International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Vocal Tone Analysis for Identification of Stuttering Levels based on Tamil Syllable

M. Manjutha, P. Subashini, M. Krishnaveni

Abstract: ***Speech is the most unique feature of humans and it is the basic mode of verbal communication. Among all, a few people face hurdles in producing normal speech because of various types of speech disorders. Stuttering is one of the disorder types mainly characterized by repetition of syllables or words and involuntary interruption during the speech. One of the unsolved problems in the realm of fluency disorder is the level identification, based upon the patient's utterance before and after the speech therapy. The main objective of the paper is to perform analysis of vocal tone to identify the major differences between a normal, moderate and severely stuttered speech, particularly for the Tamil spoken language. The stutter speech vocal tone analysis involves envelope detection based on Hilbert transform has been computed from the input of normal and stuttered speech waves in which by applying normalization to the spectrum and cepstrum of the obtained signal considering only Tamil syllable as input. The experimental outcomes are given as subjective evaluation in three categories of speech signals which results people affected by severe stuttering having low vocal tone than moderate and normal stuttering speech.***

Keywords: Cepstral analysis, Envelopes, Hilbert Transform, Normalization, Stuttering and Tamil Speech

I. INTRODUCTION

Speech disorder is possibly a disabling situation with widespread and lifelong implications which may change behaviour, cognition, social and emotional well-being. Stuttering is one such speech impaired disorder characterized by disruptions in the flow of speech or disfluencies. Census of India collected during 2011 on disability, revealed that 1.1 millions of males and 0.9 millions of females were affected by speech disability (Office of the Registrar General and Census Commissioner, India, New Delhi, 27-12-2013)[4]. Stuttering can vary significantly from one person to another and also from time to time. During recent decades, stuttering has gained much attention in the area of physiology, psychology, acoustics and signal analysis. It is also reviewed that research works are concentrating highly in the field of speech synthesis, speech diagnosis and analysis.

Revised Manuscript Received on 30 July 2019.

***** Correspondence Author

Manjutha Manavalan*, Department of Computer Science, Avinashilingam Institute Coimbatore, India.

Dr. Parthasarathy Subashini, Professor, Department of Computer Science, Avinashilingam University, Coimbatore, India.

Dr. M.Krishnaveni, Assistant Professor, Department of Computer Science, Avinashilingam University for Women, Coimbatore, India.

Generally, speech language pathologist analyses and treats the stuttered people by manual diagnoses process. To avoid manual work, efficient automatic diagnoses system introduced to identify and analyse the level of stuttering effectively. The vocal tone is one of the features used to analyse the stuttering based on the occurrence of syllable or word repetition.

Analysis of stuttered speech may have mean number of repeated units per instance of syllable, word repetition, mean duration of sound syllable repetition, prolongation and also various related measures of the frequency between and within word speech disfluencies. The speech disfluency can be characterised by normal disfluency which has minimum risk whereas stuttering has high risk. The normal disfluency has phrase repetitions, interjections like "um" and "uh", broken words between the word pause or silence whereas stuttering has repetitions of syllable or words or multi-syllabic whole word repetitions, a dysrhythmic phonation or prolongations of sounds and silent blockages of speech[5][18]. The vocal tone is one of the features used to analyse the stuttering based on the occurrence of syllable or word repetition since it has the basic unit and contains more temporal information than the phoneme. During the speech production the vocal cords get open and closed which breaks the airstream into sequences of pulses. The repetition rate of these pulses is pitch that determines the fundamental frequency of speech signal. The remaining section of the paper is structured as follows. Section 2 briefly describes contribution of various researcher related to stutter speech. Outline of the proposed work explains vocal tone analysis of Tamil stuttered syllables in Section 3. Section 4 illustrates the analysis of results and discussion. Section 5 draws the conclusion and future work.

II. LITERATURE REVIEW

This section briefly reviews the previous work carried out by various researchers related to vocal tone.

Leanne Jani et al. (2013) introduced an Audio index procedure to compute the duration and frequency of the stutter syllables. The data set consist of real time recorded sample of 200 syllables that are assed by ten Speech Language pathologists. The experimental analysis given shows the simultaneous and successive conditions of the difference among syllable, stutter syllable counts and duration made successively by t-tests[16]. The effect of syllable structure on the frequency of speech disfluencies in adults with stuttering was examined in 2015 (Elham Masumi et al.).

The main objective of the paper is to distinguish the frequency of disfluencies for 60 words and 60 non-words with the simple and difficult syllable structure of adults with stuttering. Stuttering Severity Instrument-3 (SSI-3) was used to identify the diagnosis of stuttering, and the data distribution of the normality measurement was analysed and tested using SPSS 16.0 and Kolmogorov-Smirnov respectively. The experiment result interprets that the mean frequency of disfluencies in the difficult syllable structure was more than that of the syllable structure in the list of words [9]. Peter Howell and Louise Vause (1986) explained the acoustic analyses and perception of stuttered vowels. The author proposed two analyses. In the first experimental the analysis is carried out on the utterance of stutter people often produced in the sound like 'schwas' that is a neutral vowel. The second experimental analysis is on 'schwas' of several stutterers. The outcome of the experiment is tested with ANOVA that is used to analyse the two types of experiment with spectral properties of vowels. From the analysis it reveals that the low amplitude and the short interval are the significant issue that cause the stuttered vowels sound like schwa [19].
N.Derakhshan and M.H. Savoji(2006) has proposed Hilbert transform based sub-band analysis speech that represented in time-frequency to enhance the speech using spectral subtraction[6]. Lei He and Volker Dellwo (2016) introduced Praat script based Hilbert transform signal to identify the envelope of the speech signal which has Temporal Fine Structure (TFS) and it is modulated with Low Frequency Envelope (ENV) [11]. In 1992, Peter Howell and Mark Williams analysed 24 children and eight teenage stutters syllable repetitions of vowel neutralization. The outcome of the analysis depicts that the vowels to be neutral when the differences occurred in vowel duration or the fundamental frequency. The source of excitation and low intensity cause the neutral vowels which may exist in the late childhood [20]. Spectral and cepstral based phoneme parameter framework is proposed and analysed with different noise condition as well as different feature sets. In order to evaluate the accuracy of the proposed framework two classical methods such as Support Vector Machine (SVM) and Naïve Bayes were used by Grazina Korvel et al. (2018). In which cepstral parameter gives better results (95.2%) than spectral parameters (90.8%) [10]. The spectrogram and formants were analysed with the WaveSurfer in utterance of the stuttered and non-stuttered speech that are compared and segmented. Stuttered speech repetition of syllables such as 'po' and 'do' is repeated instead of 'p' which were visually depicted by Elżbieta Szabelska(2013) [8]. The speech production, speech spontaneity in rhythmic speech for stuttering frequency was examined by Jason H. Davidow (2011). The finding depicts the mean vowel duration and also the short phonated intervals percentage for the stuttered and fluent speech. From this, it is understood that the vocal fold vibration duration gets altered during rhythmic speech [13]. Andrzej Czyzewski (2003) analysed the stuttered speech based on the machine learning algorithm such as neural network and rough set. The different analysis examined is vocal tone, formant analysis, vowel prolongation and vowel repetition. The individual series of stuttered speech analysis results given will help to diagnose the speech disorder [2].

From the above literature, it is understood that most of the authors have contributed towards the analysis of stuttered speech syllable. Still there is a need for analysing regional languages. In this paper, vocal tone analysis on Tamil regional stuttered syllable is focused which analyses the vocal tone of the normal people, moderate stutter and severe stutter speech based on Tamil syllable to identify the disfluency.

