Available online at www.elixirpublishers.com (Elixir International Journal)

Electronics Engineering

Elixir Elec. Engg. 85 (2015) 34608-34615

ISSN: 2229-712X

Evaluation of Vocal Folds Speech Disorder with Special Focus on Normalized Pitch Variation Characteristics

Manasi Dixit[1] and Shaila Apte[2]

[1]Department of Electronics K.I.T.'s College of Engineering, Kolhapur, 416234 Maharashtra, India.

[2]Rajarshi Shahu College of Engineering, Pune, Maharashtra, India.

ARTICLE INFO

Article history:

Received: 14 June 2015;

Received in revised form:

22 August 2015;

Accepted: 26 August 2015;

Keywords

Normal and Vocal Folds disordered Speech data,

Pitch variations, Jitter,

Shimmer, HNR,

Percentile f_0 Characteristics,

% CQ –Characteristics,

All Tx –Regular Tx Pitch Period Characteristics.

ABSTRACT

In this work the speech disorder due to damaged Vocal Folds is discussed. The basics of the disorder are described here. The speech parameters are identified which characterize the disorder and correction system is designed to improve the speech quality. The speech signal samples of people of age between five to eighty years are considered for the present study. These speech signals are digitized and enhanced and analyzed for the Jitter, Shimmer, HNR, Pitch variations Tx graphs, Normalized Percentile f_o characteristics and % Close Quotient EGG graphs using MATLAB, PRAAT, SFS and EXCEL platforms.

Introduction

Vocal Pathologies arise due to accident, disease, misuse of the voice, or surgery affecting the vocal folds and have a profound impact on patients' life. The modeling of normal and pathological voice source and the analysis of healthy and pathological voices is done in this work.

Most voice-related pathologies are due to irregular masses or the presence of pathologies in the larynx such as vocal fold nodules or vocal fold polyps located on the vocal folds interfering in their normal and regular vibration [1]. This phenomenon causes a decrease in voice quality, that is, usually the first symptom of this type of disorders.

Literature Review

Anandthirtha B. Gudi et.al. have developed a new speech measure which is based on parameterization of the autocorrelation envelope of the AM response for vocal fold pathology assessment. Voice conversion system is created to change the perceived speaker's identity of a speech signal, which is based on converting the LPC spectrum and predicting the residual as a function of the target envelope parameters. [1].

Zoran C´ irovic et.al. used EGG features representing the time characteristics of an idealized EGG waveform, and then concatenate both the EGG features and audio features by applying a glottal activity detector [2]. The Electroglottograpgh is a device for the measurement of the time variation of the degree of contact between vibrating vocal folds during voice production. The degree of contact is proportional to the impendence between two electrodes on the subject's neck [2].

Constantine Kotropoulos et.al. have discussed two distinct two-class pattern recognition , the detection of vocal fold paralysis in male subjects and the detection of female subjects suffering from vocal fold edema. [3].

Tomoki Toda et.al. have described a novel statistical approach to the vocal tract transfer function (VTTF) estimation of a speech signal based on a factor analyzed trajectory hidden Markov model (HMM). Speech is a quasi-periodic signal, there are many missing frequency components between adjacent fo harmonics. The proposed method determines a time-varying VTTF sequence based on the maximum a posteriori (MAP) estimation [4]. Kumara Sharma et.al. have proposed Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of speech as Acoustic Indicators of Laryngeal and Voice Pathology [5].

Daryush Mehta has discussed about Aspiration Noise during Phonation: Synthesis, Analysis, and Pitch-Scale Modification. The study investigates the synthesis and analysis of aspiration noise in synthesized and spoken vowels based on the linear source-filter model of speech production [6].

Arpit Mathur et.al. have discussed about the significance of parametric spectral ratio methods in detection and recognition of whispered speech [7].

Voice Quality Analysis

Voice production starts with the vibration of the vocal folds, which can be more or less stretched to achieve higher or lower pitch tones. In normal conditions and in spite of this pitch variation ability, phonation is considered stabilized and regular. Any transformation on the vocal fold's tissue can cause an irregular, nonperiodic vibration which will change the shape of the glottal source signal from one period to the next, introducing jitter [8]. The same problem can occur in amplitude. If, for instance, the vocal folds are too stiff, they will need a higher sub glottal pressure to vibrate. The glottal cycle can thus be irregularly disturbed also in amplitude, originating shimmer. The possible existence of high frequency noise, especially during the closed phase of the glottal cycle, originated by a partial closure of the vocal folds will cause an air leakage through the glottis, providing a turbulence effect. All these phenomena affect the glottal source signal. [8]. In this work a non-interactive approach

Tele:
E-mail addresses: manasidixit64@gmail.com

is taken without consideration of the influence of the supra glottis vocal tract or the influence of the sub glottis cavities on the glottal flow. As a consequence, it is assumed that the source and filter parameters are independent.

Details of the vocal tract and the utterances with respect to different articulators [9] are indicated in Figure 1.

