)$$

The frequency domain signal is converted to Mel frequency scale, the signal in this is more audible to human beings. This is achieved by a frequency domain signal through a Mel-filter bank, which consist of set of triangular filters. These triangular shape filters generates weighted sum of signal components .The magnitude of triangular filters equal to '1' at the center of the frequency and it approaches '0' linearly. The Mel-vectors for a given frequency is calculated by

$$\mathrm{F} = 2595 log_{10}(1 + f/700) \text{----------------(4)}$$

Then these Mel scale cepstrum coefficients are converted back to time domain by applying Discrete Cosine Transform(DCT) and its given by

$$Y_K = \alpha \sum_{n=0}^{N-1} y_n \cos\left\{\frac{(2n + 1)\, \pi k}{2M}\right\} ------(5)$$

Finally, resultant signal is MFCC vectors, called as acoustic vectors.

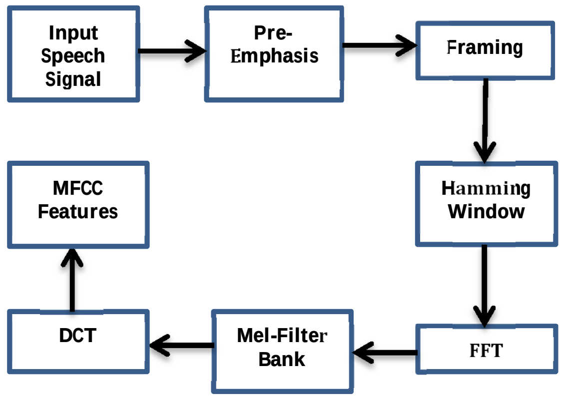

Figure 2:Block Diagram of MFCC Feature Extraction Techniques

C. Perceptual Linear Prediction (PLP)

The main aim of PLP is to determine the characteristics of human hearing accurately through feature extraction process. It is same as Linear Predictive Coding (LPC), but it is not based on long-term spectrum speech i.e. for PLP also speech signal is splitted into frames of N number of samples. Through transformations PLP changes the short frame spectrums.

The PLP parameters are generated by Bark-spaced filter bank, which consists of 18 filters to convert time domain signal to frequency domain within the range of 0 to 5KHz and to generate the PLP coefficients following steps are followed:

(i) An N-point DFT is applied for input speech signal C(n)
(ii) Discrete convolution is applied for power spectrum signal to generate critical band power spectrum.
(iii) Down-sampled signal undergoes pre-emphasis.
(iv) Compress the signal based on intensity of the loudness.
(v) IDFT is applied to get the autocorrelation functions.
(vi) Finally, PLP coefficients are generated by auto regressive process.

D. Relative Spectra Filtering (RASTA)

Speech analysis will be done in two ways one is temporal analysis and other is spectral analysis. In temporal analysis, speech signal is taken as it is for processing, but in spectrum speech is taken in the form spectrum for processing. In temporal analysis, the temporal properties are differ from speech properties and in spectrum analysis it consists of filters to strengthen the representation of speech signal. Therefore to strengthen speech recognition, RASTA is used. The block diagram representation of RASTA is shown in the figure.

The phantom estimations of info discourse flag are compacted by a nonlinear pressure run (a=2/3) preceding performing the separating operation and extended subsequent to sifting (b=3/2).

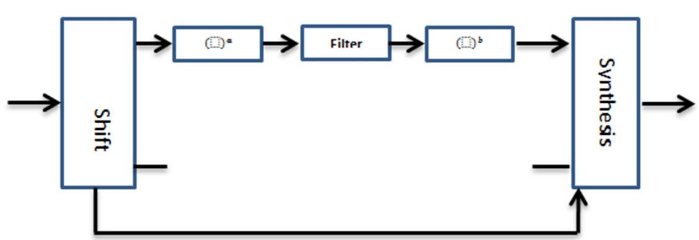

Figure 4: Block Diagram of RASTA Method

The output of each filter is given as

$$R_n(K) = \sum_{n=-N}^{N} W_n(i)\, Y_n(K-i) \text{-----------------}(6)$$

Here, $\boldsymbol{R_n(K)}$is assessment of sparkling speech in frequency box "i" and frame-index "k", $\boldsymbol{Y_n(K)}$ is noisy range, $\boldsymbol{W_n\ (i)}$are the weights of the channel and M is request of the channel.