III. VOCAL TONE ANALYSIS FOR TAMIL SYLLABLE

People who are affected by stuttering have great difficulty while communicating with others when their vocal apparatus fails. The human speech production consists of two separate processes in which one produces initial sound and another modifies it. The larynx generates a speech sound whose spectrum has many different frequencies. The articulators collectively have tongue, teeth, lips and velum that alter the spectrum of speech sound over the time. The vocal folds produce the sound. However, the tone can be modified by manipulating the different structures of the vocal tract. The proposed vocal tone analysis consists of data acquisition, logarithmic spectrum before normalization, cepstrum before normalization, envelope of the signal spectrum, normalization, logarithmic spectrum after normalization, cepstrum after normalization and vocal tone. Initially, the collected data was stored in a database in which spectrum and cepstrum of the stuttered speech signal were evaluated. The obtained result was further processed for identifying the envelope of the signal. The envelope of the signal was normalized by spectrum and cepstrum since the vowels and fricative produced by the men and women have different spectral shapes. In the proposed work, the normalization is applied to predict the vocal tone of the people affected by stuttering.

A. *Tamil Language*

Tamil Language is a Dravidian Language predominantly spoken by the Tamil people around the world of Southern India, Pakistan, Northern Sri Lanka and also in South East Asia of Malaysia, Singapore, and Fiji. South Africa parts of East Africa, Great Britain, the United States, Canada, and Trinidad, Guyana, Mauritius also Tamil is official language in India, Sri Lanka, and Singapore. Tamil alphabets consist of twelve vowels and eighteen consonants and each alphabet is represented by unique letter. Tamil is familiar for syllabic language where the vowel and consonants association is represented by a composite form of consonant and a secondary symbol for each vowel. Generally, a secondary symbol is included either before or after the consonant to make up the syllable, but in a few cases the consonant is slightly modified to form a new shape. Tamil syllables are derived by a combination of vowels (twelve) and consonants (eighteen) which form a graphemes (total of two hundred and sixteen different characters) [23].

B. *Data Acquisition*

The proposed vocal tone analysis is proved using the real time data that are recorded by using Audacity 2.1.3. People affected by stuttering and fluent speakers used to record the speech in which based on the utterance of stuttered speakers, are categorized as moderate and severe. The complex sentence and simple sentences in Tamil language were recorded. Each speaker was allowed to repeat the spontaneous speech and Tamil sentences 5 times in order to create the actual speech database. The speech files were recorded in the .wav file format has sampling frequency of 16000 kHz. Every stuttered syllable has been segmented manually by hearing the audio. The sample syllable of 'போ'(Pō) was uttered by Normal Speaker (NS), Moderate Stutter Speaker (MS) and Severe Stuttered Speaker (SS) as shown in Fig.1.

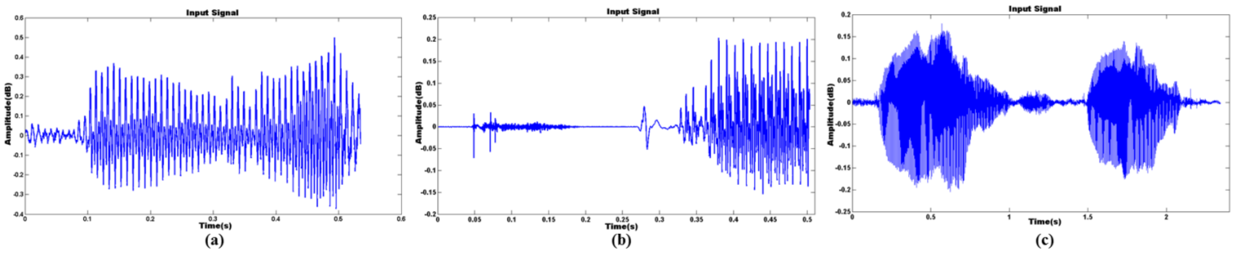


Fig.1. Sample input syllable of 'போ'(Pō)(a) Normal (NS) (b) Moderate (MS) and (c) Severe Stuttered (SS)

C. *Logarithmic Spectrum Before Normalization*

The logarithmic spectrum is necessary for analysing the vocal tone since it has the vocal tract information of the speakers as envelope and the excitation source information as pitch. As speech signal is characterized as a convolution between moderate changes in the vocal tract impulse response (filter) and rapid changes in the glottal pulse (source), the speech spectrum consists of the spectral envelope (low frequency) and the spectral details (high frequency). Logarithmic spectrum separates the low frequency and high frequency from the spectrum of speech signal.

The logarithmic spectrum of the speech is obtained by taking the logarithmic values of the original speech signal which is given in (1) and computed log spectrum of syllable uttered by three different speakers as shown in Fig.2.

$$X = \log x \qquad (1)$$

where, X represents spectrum of the signal and x represents the original speech signal.

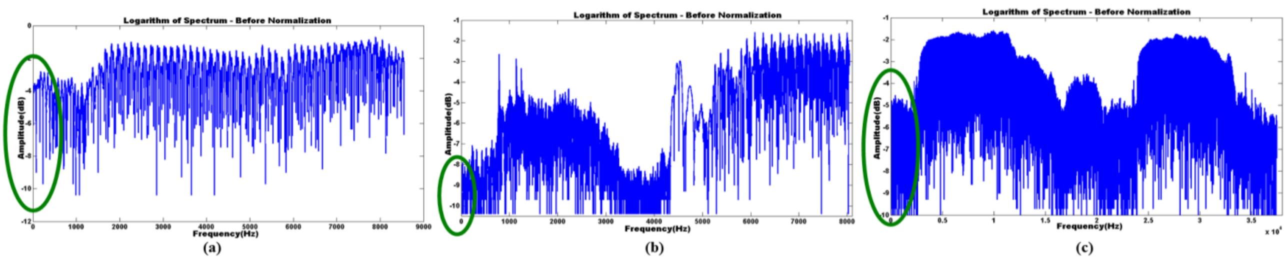


Fig.2. Logarithmic spectrum of syllable 'போ'(Pō)(a) Normal (NS) (b) Moderate (MS) and (c) Severe Stuttered (SS)

Cepstrum Before Normalization

The cepstrum can easily differentiate the difference between the normal and abnormal speech signal. Cepstrum is a reliable way of acquiring an estimation of the dominant fundamental frequency for long clean stationary speech signal. The cepstrum is a Fourier analysis of the logarithmic amplitude spectrum of the signal [22]. The speech signal is defined as the real cepstrum of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal.

Initially FFT subjected to log based transform of the frequency axis and de-correlate using Discrete Cosine Transform which is illustrated in the following steps [3][12][14].

Step1: The speech frame is multiplied by Hamming window and converted in to frequency

Step2: To extract the envelope triangular band pass filter M(f) is applied which is represented as,

$$M(f)=1127\log\left(1+{}^{f}/_{700}\right) \quad (2)$$

Step 3: Discrete Cosine Transform C_i on log energy obtained from the triangular bandpass filter have the cepstral coefficients i and M samples that is,

$$C_i=\sqrt{\frac{2}{M}}\left[k_m \cos\left(\frac{\pi i}{M}\left(m-\frac{1}{2}\right)\right)\right] \quad (3)$$

$$\text{and } k_m=\begin{cases} 1 \text{ if } m\neq 0 \\ \frac{1}{\sqrt{2}} \text{ if } m=0 \end{cases}$$

The cepstrum of the signal consists of two elements.

An element from the excitation sequence was (a pulse train for voiced speech) in the higher quefrencies, namely, glottal excitation [21]. The other element originates from the vocal tract impulse response, and is present in the lower quefrencies. The peak in the higher quefrencies indicates that there are some periodicities in the signal. The peak is located at the period of the fundamental frequency. In the cepstral domain, the signal components are additive, easily allowing for the use of linear filtering techniques. The cepstrum of the stuttered syllable periodic excitation has larger interval gaps as shown in Fig.3 (b) and (c) when comparing to the Fig.3 (a) which interprets that the stuttered syllable has larger stop gaps.