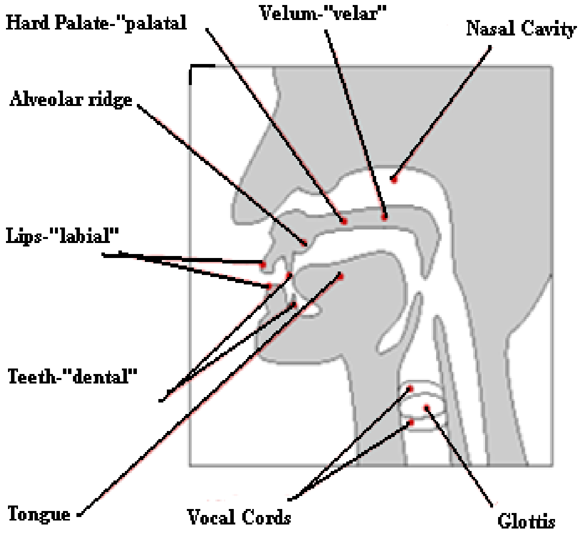

Figure 1. Vocal Tract indicating Place of different categories of Speech Production

Frequency Perturbation

The vocal fundamental frequency is reflective of the biomechanical characteristics of the vocal folds as they interact with sub glottal pressure. The biomechanical properties are determined by laryngeal structure and applied muscle forces. Adjustment of the latter, in turn, is a function of reflective, affective and learned voluntary behaviors.

The small cycle-to-cycle changes in frequency are represented by frequency perturbations or jitter [10]. The estimation of irregularities in the vibration of the vocal folds is commonly measured by the jitter parameter. Jitter measures the irregularities in a quasi-periodic signal. The jitter of a voiced speech signal is usually taken as a measure of the change in the duration of consecutive glottal cycles. Jitter is the instability of vocal fold vibration. Variations of pitch perturbation are inversely proportional to the frequency, for example when fundamental frequency increases, jitter decreases [11]. Variations of pitch perturbation is expected to change in relation to the degree of tension present in the vocal folds, where high tension reflects lower perturbation values and low tension reflects higher perturbation values. The higher jitter ratios are related to low pitch and intensity of sustained vowels. The Vocal intensity is dependent on an interaction of sub glottal pressure and the adjustment status and aerodynamics at the level of the vocal folds, as well as vocal tract status.

Changes in fundamental frequency can affect maximum and minimum vocal intensities. The fundamental frequency is raised when asked to speak louder. Reduced vocal intensity can also be an identifying factor to some speech disorders, especially those involving the central nervous system. Values of amplitude perturbation or "shimmer" are similar to jitter values. Shimmer measures the cycle-to-cycle variations of the vocal signal. Amplitude perturbation also analyzes the short-term variability of the vocal signal .Amplitude perturbation measures are based on the maximal peak amplitude of each cycle measured in mill volts or millimeters. That the forces of vocal fold tension, mass, length and subglottic pressures can affect shimmer as well as jitter scores [8]. Shimmer is also better analyzed in prolongations of single sounds rather than in spontaneous or connected speech.

The vocal folds are driven by the cranial nerves originating within this system. If the cranial nerves which innervate speech articulators are destroyed or damaged, then the transmission to the vocal folds will be disrupted. This subsequently may cause changes in the vocal fold vibratory pattern [8]. Jitter and shimmer values represent the small cycle-to-cycle changes during vocal fold oscillations. Speech portion of large duration ≥ 10ms is used to calculate frequency, jitter and shimmer. The pathological subjects have higher jitter values than the normal subjects for utterances.

Vocal Tract Model

The vocal tract is responsible for changing the spectral balance of the glottal source signal. By changing the vocal tract shape the speaker can modify its resonance frequencies to produce a wide variety of different sounds [12]. Humans use the evolution in time of the resonance frequencies to produce speech .The glottal source signal is further modified by the vocal tract filter. The contribution of the vocal tract resonances can be removed from the speech signal by inverse filtering.

HNR Estimation

A sustained vowel can be assumed to be periodic signal .The signal-to-noise ratio for a periodic signal is equal to the harmonics-to-noise ratio (HNR). According to the threshold values declared by Praat software a healthy normal voice should have an HNR of 20 for the phonation of /a/ or /i/ vowels, and an HNR of 40 for the phonation of the vowel /u/ [13]. The pathological subjects have lower HNR values than the normal subjects for utterances and hence the intelligibility of vowels or consonants produced by them are very low.

System Implementation

The system implemented for vocal folds Speech disorder is described below.

Procedure

The present work attempts to find few parameters from pathological speech for confirmation of Vocal Folds disorder. Researchers have used lot many parameters. We try to reduce the computational cost and reduce the number of parameters. The present work is based on study of children, some adult male and female subjects and some elderly male and female subjects speaking Marathi as their mother tongue. The speech data of normal subjects and pathological subjects of the same age group between 8 to 85 years is collected.

Database

The standard database is not available. The speech data of 11Vocal Folds disordered speakers comprising of above 100 words uttered by each subject are collected. The speech database consists of isolated words, connected words, fast uttered sentences and songs for e.g. -School-Prayer, National anthem and Pledge ,Nursery Rhymes ,famous film songs etc. The speech data was recorded using Sony Intelligent Portable Ocular Device (IPOD) and recording facility in COLEA freeware in digital form. The recording was carried out in a pleasant atmosphere and maintaining the children and other subjects in tension-stress free environment. The recorded signal is transformed into '.wav 'file by using GOLDWAVE software. The data was collected at Chetana Vikas Mandir, a special school established to educate Mentally Retarded children as well as children with various disorders. It is located at Kolhapur, India. The data is also collected from the patients under the treatment of speech therapists and ENT specialists in Kolhapur city. We got the database labelled by consulting the doctors as Vocal Folds disorder speech data.