This RASTA is implemented along with PLP to increase the recognition accuracy. The procedure for the implementation of RASTA-PLP is as follows:

(i) As in PLP, critical-band power spectrum is determined.
(ii) Static non-linear transformation is used to transform the spectral amplitude.
(iii) For each transformed spectral signal eliminate the time trajectory through filters.
(iv) Again apply static non-linear transformation to transform the filtered signal.
(v) To obtain the power law of human hearing, multiply the signal having equal loudness.

E. Zero Crossings with Peak Amplitudes (ZCPA)

This procedure depends on Human Auditory System. In this system, zero-intersection interim is utilized to speak to frequency and amplitude data, these frequency and amplitude represent intensity information. At the yield side, these frequency and amplitude data is joined to shape the entire element yield. The accompanying figure demonstrates ZCPA standard outline for feature extraction.

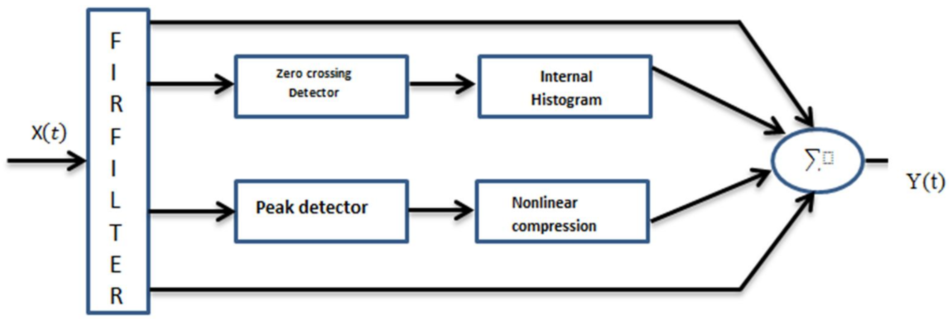

Figure 5: Block Diagram of ZCPA Method

In ZCPA FIR filter consists of 16 Band Pass Filters (BPF), which works on human audible frequency range of 200 to 4 KHz. These 16 BPF transforms original speech data to 16 different data processing paths. Then the signal is divided into N number of samples and these samples are applied to zero crossing detector for the detection zero crossing interval, from each intervals pitch of the speech is determined.

The non-linear compression of speech signal is given by the equation. The equation is a repetitiveness work; x represents a peak in the upward going zero-intersection interim. After it is packed logarithmically the outcome is Z(x).

$$Z(x) = \log(1.0 + 20\pi)\text{----------------------------------}(7)$$

F. Dynamic Time Warping (DTW)

This algorithm is based on time wrapping program, which compares the signal in time and speed to give optimized alignment between the signals. This wrapping process includes expansion or contraction of a signal in time axis is non-linear. This wrapping operation of a signal gives the information about signal similarities.

To adjust two successions utilizing DTW, a n*m lattice where the (x^{th}, y^{th}) component of the framework contains the separation T(am, bn) between the two focuses qi and cj is developed. At that point, the supreme separation between the estimations of two groupings is figured utilizing the Euclidean separation calculation as appeared in equation:

$$T\ (a_m, b_n)= 2*(a_m-b_n)\text{-------------------------------}(8)$$

Each matrix element (x, y) corresponds to the alignment between the points q_i and c_j. Then, collective distance is measured at each interval of time by:

$$T(m,n) = \min[T(m-1, n-1), S(m-1,n), S(m, n -1)] + T(m, n)\text{----------------------} (9)$$

To keep the signal properly in time axis , this algorithm maps the signal frame by frame. This is achieved by using grids, these grids reduces the total distance between the signals and this minimum distance called as Overall distance of a signal. This overall distance is calculated by passing the signals through the grids and each grid gives the overall distance. The overall distance between each individual elements called Euclidean distance. Here overall distance is measured by:

$$OD_{ab} = ED_{ab} + \min (OD_{a-1\ b-1}, OD_{a-1\ b}, OD_{x\ y-1}) \text{---}(10)$$

Here, OD = Overall distance

ED=Euclidean distance.