D. *Envelope of the signal*

While diagnosing stuttering, it is necessary to analyse not only the non-fluent fragments of utterance since whole structure of the utterance has been disturbed.

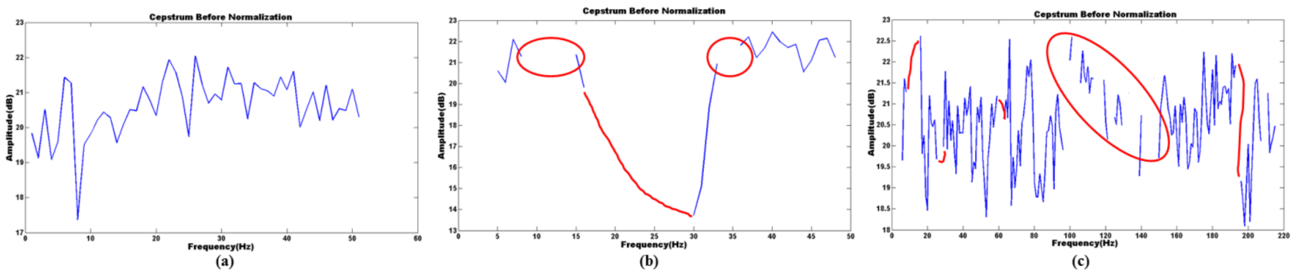


Fig.3.Cepstrum before normalization of syllable ‘போ(Pō) (a) Normal (NS) (b) Moderate(MS) and (c) Severe Stutter (SS)

Hence the envelope of the signal contains both fluent and characteristics of non-fluent fragments of speech signal. Envelope of the signal spectrum is represented by absolute value of the analytic extension of the speech signal which has been presented along with the original signal [11]. The speech signal envelope is an average of upper and lower envelope of signal, using Hilbert transforms $\hat{h}(t)$ act as imaginary part and the original signal $h(t)$ act as real part also the analytical signal is a complex time domain signals $\tilde{h}(t)$ which is represented in (4).

$$\tilde{h}(t)=h(t)+j\hat{h}(t) \quad (4)$$

In (4), j is the imaginary number $\sqrt{-1}$ and $H\{\cdot\}$ denotes the operation of the Hilbert transform and $\hat{h}(t)=H\{h(t)\}$[17].

The real part such as upper envelope is $U(\omega)$ of the Fourier transform can be expanded in (5) where, T is total duration of the signal and $t\in[0,T]$.

$$U(\omega)=+\int_0^T h(t)\cos(\omega t)dt-j\int_0^T h(t)\sin(\omega t)dt \quad (5)$$

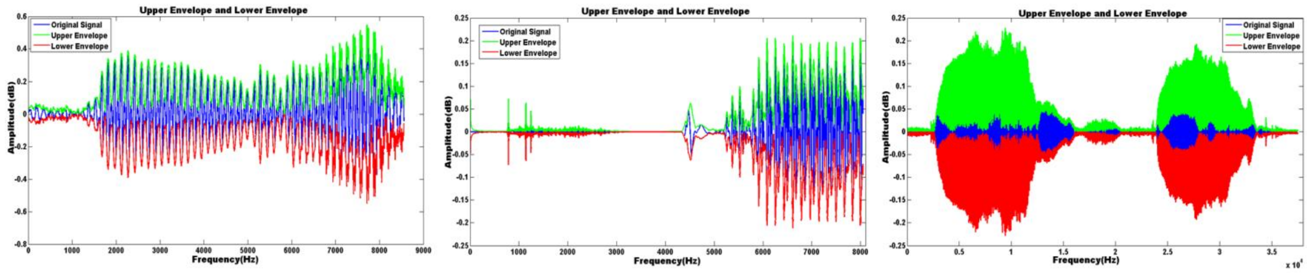
The lower envelope of the signal such as Hilbert spectrum $\hat{L}(\omega)$ is represented in the negative part of the frequency domain which is lesser than zero. The lower envelope of the signal is evaluated based on (6) and the computed imaginary part of the signal as shown in Fig.6.

$$\hat{L}(\omega)=-\int_0^T h(t)\sin(\omega t)dt-j\int_0^T h(t)\cos(\omega t)dt \quad (6)$$

The spectrum properties hold by $h(t)$ which act as an envelope that shapes the exterior boundary of signal’s waveform, and the phase term defines signal’s variations inside this boundary. The upper and lower envelope of the signal together is shown in Fig.4.

Retrieval Number: B3211078219/19©BEIESP
DOI: 10.35940/ijrte.B3211.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



(a) (b) (c)

Fig.4. Envelope of syllable 'போ'(Pō)(a) Normal(NS) (b) Moderate (MS) and (c) Severe Stuttered(SS)

E. *Normalization*

The normalization typically focused on the male or female changes in the vocal tract size and shapes. Since the individual has the difference in the speech production vocal tone normalization is necessary which examined by the listeners perceptually in order to evaluate the fluent and non-fluent speech. The normalization such as envelope of both upper and lower envelope spectrum is evaluated. The obtained envelope of the signal was normalized with upper and lower envelope of spectrum. The analysis of the vocal tone is done by changing the vocal pitch through the modified cepstral analysis. The cepstral analysis consists of three major steps that is the compression of dynamic spectrum, spectrum band reduction for cepstral analysis and frequency of the vocal tone estimation [2]. The modification in the cepstral analysis will lead to enhancing the visual perception of the cepstral maximum. Compression of the dynamics spectrum is evaluated through an envelope on maxima and minima of the spectrum which is calculated (7) and (8).

$$U(\omega)_i=\begin{cases} X_i \text{ for } X_i>U(\omega)_{i-1} \\ U(\omega)_{i-1}+\left[1-\exp\left(-\frac{r_f}{c_f}\right)\right].(X_i-U(\omega)_{i-1}) \text{ for } X_i\leq U(\omega)_{i-1} \end{cases} \quad (7)$$

$$\hat{L}(\omega)_i=\begin{cases} X_i \text{ for } X_i<\hat{L}(\omega)_{i-1} \\ \hat{L}(\omega)_{i-1}+\left[1-\exp\left(-\frac{r_f}{c_f}\right)\right].(X_i-\hat{L}(\omega)_{i-1}) \text{ for } X_i\leq\hat{L}(\omega)_{i-1} \end{cases} \quad (8)$$

where, $U(\omega)_i$ is upper envelope values, $\hat{L}(\omega)_i$ is lower envelope, X_i is spectrum logarithm, i is spectrum coefficients from i = 0, 1, 2 etc, r_f is frequency analysis resolution and c_f is spectral components integration constant (r_f / c_f ratio was set to –1/10).

The obtained upper and lower envelope values are utilized to evaluate the normalized spectrum according to (9).

$$N_i=\frac{X_{max}-X_{min}}{U(\omega)_i-\hat{U}(\omega)_{i-1}}(X_i-\hat{L}(\omega)_{i-1})+X_{min} \quad (9)$$

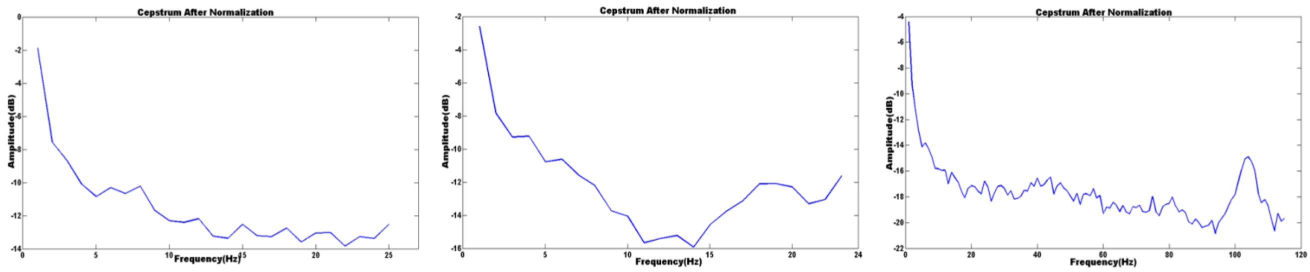
where, X_{max} is maximum of the spectrum logarithm, and X_{min} is minimum of the spectrum logarithm.