Evaluation of Speech

The present work attempts to confirm the Vocal Folds disorder from the speech signal by extracting only important segmental and supra segmental acoustic indices. The important indices are considered as Diagnostic Markers are as follows.

1. Evaluation of Fundamental frequency f_o, Jitter, Shimmer, and HNR for the analysis of harshness and breathiness in the voice to be done in the Training Phase to confirm the speech as of pathological speech category. Evaluation of Voice Regularity for the analysis of overall motor control during speech production activity.

2. Fundamental Frequency Analysis- The fo mean value lies in the high range between 180 Hz to 440 Hz for adult male female speakers and children.

3. Glottal frequency variations – the source frequency fo variations with respect to percentile values from 0 to 100. The Linear or nonlinear nature of the characteristics is important. The Vocal Folds disorder is confirmed by low gradient index in 0 to 50 fo percentile range and high gradient index and nonlinear nature of the characteristics in 50 to 100 fo percentile range. % Close Quotient (CQ) graph simulates the Laryngograph and indicates the close phase of Glottis pulse signal or vocal folds vibrational cycle. The mean CQ, range of variation of CQ and CQ variations with respect to total time duration of speech sample are important parameters. f0

System Block Schematic

Diagram of the system implemented for evaluation of Vocal Folds disorder is as shown in Figure 2.

Pre-Processing

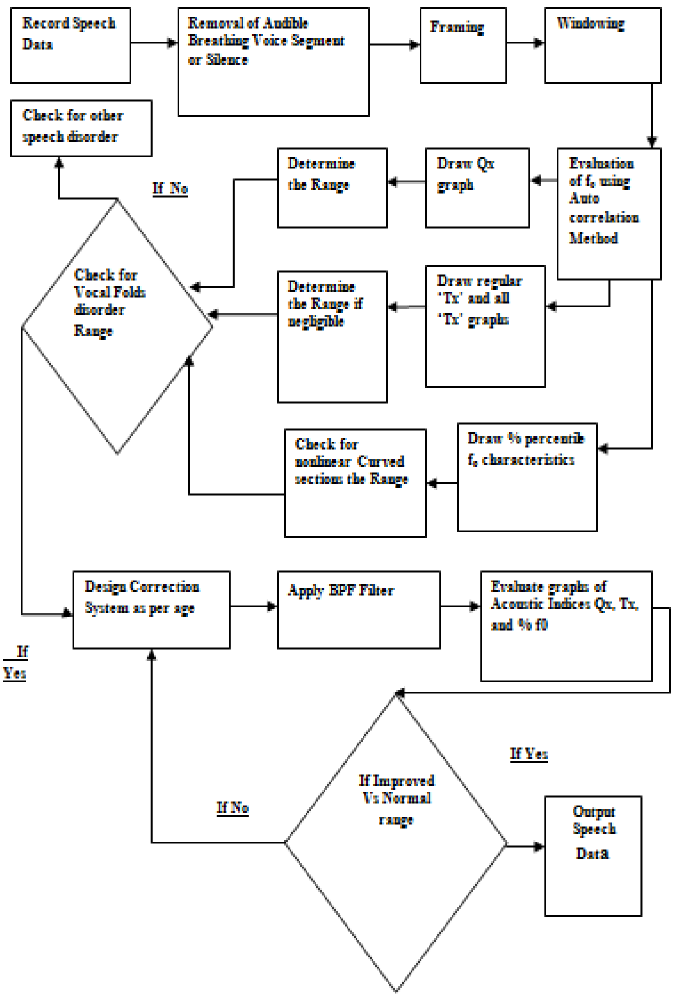

Figure 2. Block Diagram for Evaluation of Vocal Folds Disorder

System Development

The system is developed using two modes training mode and testing mode.

Training Mode- In training mode 50 speech samples are used to train the system. The Laryngograph comprising of % Close Quotient (CQ) with respect to time indicative of the close phase of Glottis pulse signal or vocal folds vibrational cycle and the comparison of histograms of all 'Tx ' which is all pitch periods with respect to regular 'Tx' pitch periods are plotted for these 50 speech samples to confirm Vocal Folds disordercharacteristics. The observations for Tx are as follows.

1. A pitch period is defined as regular if it deviates in duration in the range less than +/- 10% from the successive time periods before or after it. In case of normal persons, regular Tx graph matches with all Tx graph.

2. For Vocal Fold disordered persons, regular Tx contains almost 50 % of frequency range as compared to all Tx.

Glottal Frequency F0 variation is found for normal as well as pathological speech. It is evaluated in transformed percentile domain. The fo values for a speech data file are calculated by using Framing and Windowing Algorithm in MATLAB. Then the Normalization routine is developed in Microsoft Excel. The algorithm for evaluation of Normalized fo Variation in Percentile Domain is as follows.

1. The percentile value of f_0 maximum level is considered to be 100 percentile.

2. Hence according to the data values of f_0 variations as per the speech sample 0-5-10-15-----95-100 percentile values of every speech sample are calculated.