IV. Classification Techniques

A. Artificial Neural Network(ANN)

A mathematical model or computational model forms the ANNs ,many researchers are designing ANNs to solve problems in pattern recognition, anticipation, effective use of resources. Neural networks play an important role both in the speech[11] and the speaker recognition[12] and also it is suitable tool in distinguishing between similar signals[13-15]. In recent research, ANNs are widely used to classify continuous and discontinuous speech signal. The basic structure of ANN is shown in the figure.

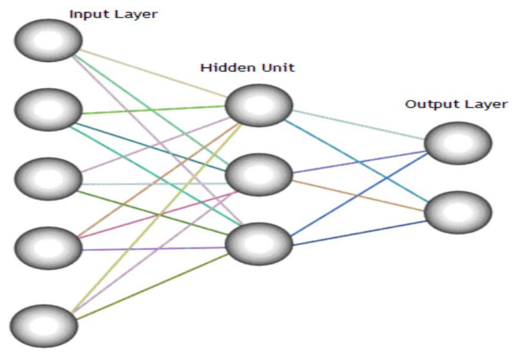

Figure 5. The basic architecture of multilayer feed forward neural network

The neural networks are trained in a such a way that, it can automatically adjusted to find the relationship between input and output signals[18].

B. Hidden Markov Mode(HMM)

A Hidden Markov Model (HMM) is a factual Markov display in which the framework being displayed is thought to be a Markov procedure with in secret (shrouded) states. Well utilize Markov procedure to show the changing measurable qualities that are just probabilistically showed through genuine perception [2]. Discourse flag is accepted as a stochastic capacity of the state succession of the Markov Chain. The state arrangements itself is covered up. There are 2 sorts of HMM separated by its perception likelihood capacities called discrete HMM (DHMM) and nonstop thickness HMM (CDHMM). Figure demonstrates the sort of HMM is utilized by them. It is a 5 state left-to-right model. The database comprises of 20 tests of ordinary discourse information and 15 tests of counterfeit falter discourse information. 10 tests of every typical and counterfeit falter discourse were utilized to produce a discourse show separately. Remaining 5 tests of ordinary discourse information and simulated stammered discourse information were utilized to test on HMM models.

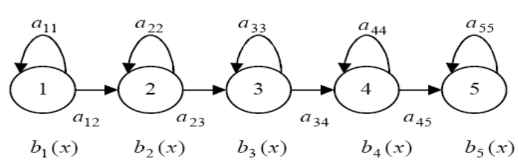

Figure 6. Representation of left-to-right HMM

C. KNN

KNN characterizes new case question in light of nearest preparing cases in the element space [7]. KNN is a kind of occasion based learning, or apathetic realizing where the capacity is just approximated locally and all calculation is postponed until the arrangement is finished. Each question protest (test discourse flag) is contrasted and each of preparing item (preparing discourse flag). At that point the protest is arranged by a lion's share vote of its neighbors with the question being allocated to the class most basic among its k closest neighbors (k is a positive whole number, regularly little). On the off chance that k = 1, then the question is essentially doled out to the class of its closest neighbor [8] [10]. The flow chart is shown in the figure.