F. *Logarithmic Spectrum after Normalization*

In order to analyse the speech signal, two important features such as spectral and temporal features are necessary. In this paper, spectral feature was evaluated and the logarithmic spectrum of the three different types of speech signal was dynamically compressed after the normalization. The logarithmic spectrum of the signal is normalized with upper and lower envelop of the sampled speech as in (1).

G. *Cepstrum after Normalization*

In speech signal processing, cepstrum method is a traditional characteristic parameter detection algorithm. A cepstral feature was obtained from cepstrum of the speech signal which was normalized to detect the glottis incentive of pitch information and produces the smooth cepstral envelope signals. Cepstrum is obtained by Inverse Discrete Fourier Transform (IDFT) of the log magnitude of the Discrete Fourier Transformation (DFT). In order to reduce the computational complexity in the Fourier transformation, cosine transformation is applied as followed in the cepstral before normalization. Normalized cepstral analysis is applied to reduce the signal discontinuity in the periodicity of the syllable, and hence the cosine transform will exist in the normalized cepstrum as shown in Fig.5.



(a) (b) (c)

Fig.5.Cepstrum after normalization of syllable 'போ'(Pō) (a) Normal (NS) (b) Moderate (MS) and (c) Severe Stuttered (SS)

H. *Vocal Tone*

The abnormalities of the vocal tone cause the speech disorder. Voice depends on the vibratory characteristics of the vocal folds, setting the air above the level of the larynx into vibrations as well. The vocal fold generates the sound. However the tone could be modified by manipulating the various structures of individual vocal tract [15]. The vocal tone frequency is predicted as a centre of gravity of a fragment of the calculated cepstrum according to (10).

$$\hat{f} = \frac{1}{r_c}\frac{\sum_{i=m}^{n} W_i}{\sum_{i=m}^{n} i.W_i} \qquad (10)$$

where $\hat{f}$ is estimated frequency, and r_c is resolution of cepstral analysis that is $2^{\text{bits per sample}}$. The bit per sample of each signal is 16. The W_i is cepstrum coefficient number i and m, n are number of cepstral coefficient which contains the maximum and minimum related to the vocal pitch. Number of maximum cepstral coefficient k is estimated by $m = k - 1$ and $m = k + 1$. The sequence of values obtained from the successive frames of the speech signal vocal tone can be evaluated by the estimated frequency as shown in Fig.6.

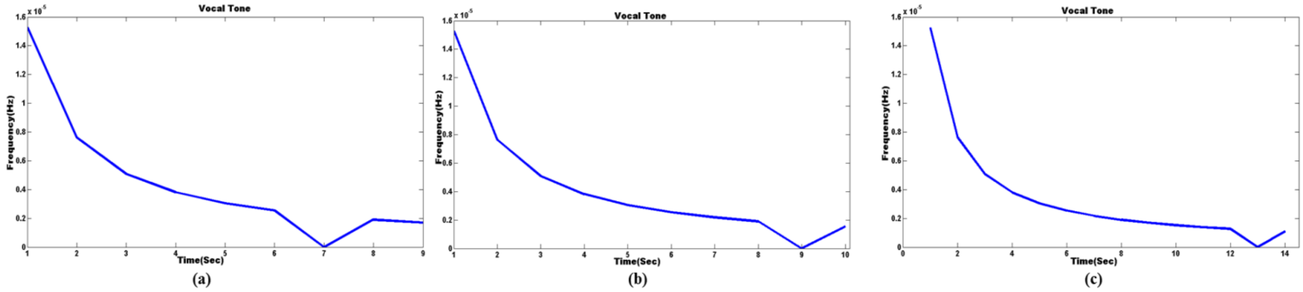


Fig.6. Vocal tone of syllable 'போ'(Pō)(a) Normal (NS) (b) Moderate (MS) and (c) Severe Stuttered (SS)

IV. ANALYSIS AND DISCUSSION

The proposed work is implemented in MATLAB computational platform and in version MATLAB R2013a [1][7]. The presented experimental results indicate the analysis of vocal tone especially for Tamil language, comparing with three types of speakers, namely, fluent speaker, moderate stutter speaker and severe stutter speaker. The original speech of three speaker syllable results show that the severe stutter uttered syllable 'போ'(Pō)has more silence than the moderate stutter and normal speaker with respect to time. The total length of the NS, MS and SS syllable are 0.5347 sec, 0.5030 sec and 2.3428 sec as indicated in Fig.1. The stuttered syllable of 'போ'(Pō) was initially stuttered by MS and SS with silence. The sever stutter, repeated the syllable 'போ'(Pō) twice as shown in Fig.1 (c). The speech signal is represented in the time domain with respect to the amplitude. The logarithmic spectrum of NS, MS and SS of syllable 'போ'(Pō) were false in the negative range which has been highlighted with circle in the amplitude domain as shown in Fig.2. The silent part of the logarithmic spectrum moves down and gradually moves up and down and it remains until the silence region was terminated as in Fig.2 (b) and (c). The voice part of the logarithmic spectrum gradually moves towards the upper level that is represented in the frequency domain with respect to the amplitude. The NS does not have long silence pause so that the logarithmic spectrum of the signal intensively maintains the spectrum peak in the upper region as shown in Fig.2 (a). In the log magnitude spectrum, the slow varying components represent the envelope corresponds to the vocal tract and the fast varying components to the excitation source. The outcome of the result interprets that the NS of logarithmic spectrum and its fine detailed features wave patterns are mostly repeated at regular interval of time. On the other side, the stuttered speech has some perturbation which also has some noises. Every recurrence or duration of these patterns resides to one glottal cycle, or one cycle of vocal fold opening and closing in the larynx. The cepstrum of the severe stuttered syllable signal periodic excitation has more interval gaps than the moderate stutter syllable and normal speaker as shown in Fig.3 (a, b and c) which interprets that the stuttered speech has larger stop gaps highlighted with circle. The envelope of the signal interprets that, the upper and lower envelope of the signal travels towards the boundary of positive and negative region of the signal. It matches state that frequency information will be high when the excitation of the vocal track is larger. The silence and the noise region of the signal have low frequency information. Depending upon the signal information, the original signal of the starting region will vary, and that will be originated either in the upper envelope or in the lower envelope region. The upper envelope of the signal is represented in green color, the lower envelope of the signal is represented in red color and the original signal is represented in blue color as shown in Fig.4 (a, b and c). The outcome of the cepstrum normalization results of syllable 'போ'(Pō) doesn't have any signal discontinuity in the intervals of each frames. Fig.5. shows the grass shape wave present in the cepstrum is flattened in the normalization of the speech syllable. The interpretation of the vocal tone from Fig.6 is that the MS and SS syllable pitch period is not perfectly regular compared to the normal speech. The vocal tone pitch and the frequency of normal syllable have a mixture of low and higher frequencies compared to the MS than SS. Severe and moderate stuttered syllable have the resonant stricture of vocal tone shapes whereas normal speaker has fricative stricture which consists of a very narrow opening vocal tone shapes as shown in Fig.6 (a, b and c).