3. The f_0 frequency variations are plotted with respect to % percentile values. This graph provides a very good measure of fundamental frequency analysis to differentiate between normal speech and pathological speech. It also confirms Vocal Folds disorder.The graph is linear for certain range with high or low gradient and it exhibits curved nature in the other range of fo for Vocal Folds disorderedspeakers. For normal speech percentile fo track graphs are linear from 5 to 95 range with very low gradient 0.05 to 0.3

The observations for CQ are as follows.

1. CQ is a time Vs CQ graph. It indicates that the CQ values are present in the range 15 % to 75 % for Vocal Folds disordered speakers.

For normal speech, CQ variation is observed in the range 15 % to 60 %

The observations for Regular Tx pulses are as follows.

1. % Regular Tx pulses is less than 50 % or sometimes even less than 50 % of the frequency spectrum as compared to all indicating loss of regular pitch periods.

Graphs of some of the Diagnostic Markers

Figure.3 to Figure 5 show the all Tx and regular Tx graphs for 3 disordered speakers. The observations are indicated in the notes. The general observation is that the time duration of regular Tx pulses is less than 50% of All Tx graph and it is sometimes even not produced at all.

Diagnostic Markers – dudh1.wav

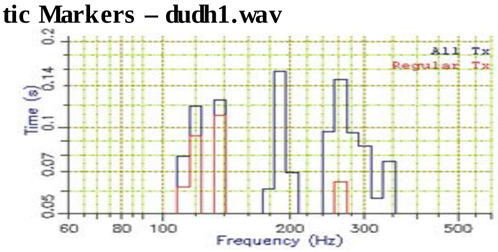

Figure 3. % Regular Glottal Pulses dudh1

Notes

The regular Tx pulses are produced in the 110 Hz to 140 Hz and 258 Hz to 278 Hz pitch spectrum. The time duration of regular Tx pulses is almost 50% of All Tx graph.

Diagnostic Markers-dr1.wav

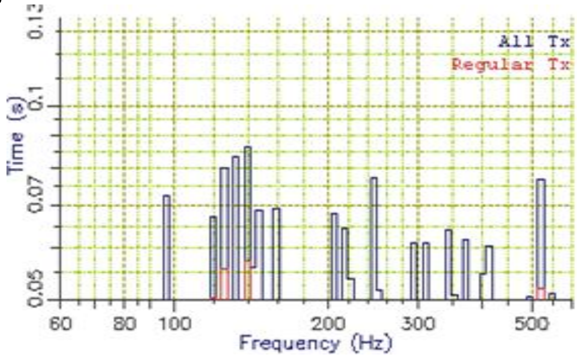

Figure 4. % Regular Glottal Pulses -dr1

Notes-1. The regular Tx pulses are produced in the 120 Hz to 140 Hz and 500 Hz to 510 Hz pitch spectrum. The time duration of regular Tx pulses is almost 30% of All Tx graph

Diagnostic Markers-par 16.wav

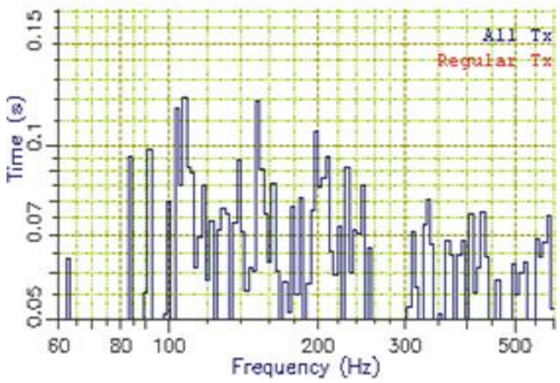

Figure 5. % Regular Glottal Pulses -par 16

Notes

1. The regular Tx pulses are not produced in the entire 60 Hz to 600 Hz pitch spectrum.

Figure.6 to Figure 8 show the % CQ graphs for 3 disordered speakers. The observations are indicated in the notes. The general observation is that % CQ graphs indicate that Qx pulses spread in 15% to 77% CQ range.

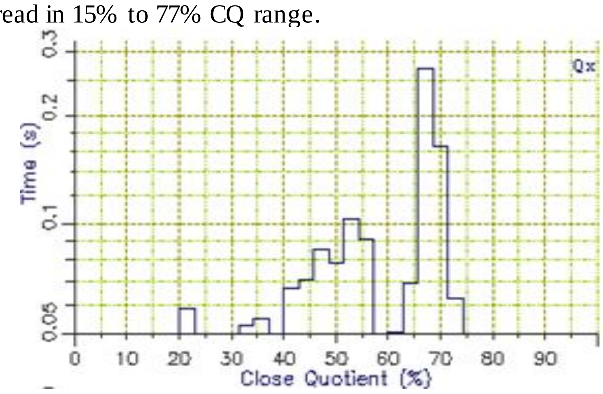

Figure 6. % Close Quotient wrt Time for speaker dudh1

Notes

% Qx graph indicates set of short Qx pulses spread in 20% to 75% CQ with mean value %CQ – 54.7

Figure 9 to Figure 11 show the percentile f_0 graphs for 3 disordered speakers. The observations are indicated in the notes. The general observation is that percentile f_0 graphs indicate that it has low index of < 1 between 5 to 50 and high index of > 3 between 55 to 95 . Sometimes we observe that the high gradient is seen for 5 to 95.