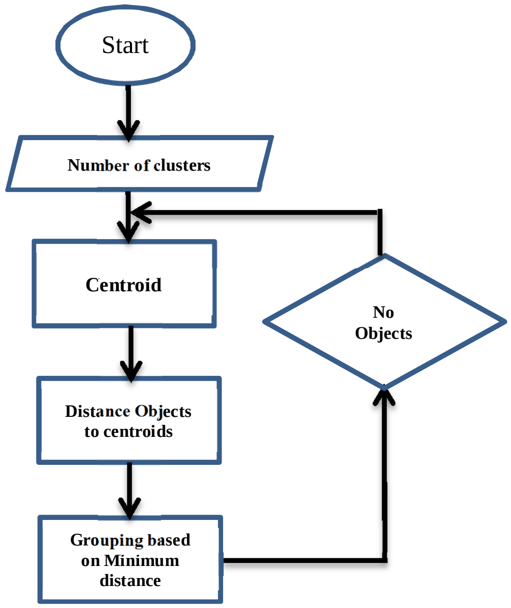

Figure 7: Flow chart of KNN Algorithm

D. Algorithm

1. Put K focuses into the space spoken to by the items that are being grouped. These focuses speak to starting gathering centroids.
2. Dole out each question the gathering that has the nearest centroid.
3. At the point when the sum total of what items have been appointed, recalculate the places of the K centroids.
4. Recap Step 2 and 3 until the centroids never again move.

The Euclidean distance measure equation is given by:

$$d_e(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} ---------(11)$$

E. Support Vector Machine(SVM)

A SVM is a characterization method in light of the factual learning hypothesis [17, 18]. It is administered learning method that uses a marked informational collection for preparing and tries to discover a choice capacity that characterizes best the preparation information. The reason for the calculation is to discover a hyperplane to characterize choice limits isolating between information purposes of various classes. SVM classifier finds the ideal hyperplane that Correctly Separates (characterizes) the biggest portion of information focuses while expanding the separation of either class from the hyperplane. The hyper plane condition is given by

$$W^T X + b ------------- (12)$$

where *w* is weight vector and b is bias. Nonlinearity is satisfied by mapping the input features *x* into higher dimensions using a function

$$min_{\omega,\xi,b}\ (\omega,\xi) = \frac{1}{2}\omega^T\omega + C\sum_{i=0}^{m}\xi_i$$

And hence the hyperplane becomes:

$$min_{\omega,\xi,b}\ (\omega,\xi)=\frac{1}{2}\omega^T\omega + C\sum_{i=0}^{m}\xi_i ----(13)$$

Such that

$$Y_i(\omega^T\varphi(X_i) + b) \geq 1 - \xi_i \quad \text{i=1,----,N---------(14)}$$

$$\xi_i \geq 0, i = 1, ---, N\text{------------------------------(15)}$$

The obliged enhancement issue in condition 11, 12 and 13 is alluded as the primal improvement issue. The streamlining issue of SVM is typically composed in double space by presenting limitation in the limiting utilitarian utilizing Lagrange multipliers. The dual formulation of the problem is

$$max_a \sum_{i=1}^{\pi} \alpha - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i\, \alpha_j y_i y_j (x_i, x_j) --- (16)$$

Subject to $a_i \geq 0$ for all i=1, ----, N

$$\sum_{i=0}^{\pi} \alpha_i y_j = 0 ------(17)$$

Thus the hyperlane is

$$f(x) = sgn[\sum_{i=1}^{m} y_i\ a_i(X_i, X) + b ----- (18)$$

V. CONCLUSION

Recognition of disfluencies in the passage is a time consuming one and also results in poor assessment. This recognition process involves feature extraction and classification techniques. In this paper few techniques are discussed for feature extraction and classification process, LPC and MFCC methods are suitable for feature extraction process. Among LPC and MFCC, MFCC is best because the basic idea if this to five accurate audible signal and for classification DTW can be used for score matching , because it works on time and speed of speech and it can classify continuous and discontinuous speech.

REFERENCES

[1] R. C. Rose and G. Riccardi, “Modeling Disfluency and Background Events In ASR for a Natural Language Understanding Task"IEEE”, Vol X, 0-7803-5041 -3/99 -1999 .