Retrieval Number: B3211078219/19©BEIESP
DOI: 10.35940/ijrte.B3211.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

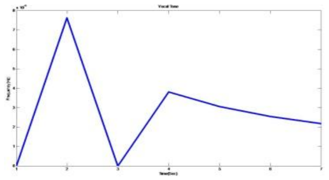
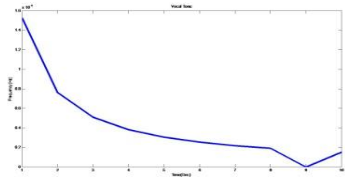
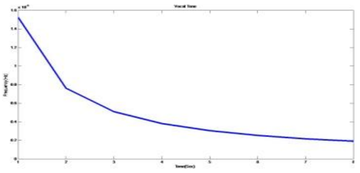
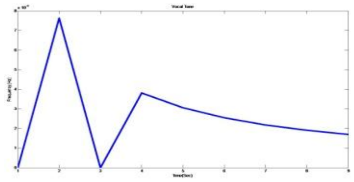
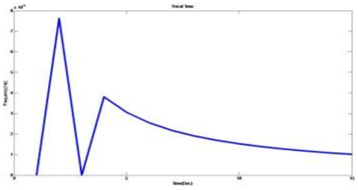
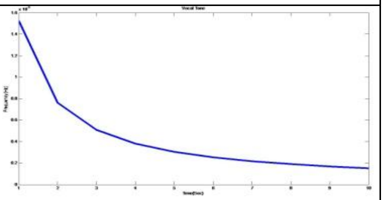
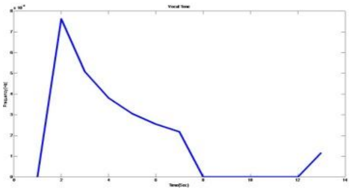
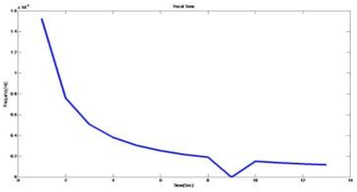
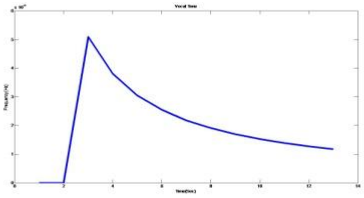
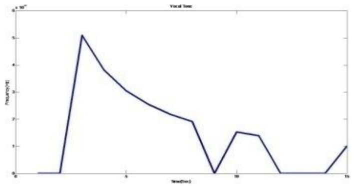
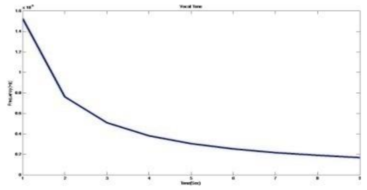
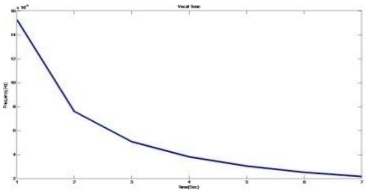
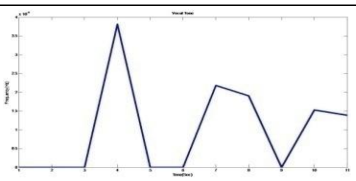
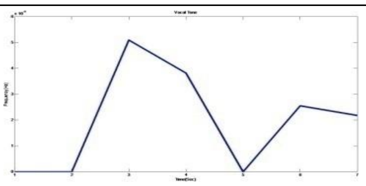
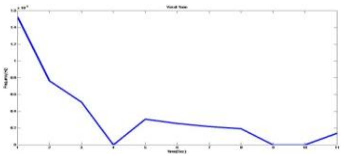
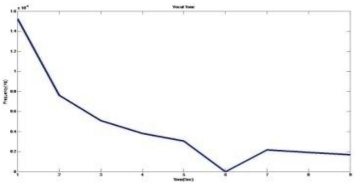
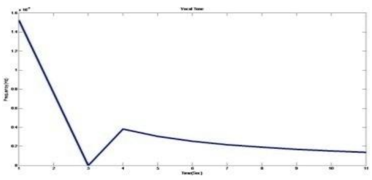
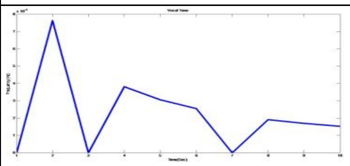
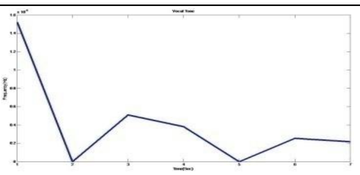
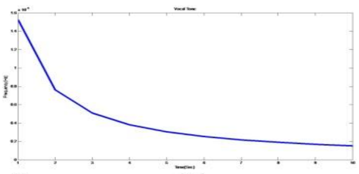
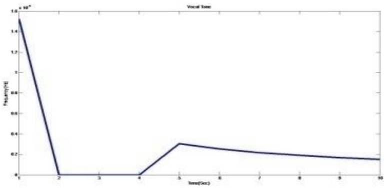
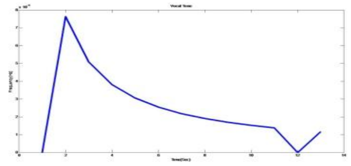
Table-I shows that the visualization of the vocal tone analysis distinguish with normal speech, moderate stuttered speech and severe stuttered speech. While recording, people often stuttered the specific syllables like கு(Ku),த(Tha) and ம(Ma). Some of the stuttered syllables obtained from the speaker is இ(I), உ(U), எ(E), ஓ(Ō), க(Ka), கு(Ku), த(Tha), ந(Na), ப(Pa), பெ(Pe),பொ(Pho), போ(Pō), ம(Ma), மா(Mā), மே(Mē), வா(Vā), வீ(Vī) which was used for the vocal tone analysis.

From the analysis of the vocal tone, the normal, moderate and severe stuttered speaker produced different syllable which has the various types of vocal tone shapes. In Table-I, Syllable இ (I) is again repeated by the speaker in which the MS vocal tone has curve shape. It shows that the stuttering level varies for both MS and SS when the utterance goes difficult. In syllable உ (U), NS and MS has the tap stricture which was extremely brief stop when comparing to the SS. In syllable எ (E) the vocal tone of the MS and SS shape is changed comparing to NS.

The NS has stop stricture since the long prolongation. In the syllable க (Ka) of NS has the tap stricture. However, the MS and SS vocal tone pitch was changed during the utterance. In NS of syllable த (Tha) has the tap stricture followed by stop stricture and the MS has produced the sound 'schwas' before the syllable த (Tha) so that the vocal tone shape was changed. Also, the SS uttered through the nasal cavity. The NS of syllable ந (Na) has greater opening with approximant stricture but the MS produced the sound 'schwas' before the syllable. Also, the SS uttered the syllable ந (Na) with larger prolongation. The syllable ப (Pa) produced by NS and MS was same. Hence it has moreover equal vocal tone shape and the SS has repeated the syllable twice since the vocal tone shape was changed. The syllable பெ (Pe) of both NS and SS has tap stricture but the MS stuttering level varies since the 'schwas' sound was obtained during the utterance. The next syllable பொ (Pho) uttered by the NS has stop stricture and MS utterance goes difficult since the 'schwas' sound and the SS have long prolongation.