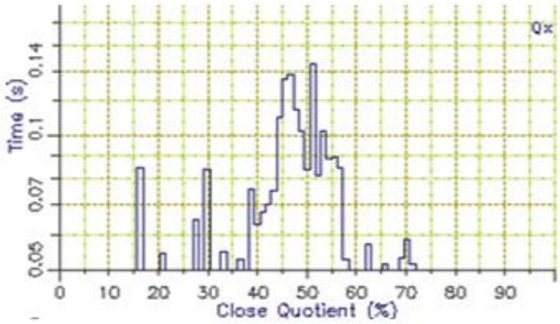

Figure 7. % Close Quotient wrt Time –dr1

Note-% Qx graph indicates set of short Qx pulses spread in 15% to 72% CQ with mean value %CQ – 44.5

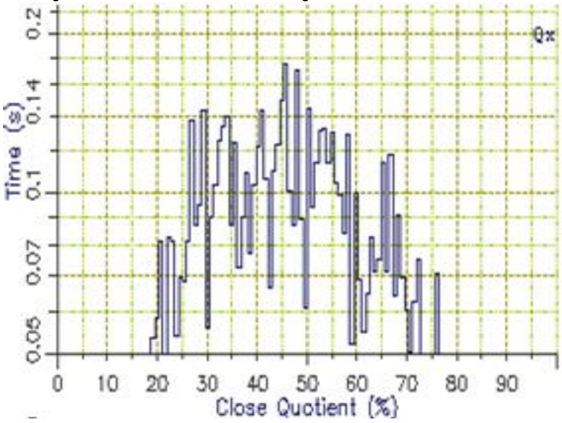

Figure 8 . % Close Quotient wrt Time-for speaker par 16

Note- % CQ graph indicates Qx pulse 18% to 77% wide with CQ with mean value %CQ – 46.2

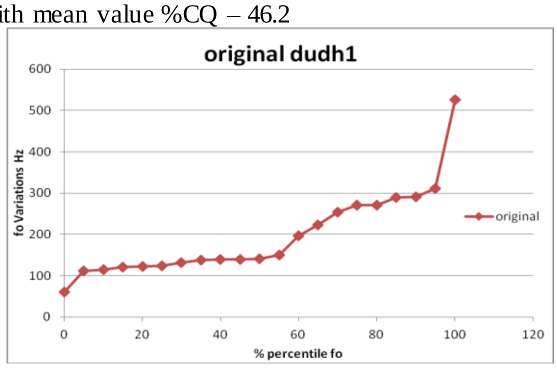

Figure 9. Percentile fo track variations for speaker dudh1

Note-Percentile fo track indicates initial linear section from 5 to 50 percentile range with 0.07 -low gradient index, and second nonlinear curved section graph from 55 to 95 percentile range and from 95 to 100 range– 4.31 very high gradient index

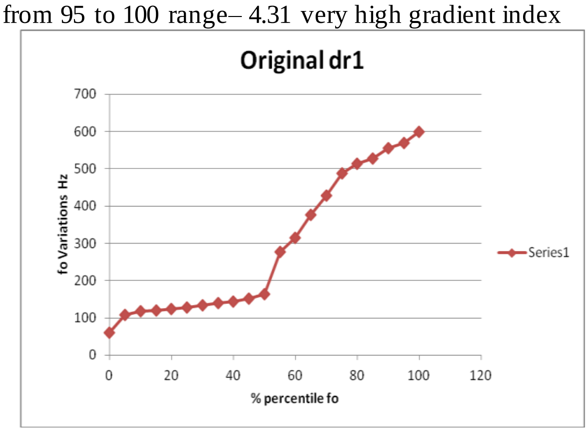

Figure 10 Percentile fo track variations for speaker dr1

Note- Percentile f_o track indicates linear graph from 5 to 50 percentile range –0.12 low gradient index, nonlinear curved graph from 55 to 95percentile range with very high gradient

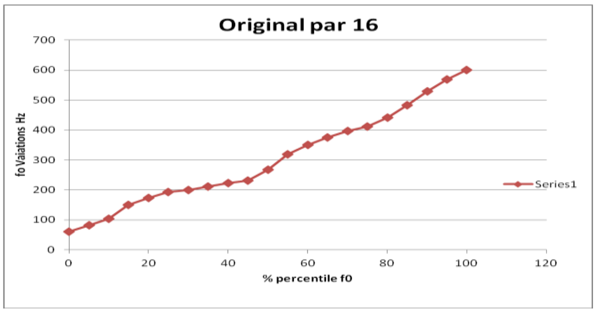

Figure 11. Percentile fo track variations for speaker par16

Note-Glottal frequency fo Characteristics in Percentile Domain indicates continuous rising locally nonlinear graph from 0 to 100 percentile range –0.55 high gradients

Testing Mode- In testing mode remaining 50 speech samples are used for confirmation of Vocal Folds disorder. The testing mode checks the Laryngograph characteristics, CQ graph and F0 track to confirm Vocal Folds disorder. The following observations are made.

Fundamental Frequency f_o mean is in the range 120 Hz to 440 Hz as per the categories like adult male, female, children or elderly speakers.

1. It is observed that Laryngograph comprising of regular Tx pulses contains less than 50 % or even negligible of the frequency spectrum as compared to all Tx.
2. Time Vs closed quotient graph indicates closed quotient range more than 50 % wide with mean %CQ within 42% to 48%.
3. f_o track has a mixed linear and nonlinear curved nature with high gradient in higher percentile range.