[2] Carlos Hernindez-Espinosa, Mercedes Fernindez-Redondo, Pedro G6mez-Vilda, and Juan,” Diagnosis of Vocal and Voice Disorders by the Speech Signal”, IEEE ,0-7695-0619-4/00 -2000.

[3] Jagadish Nayak and P Subhanna Bhat, “Identification of Voice Disorders using Speech Samples”, lEEE, 0-7803-765l-Xl03 -2003.

[4] Dagen Wang, Shrikanth S. Narayanan, “A Multi-Pass Linear Fold Algorithm For Sentence Boundary Detection Using Prosodic Cues” ,0-7803-8484-9/04 ICASSP-2004,IEEE.

[5] MatthiasHonal, and Tanja Schultz, “Automatic disfluency removal on recognized spontaneous speech -Rapid adaptation to speaker-dependent disfluencies”, 0-7803-8874-7/05 IEEE-ICASSP 2005.

[6] Chung-Hsien Wu, Senior Member, IEEE, and Gwo-Lang Yan, “Speech Act Modeling and Verification of Spontaneous Speech With Disfluency in a Spoken Dialogue System”, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 13, NO. 3, MAY 2005.

[7] Rodrigo C. Guido' Jos& C. Pereimi Everrhon Fonsecat, “TRYING DIFFERENT WAVELETS ON THE SEARCH FOR VOICE DISORDERS SORTING”, 0-7803-8808-9flS/$?O.D0 2005 IEEE.

[8] Matthew Lease, Mark Johnson, and Eugene Charniak, “Recognizing Disfluencies in Conversational Speech”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 5, SEPTEMBER 2006.

[9] Jui-Feng Yeh and Chung-Hsien Wu, Senior Member, IEEE, “Edit Disfluency Detection and Correction Using a Cleanup Language Model and an Alignment Model”, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 5, SEPTEMBER 2006.

[10] Tatsuya Kawahara, Intelligent Transcription System based on Spontaneous Speech Processing”, Second International Conference on Informatics Research for Development of Knowledge Society Infrastructure (ICKS'07)-IEEE,2007

[11] Mauricio Alvarez, Ricardo Henao, Germ´an Castellanos, Juan I. Godino, Alvaro Orozco, “Kernel Principal Component Analysis through Time for Voice Disorder Classification”, Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006

[12] Serdar Yildirim and Shrikanth Narayanan, Senior Member, IEEE, “Automatic Detection of Disfluency Boundariesin

Spontaneous Speech of Children Using Audio–Visual Information", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 17, NO. 1, JANUARY 2009.

[13] Lotfi Salhi and Talbi Mourad, "Voice Disorders Classification Using Multilayer Neural Network" , 2008 International Conference on Signals, Circuits and Systems-IEEE.

[14] Che-Kuang Lin, Student Member, IEEE, and Lin-Shan Lee, Fellow, IEEE, "Improved Features and Models for DetectingEdit Disfluencies in Transcribing Spontaneous Mandarin Speech", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 17, NO. 7, SEPTEMBER 2009.

[15] Wen Wang, Gokhan Tur, Jing Zheng, Necip Fazil Ayan, "AUTOMATIC DISFLUENCY REMOVAL FOR IMPROVING SPOKEN LANGUAGE TRANSLATION", IEEE-ICASSP 2010.

[16] Zeinab Mahmoudi1, Saeed Rahate, Mohamad Mahdi Ghaseme, "CLASSIFICATION OF VOICE DISORDERS IN CHILDREN WITH COCHLEAR IMPLANTATION AND HEARING AID USING MULTIPLE CLASSIFIER FUSION", 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010).

[17] C. Shahnaz,,S. A. Fattah and U. Mahbub, "Detection of Voice Disorders based on Wavelet and Prosody-related Properties", 978-1-4577-1124-4/11, 2011 IEEE.

[18] Guoping Li , Mark E. Lutman , Shouyan Wang & Stefan Bleeck, "Relationship between speech recognition in noise and sparseness", International Journal of Audiology 2012; 51: 75–82.