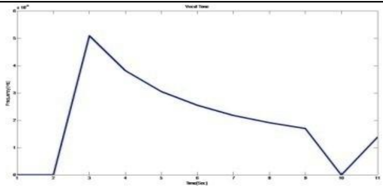
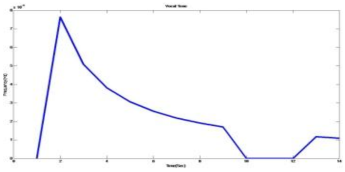
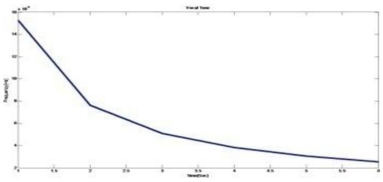
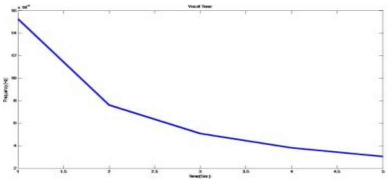
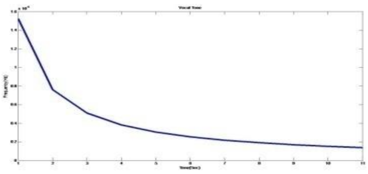
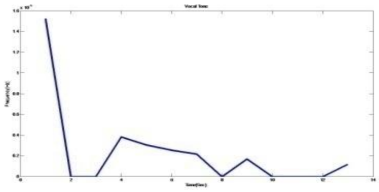
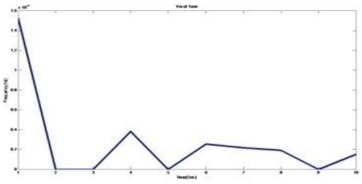
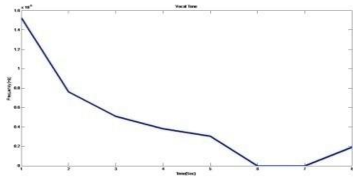
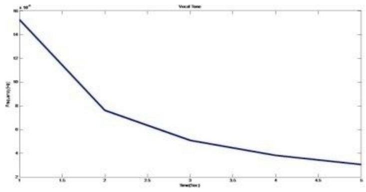
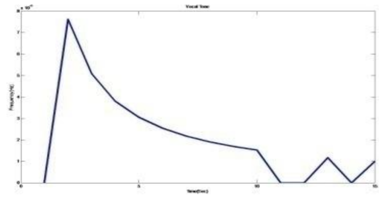
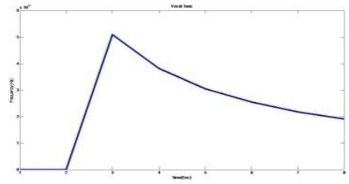
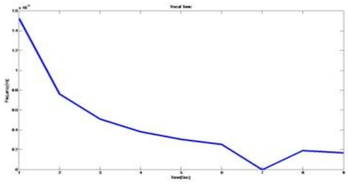
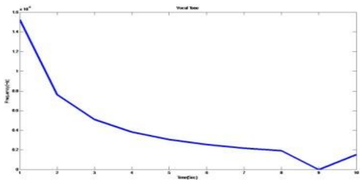
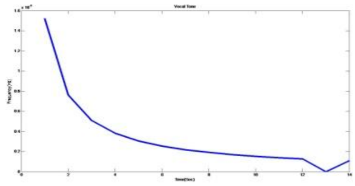
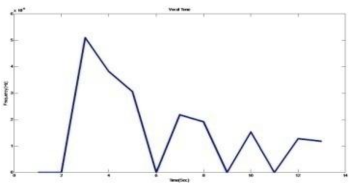
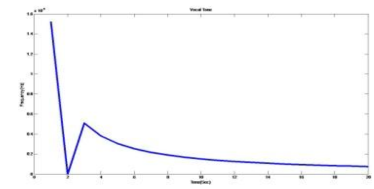
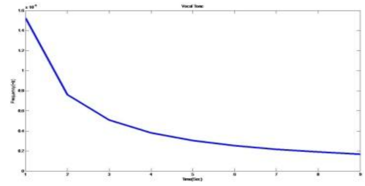
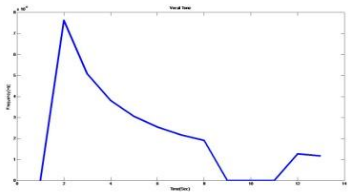
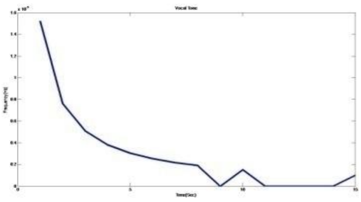
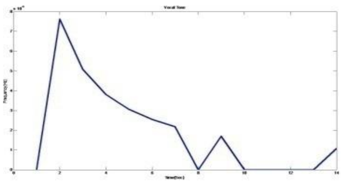
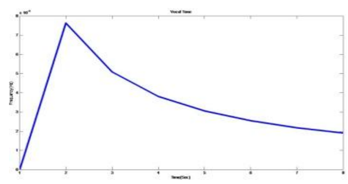
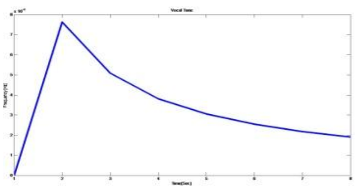
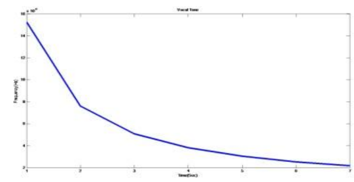
Hence the vocal tone shape differs when comparing to the NS of syllable பொ (Pho). In syllable போ (Pō) of NS and MS has same stricture when comparing to the SS since the SS produced by the speaker has repeated the syllable twice. The NS produced syllable ம (Ma) has stop stricture and the MS has difficulty while uttering the syllable since the 'schwas' sound and the SS has longer prolongation. Hence the vocal tone shape has been changed. The same syllable ம (Ma) uttered by the NS vocal tone frequency was changed and has stop stricture and the MS added the non-lexical verb 'hmm' before the utterance hence the vocal tone shape was changed when compared to the NS. The SS of syllable ம (Ma) has larger prolongation. The vocal tone of syllable ம (Ma) of NS, MS and SS has moreover equal shape because each speaker uttered the syllable with same prolongation. When comparing to the NS and SS of syllable ம (Mā), MS has non-lexical verb like 'hmm' added before the utterance of the syllable and the NS and SS vocal tone was similar which has stop stricture. The syllable மே (Mē) of NS and MS has the tap stricture of vocal tone shape whereas the SS syllable prolongation was too large since the vocal tone shape differs. The NS of syllable வா (Vā) has stop stricture vocal tone shape but the MS complex produce the syllable வா (Vā). Hence the vocal tone shape varies comparing to the NS. Also the prolongation of SS is larger since the air flow generates random sounds so that the vocal tone differs. The next syllable வீ (Vī) was uttered by both NS and MS vocal tone shape has approximant stricture while the SS level varies when the utterance goes difficult since the vocal tone shape has been changed.

Table- I:Vocal tone analysis of normal speech, moderate stuttered speech and severe stuttered speech

Tamil Spoken Syllable	NS	MS	SS
இ (I)			
உ(U)			
எ(E)			
ஓ(Ō)			
க(Ka)			
கு(Ku)			
த(Tha)		 	

Retrieval Number: B3211078219/19©BEIESP
DOI: 10.35940/ijrte.B3211.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

ந(Na)			
ப(Pa)			
பெ(Pe)			
பொ(Pho)			
போ(Pō)			
ம(Ma)			
மா(Mā)			
மே(M ē)			