System Developed for correction –

We have used MATLAB based open source developmental tool E-System compatible with SFS developmental software for trying methods for correcting Vocal Folds disordered speech. The developmental tool COLEA and Adobe Audition are used for preprocessing the speech samples. In preprocessing the silence zone or the audible breathing voice segments are removed. Using E-System following processing blocks can be designed.

- Amplifier /Attenuator – Design specifications are gain and Bandwidth
- Low Pass Filter- Design specifications are Cut off Frequency
- High Pass Filter- Design specifications are Cut off Frequency
- Band Pass Filter- Design specifications are lower and upper Cut off Frequency
- Band Stop Filter- Design specifications are lower and upper Cut off Frequency
- Vocal Tract Filter- Design specifications are f_1,f_2 and f_3 formant Frequencies.
- Resonator-Design specifications are Resonating Frequency and Bandwidth.

The system applied for correction is developed with the help of following filters.

Band Pass Filter- The lower cut off frequency is in the range 10 Hz to 100 Hz and the upper cut off frequency selected should be such that the second formant frequency f_2 should lie in the pass band. Hence it is selected as 1500 Hz, 2000 Hz or 2500 Hz as per male, female or children based on pitch frequency range.

Resonator - The resonating center frequency selected should be such that the second formant frequency f_2 should lie in the pass band. Hence it is selected as 1500 Hz, 2000 Hz or 2500 Hz as per male, female or children based on pitch frequency range.

Vocal Tract Filter – It is realized as a cascaded combination of three resonators acting as per three formant frequencies. The standard adult male formant frequencies are 500 Hz, 1500Hz, and 2500 Hz. The first formant frequency is amplified by 20 dB, the second formant frequency is amplified by 10 dB and the third formant frequency is maintained at 0 dB. Hence this filter boosts up the input speech signal spectrum as per the formant frequencies. In case of pathological speech the amplitudes of upper formants are degraded .Hence the VTF is the better solution to lift up the second formant spectrum.

System Applied for Correction of Vocal Folds disorder

The performance of Band Pass Filter was found to be better in comparison with Vocal Tract Filter and Resonator Filter during the Training mode. Hence the BPF is applied for correction of Vocal Folds disorder speech. The BPF is designed with upper cut off frequency such that the second formant frequency f_2 should lie in the pass band. Hence it is selected as 1500 Hz, 2000 Hz or 2500 Hz as per male, female or children based on pitch frequency range.

After applying the BPF to Vocal Folds disorder the following observations are made.

4. It is observed that Laryngograph comprising of regular Tx pulses graph shows improvement and contains more than 50 % of the frequency spectrum as compared to all Tx
5. Time Vs closed quotient graph indicates closed quotient range more than 50 % wide with mean % CQ within 42% to 48%.
6. Percentile fo track has a low gradient (slope)- In 10 to 50 percentile range – 0.17 to 0.24 and a high gradient (slope) in 55 to 100 percentile range –0.55 to 1.94. The nonlinear curved section is improved and converted to approximately linear with comparatively low gradient levels.

Improvement in the diagnostic markers due to application of BPF Filter is indicated below with the help of Regular Tx graph, % Close Quotient (CQ) graph and % percentile f_o track graphs through Figure 12 to Figure 18 for speaker dudh1. Figure 12 shows the graph of regular Tx after filtering using BPF. It is seen that the regular pulses are now produced in 110 Hz to 160 Hz range. Figure 14 shows the graph of % CQ after filtering using BPF. It is seen that the range of CQ is now varies from 15 to 60. Figure 16 shows the graph of percentile f0 track variation after BPF. It is observed that the nonlinear curved graph from 55 to 85 percentile range is now smooth. It becomes approximately linear from 5 to 85 percentile range with 0.06 very low gradients.

Diagnostic Markers-for speaker dudh 1BPF.wav

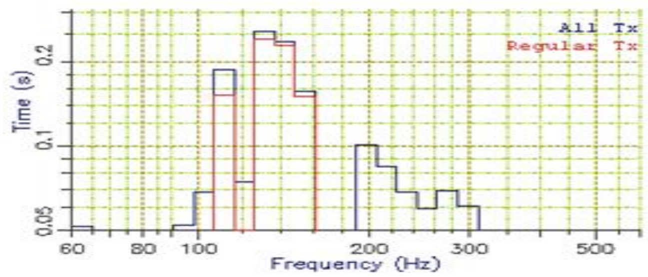

Figure 12. % Regular Glottal Pulses-dudh1 BPF processed

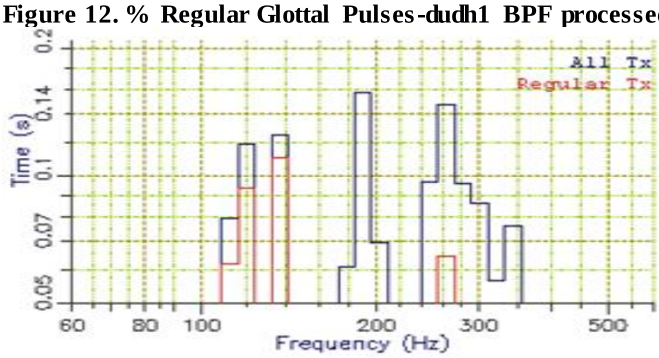

Figure 13. % Regular Glottal Pulses-dudh1 Original

Notes-

The regular Tx pulses are produced in the 110 Hz to 140 Hz and 260 Hz to 275 Hz pitch spectrum in original where as in 110 Hz to 160 Hz pitch spectrum only in BPF processed graph.