[19] Adam Jacksa and Katarina L. Haleya, "Auditory masking effects on speech fluency in apraxia of speech and aphasia: comparison to altered auditory feedback", Journal of Speech, Language, and Hearing Research • Vol. 58 • 1670–1686 • December 2015

[20] Sunday Iliya and Ferrante Neri, "Towards Artificial Speech Therapy:A Neural System for Impaired Speech Segmentation", International Journal of Neural Systems, Vol. 26, No. 6 (2016) 1650023 (16 pages) World Scientific Publishing Company.

[21] Suszyński W., Kuniszyk-Jóźkowiak W., Smołka E., Dzieńkowski M., Prolongation detection with application of correlation procedure, Annales Informatica UMCS, Lublin,2 (2004) 183.

[22] Suszyński W., Kuniszyk-Jóźkowiak W., Smołka E., Dzieńkowski M., Automatic recognition of non-fluent stops, Annales Informatica UMCS, Lublin, 1 (2003) 133.

[23] Andrzej Czyzewski, Andrzej Kaczmarek1and Bozena Kostek, "Intelligent processing of stuttered speech", Journal of Intelligent Information Systems, 10.1.1.109.715, 2003

[24] P.Howell, S. Sackin and K. Glenn,"Development of a two stage procedure for the automatic recognition of disfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers," Journal of speech Language and Hearing Research, Vol 40, pp. 1085, 1997.

[25] Pratik K. Kurzekar, Ratnadeep R. Deshmukh , Vishal B. Waghmare , Pukhraj P. Shrishrimal, "A Comparative Study of Feature ExtractionTechniques for Speech Recognition System", International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization).Vol. 3, Issue 12, December 2014. ISSN: 2319-8753.

[26] Zebrowski, P.M. (1991). "Duration of the speech disfluencies of beginning stutterers", J. Speech-Hear-Res., 34, 3.

[27] Michael D. McClean, D.C. Stephen M. Tasko,MI Charles M. Runyan James "Orofacial Movements Associated With Fluent Speech in Persons Who Stutter" Journal of Speech, Language, and Hearing Research,Vol. 47, pp 294–303, April 2004 ©American Speech-Language-Hearing Association 1092-4388/04/4702-0294

[28] Waldemar Suszyńskia*, Wiesława Kuniszyk-Jóźkowiaka, Elżbieta, Smołkaa, Mariusz Dzieńkowski,, "Speech disfluency detection with the correlative method", Annales UMCS Informatica AI 3 (2005) 131-138.

[29] K.M Ravikumar, R.Rajagopal, H.C.Nagaraj, "An Approach for Objective Assessment of Stuttered Speech Using MFCC Features", DSP Journal, Volume 9, Issue 1, June, 2009.

[30] E.Yairi and B.Lewis, "Disfluencies at the onset of stuttering," Journal of speech & Hearing Research, vol.27, pp. 154-159, 1984.

[31] Aini Hafizah Mohd Saod, Dzati Athiar Ramli, "Preliminary Study on Classification of Apraxia Speech using Support Vector Machine", Australian Journal of Basic and Applied Sciences, 5(9): 2060-2071, 2011ISSN 1991-8178.

[32] Monica Mundada, Sangramsing Kayte, Sumegh Tharewal, Dr. Bharti Gawali, "Classification of Fluent and Dysfluent Speech Using KNN Classifier", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 9, September 2014 ISSN: 2277 128X.

[33] G. Manjula, Dr. M. Shiva Kumar, "Stuttered Speech Recognition For Robotic Control", ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 12, June 2014.

[34] Jiang J, Lu C, Peng D, Zhu C, Howell P, "Classification of Types of Stuttering Symptoms Based on Brain Activity", PLoS ONE, e39747. doi:10.1371/journal.pone.0039747,2012 Arghavan Bahadorinejad, Farshad Almasganj.