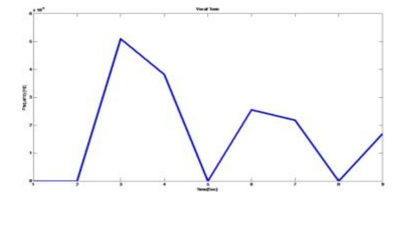
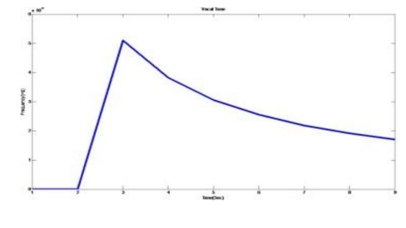
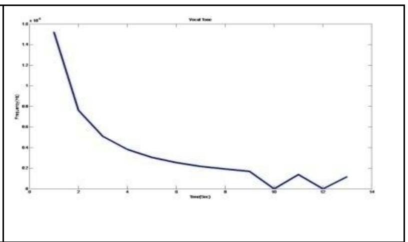
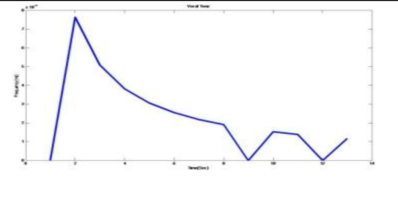
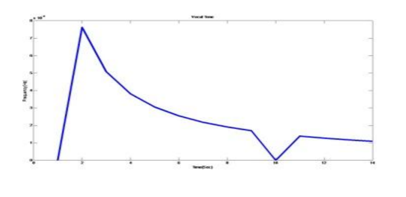
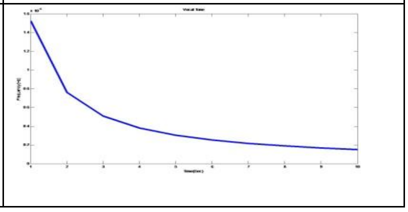
வா(*Vā*)			
வீ(*Vī*)			

Fig.7 and Fig.8 shows the similarity between short syllable (Kuril) and long syllable (Nedil) of NS, MS and SS. However, long syllables (Nedil) are much higher in frequency comparing to short syllables.

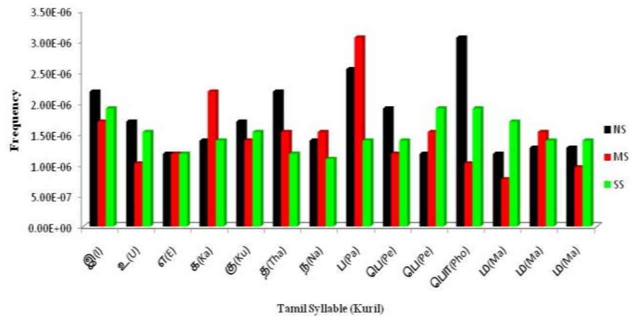


Fig.7. Tamil short syllables (Kuril)

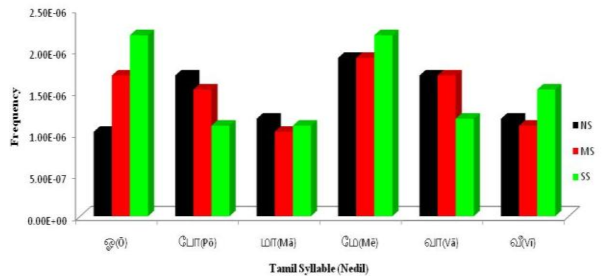
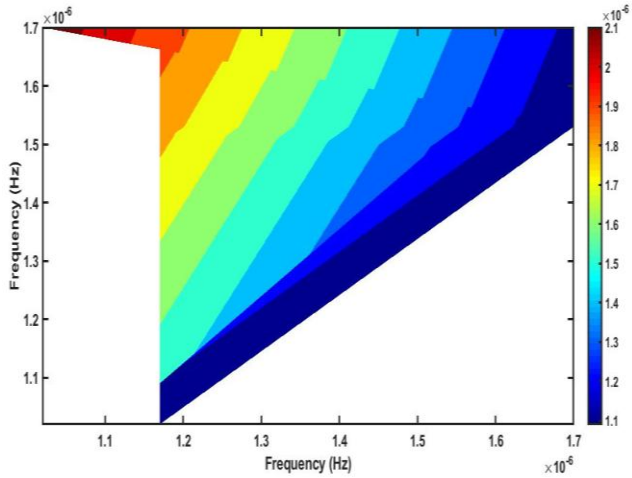


Fig.8. Tamil long syllables (Nedil)

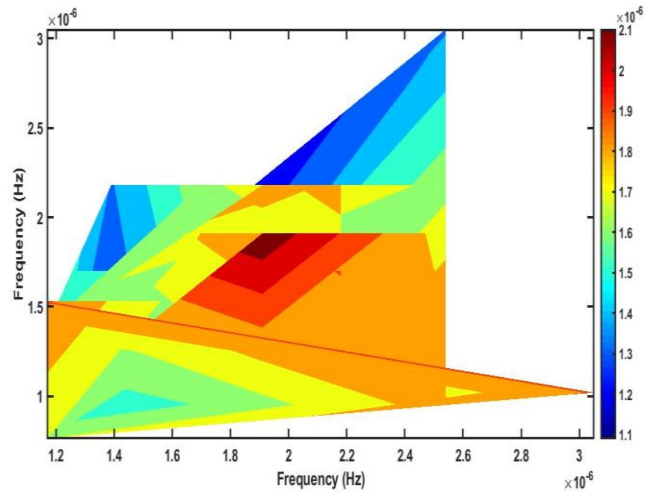
Fig.9 shows the comparison between the minimal vocal tone frequencies of the long and short Tamil syllables for all three speech spectrum respectively. The figures assist in identifying the possible correlation between the NS, MS and SS speech spectrums as well the persistence of their frequencies for different syllables. The orders of the short and long syllables were sufficiently selected signifying the learning approach of the speakers corresponding to the Tamil alphabets.

Fig.9 (b) with Tamil short syllable shows non-sequenced correlation between the three speakers, increasing the difficulty. This variation of frequencies over both syllables could be mainly due to the delayed pick up of new syllable with repetition. Severely stuttered speaker find it more repeated to a particular Tamil short syllable.

The combined Tamil syllable Fig.10 also shows that severely stuttered people extends their repetition of the same syllable only for a short duration of time and recovers as much quick as a moderate speaker. This method also provides a key point on the time scale at which this recovery subsists.



(a)



(b)

Fig.9. Tamil long (a) and short (b) syllables vocal tone frequency (minima)

Retrieval Number: B3211078219/19©BEIESP
DOI: 10.35940/ijrte.B3211.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

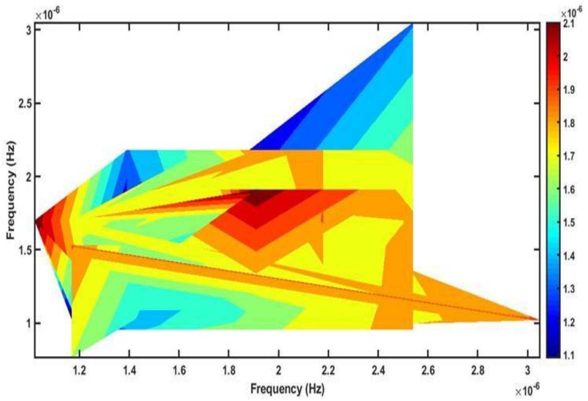


Fig.10. Selected 20-Tamil long and short syllables vocal tone frequency (minima)

V. CONCLUSION AND FUTURE WORK

Analysis and diagnosis of dysfluent speech are very essential for the speech pathologist in order to identify the severity level of stuttered speech. This research paper has proposed a detailed analysis of vocal tone of different speakers, particularly for the Tamil syllable, which enables the understanding of speech visually itself by comparing fluent speech, moderate stuttered speech and severe stuttered speech. The detailed outcomes of the analysis were concluded as follows:

a) The vocal tone analysis of the speech spectrums of different speaker viz. normal speaker, moderately stuttered speaker and severely stuttered speaker were performed subjecting to the criteria of similar selective syllable from Tamil literature.

b) It was observed that the severe stuttering speaker has expressed more short frequency variations and the fluctuations in time interval during their speech.

c) The acoustic analysis of severe and moderate stuttering children will provide useful in-depth analyses of stuttering manifestations at its early stage.

d) The result depicts the stuttering level of a moderate speaker exposed with Tamil syllable performing better with non-repeating syllables whereas severely stuttered spectrum finds more uncomplicated with the same.

e) All three speakers find it easier to repeat the long Tamil syllable but short syllable shows poor recital that depicts the difficulty especially for the moderate and severely stuttered speakers.

f) Overall performance of the speakers without differentiation also shows that stuttered speakers construct difficulties in reproducing the spectra as a normal speaker.

g) The characteristic recovery time required for severely stuttered speaker to outperform as normal depends on the type of syllable in Tamil language.