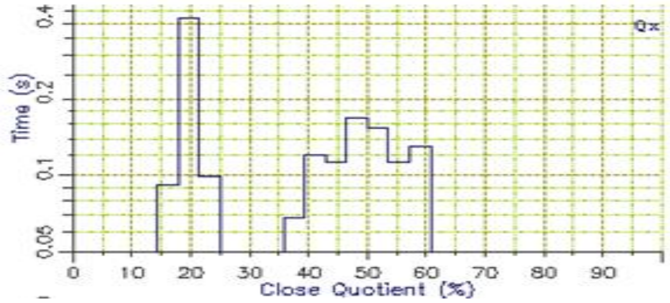

Figure 14. % Close Quotient wrt Time for speaker dudh1 BPF Processed

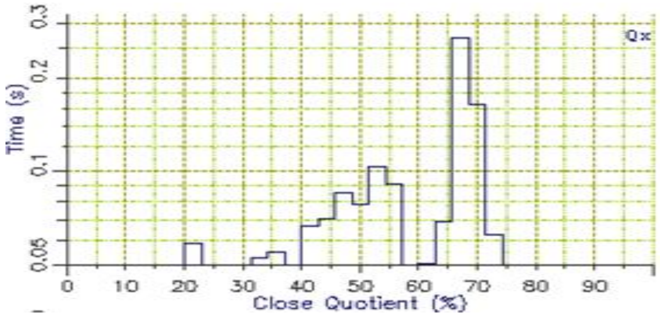

Figure 15. % Close Quotient wrt Time dudh1 Original

Notes-

The % CQ graph lies within 20 to 75% range in Original where as it lies in 15% to 60 % range in BPF processed.

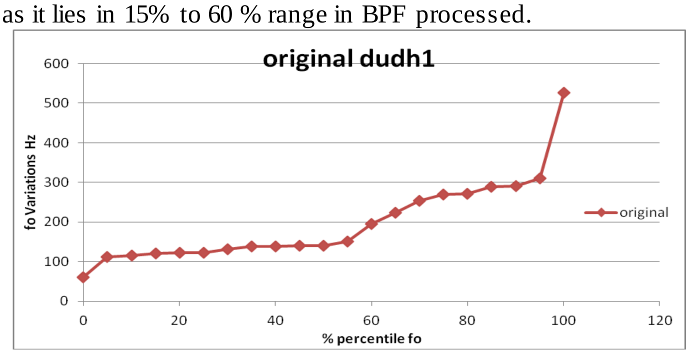

Figure 16. Percentile fo track variations –dudh1 Original

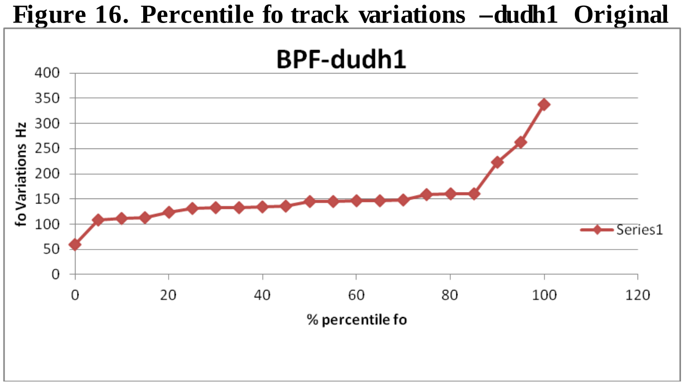

Figure 17. % Percentile fo track variations –dudh1 BPF Processed

Notes- Original **percentile fo track indicates** linear graph from 5 to 50 **percentile range** –0.077 low **gradient indexes. In BPF processed the** nonlinear curved graph from 55 to 85 **percentile range is** smoothed, approximately linear from 5 to 85 percentile range with 0.06 very low gradient.

Figure 18 shows comparison of fo track variation using different filters. It is observed that Band Pass Filter (BPF) performs better giving larger linear range with a low gradient.

Percentile of Track Comparison dudh1 Vocal Folds Disorder

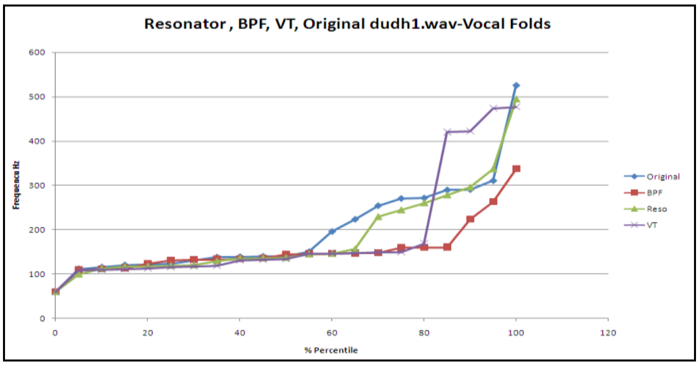

Figure 18. Comparative Response of the Correction System for Vocal Folds Disorder

Segmental and Supra segmental Acoustic indices Analysis

The analysis of segmental and supra segmental acoustic indices was carried out for particular isolated words and continuous speech data. The isolated word data above 100 words uttered by each of 25 normal subjects and12 Vocal Fold disabled subjects were analyzed and reference /threshold level was considered for each isolated word. Various Misarticulation or wrong pronunciation cases were analyzed in case of pathological subjects. The spectrograms were studied for the purpose of formant analysis. Continuous speech exhibits greater complexities and it is less intelligible.