The pathologist on detecting and diagnosing the nature of stuttered phenomena with respect to the level of disorder will enhance the performance of patients getting more open to interactions with other persons and their social isolation decreases. It was remembered that most of the patients were school children. So it is of utmost importance that difficulties in learning should be eliminated in the earliest stage of their education and also based on the vocal tone of analysis pathologist can easily diagnose the dysfluent speakers. The proposed vocal tone analysis will be strengthened in future by extending the analysis to formant which can help substantially to identify the depth of the severity level and also enhance the level of understanding the stuttering phenomena.

ACKNOWLEDGMENT

This work is supported by Science and Engineering Research Board (SERB) - Department of Science and Technology (DST) under grant (No. EEQ/2016/000110) "*Technology development for the support and improvement of speech language impaired people*". The authors also thankful to Office of the Registrar General and Census Commissioner, New Delhi, India for providing the census data on speech disorders.

REFERENCES

1. A.Gilat, *MATLAB An Introduction With Applications*, John Wiley and Sons, 2004.
2. Andrzej Czyzewski, Andrzej Kaczmarek and Bozena Kostek (2003), "Intelligent Processing of Stuttered Speech", *Journal of Intelligent Information Systems*, vol.21, no.2, pp.143–171.
3. Barbosa, F and Silva W (2015), "Support vector machines, Mel-Frequency Cepstral Coefficients and the Discrete Cosine Transform applied on voice based biometric authentication", In the proceedings of *IEEE SAI Intelligent Systems Conference (IntelliSys)*,London, UK, pp. 1032-1039.
4. Census of India 2011, Office of the Registrar General and Census Commissioner, New Delhi, India (27-12-2013), http://www.censusindia.gov.in
5. Chagit E. Clark, Edward G. Conture, Tedra A. Walden and Warren E. Lambert (2013), "Disorders Speech sound articulation abilities of preschool-age children who stutter", Journal of Fluency Disorders, vol. 38, pp. 325–341.
6. Derakhshan, N., and Savoji, M. H (2006),"Perceptual Speech Enhancement Using Hilbert Transform", In the proceedings of *IEEE International Symposium on Industrial Electronics*, Montreal, Quebec, Canada, pp.681-684.
7. E. S. Gopi, *Digital Speech Processing Using Matlab*, Signals and Communication Technology Springer, 2014.
8. E. Szabelska, A Kruczyńska (2013), "Computer-based speech analysis in stutter", Applied Computer Science, vol. 9, no.2, pp.34-42.
9. Elham Masumi, Zohre Arani Kashani, Nafise Hassanpour and Mohammad Kamali (2015), "The Effect of Syllable Structure on the Frequency of Disfluencies in Adults with Stuttering", *Middle East Journal of Rehabilitation and Health*, vol.2, no.2, e26436, pp.1-5.
10. Grazina Korvel, Olga Kurasova and Bozena Kostek(2018), "Comparative Analysis of Spectral and Cepstral Feature Extraction Techniques for Phoneme Modelling", In the proceedings of *Multimedia and Network Information Systems*, pp. 480–489.
11. He, Lei; Dellwo, Volker (2016), "A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform", In the proceedings of *INTERSPEECH*, San Francisco, pp. 530-534.
12. IngYann Soon, SooNgeeKoh and Chai Kiat Yeo (1998),"Noisy speech enhancement using discrete cosine transform", *Speech Communication*, vol. 24, pp. 249-257.
13. Jason H. Davidowa, Anne K. Botheb and Jun Ye (2011), "Systematic studies of modified vocalization: Speech production changes during a variation of metronomic speech in persons who do and do not stutter", *Journal of Fluency Disorders*, vol. 36, no.2, pp. 93–109.
14. Jingdong Chen, Kuldip K. Paliwal and Satoshi Nakamura (2000), "A block cosine transform and its application in speech recognition", In the proceedings of Sixth *International Conference on Spoken Language Processing*, Beijing, China, pp.117-120.
15. Keng-hao Chang (2012), "Speech Analysis Methodologies towards Unobtrusive Mental Health Monitoring", *UC Berkeley*. ProQuest ID: Chang_berkeley_0028E_12254.
16. Leanne Jani, Mark Huckvale and Peter Howell (2013), "Procedures used for assessment of stuttering frequency and stuttering duration", *Clinical Linguistics & Phonetics*, vol.27, no.12, pp.853–861.

17. N. Thrane (1984), *The Hilbert transform*, Technical Review(Brüel&Kjær), no. 3, pp. 3-15.
18. Norimune Kawai, E. Charles Healey and Thomas D. Carrell (2012), "The effects of duration and frequency of occurrence of voiceless fricatives on listeners' perceptions of sound prolongations", *Journal of Communication Disorders*, vol.45, pp. 161–172.
19. Peter Howell and Louise Vause (1986), "Acoustic analysis and perception of vowels in stuttered speech", *The Journal of the Acoustical Society of America*, vol.79, no.5, pp.1571-1579.
20. Peter Howell and Mark Williams (1992), "Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech", *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp.1697-1706.
21. PetreStoica and Randolph Moses, *Spectral Analysis of Signals',* Prentice Hall, 2005.
22. Syed Mohammad Ali and Pradeep Tulshiram Karule (2013), "Spectral Analysis of Pathological & Normal Speech Signal", *International Journal of Scientific & Engineering Research*, vol.4, no, 2, pp.1-4.
23. http://www.southasia.sas.upenn.edu/tamil/tamilsound.html

AUTHORS PROFILE

Manjutha Manavalan has obtained her Masters in Computer Science from Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India. She is pursuing her Ph.D., in the Department of Computer Science, from same Institute. Her research interest includes Speech Analysis and Recognition, Speech signal, Image and Video Processing.

Dr. Parthasarathy Subashini is working as Professor in the Department of Computer Science, Avinashilingam University, Coimbatore, India. Professor Subashini's research has spanned a large number of disciplines like Image analysis, Pattern recognition, Neural networks, and Computational Intelligence. She has authored or co-authored 2 Books, 4 Book chapters, 1 Monograph, 145 papers, including IEEE, Springer's in various international, national journals and conferences. To add up her credit, recently she has contributed a Nature Inspired Optimization algorithm called 'Synergistic Fibroblast Optimization(SFO)'. The codes and published articles are publicly available for the perusal of the Research community in MATLAB central, Researchgate and Adademia.edu.

Dr. M. Krishnaveni is a Assistant Professor of Department of Computer Science, Avinashilingam University for Women, Coimbatore. She has 6 years of research experience working as a researcher in Naval Research Board, Defence Research Development Organisation, Ministry of Defence. She was been a consultant for Military related application project. She has published several papers in her research specialization with an overall count of about 71 both nationally and internationally. She focuses on research and experimental processes with the use of digital technologies such as: MATLAB, Computer Science, Biological computation along with machine intelligence and real-time physical computation. She also acts as a resource person and holds national and international funded research projects which comes under University Grants Commission (India) and Department of Science and Technology and DRDO-NPOL. She is a recipientawards such as bestyoung teacher award IASTE 2017,best NSS Programme officer award ,NYLP 2016,Goverment of India