Classification of Speech on the basis of Segmental and Supra segmental acoustic indices

The following parameters are extracted by using the developmental tool available as an open source software SFS and PRAAT from every speech data to classify whether the speech is Normal or Pathological. The Normal range and Pathological range of the parameter values are as below.

Considering the observations for % CQ variation, Tx variation and fo track variation observations the system is designed for confirmation of Vocal Folds Disorder. Results for training system are shown in Table 2.

Results for testing of remaining 50 % samples are shown in Table 3 as % samples confirmed for Vocal Folds Disordered speech and normal speech. The results for the processing using BPF are shown in Table 4. The table indicates that all parameters are improved using BPF.

Observations

The existing speech enhancement algorithms like spectral subtraction do not help in enhancement of pathological speech . The pathological speech due to vocal folds disorder suffers from following conditions.

- Breathing voice segments are audible in speech because the subjects are under stress when they speak .When the speakers are supposed to take pause in between utterances of two successive words generally the breathing voice segment is heard.
- The minimum intensity level does not drop much as there is no silence region due to the presence of breathing voice segments.
- The speakers have to put more efforts for the motor movements of articulators. Hence the utterances of different words are not appropriate.
- Due to low HNR levels below the pathological threshold of 12 dB the speech indicates harshness.

Our Contribution to present work

The Vocal Folds disordered speech database is not available.

1. We got the database labeled by the doctors.

Table 1. Range of Segmental and Supra segmental Acoustic Indices for Classification of Normal Speech and Pathological Speech

Parameter	Normal Speech Range			Pathological Speech Range
Fundamental Frequency-Pitch	Children	Adult Male	Adult Female	for VPD* 300 -550 Hz otherwise same as Normal Speech
	208-440 Hz	85-196 Hz	155-334Hz	
Jitter (mean)	Range 0.0 % to 18%			Range 14% to 45%
Shimmer (mean)	Range 0.0 % to 5 %			Range 0.0% to15%
HNR (mean)	Range 12 dB to 45 dB			Range 5dB to 11 dB
Voice Regularity	Range 50% to 95 %			Range 5 % to 45%
Audible Breathing Voice Segment	Not Present			Always Present

Table 2. Range of diagnostic markers for 50 % of Vocal Folds Disordered speech samples and 50 % of normal speech

Diagnostic marker	Range of values for Vocal Folds Disordered speech	Range of values for normal speech
Time Vs % CQ	15 % to 75 %	10 % to 60 %
Time periods Vs frequency TX graph	Regular Tx has less than 50 % or negligible frequency range as compared to all Tx graph with intermittent pulses	Regular Tx matches with all Tx for more than 90 %.
Percentile Glottal Frequency fo track variation	fo track has linear nature in initial range and nonlinear, curved nature in higher range in percentile domain	Linear Characteristics with very low gradient for almost 90% range

Table 3. Results of testing mode for Disordered speech and normal speech samples

Parameter used	% samples confirmed for Vocal Folds Disordered speech	% samples confirmed for normal speech
Time periods Vs frequency TX graph	98 %	100 %
Percentile fo track variation	100 %	100 %
All 3 parameters % CQ, Tx, fo	97 %	100%

Table 4. Results of processing speech samples using BPF

Diagnostic marker	Range of values for Vocal folds speakers	Range of values for VFs after applying correction using BPF
Time Vs % CQ	15 % to 75 %	20 % to 60 % graph is more smooth
Time Vs Frequency TX graph	Regular Tx pulses occur from 110 Hz to 140 Hz and 260 Hz to 275 Hz	Regular Tx graph show pulses from 260 Hz to 275 Hz
Percentile Glottal Frequency fo variation	Percentile fo track indicates linear graph from 5 to 50 percentile range –0.077 low gradient indexes.	After BPF processing the f_0 graph that was nonlinear curved graph from 55 to 85 percentile range is smooth and approximately linear from 5 to 85 percentile range with 0.06 very low gradient.

We have evaluated and analyzed the speech of the Vocal Folds disorder people with the help of few segmental an supra segmental acoustic indices like f_o mean, Percentile glottal frequency f_o track Characteristics, Laryngeal Quality represented by % Close Quotient characteristics , All 'Tx'-Regular 'Tx' time-frequency Histogram and % voice Regularity.

2. Evaluation and confirmation of Vocal Folds disordered speech using the present theory is done for the first time by us and it is not done by any one before.

Concluding Remarks

The Vocal Folds disorder can be identified by evaluation of speech of 25 normal and 12 pathological subjects consisting of over 100 speech samples. Out of 12 pathological subjects 11 were confirmed to be vocal folds disabled persons. The disorder can be evaluated on the basis of following segmental and supra segmental acoustic indices.

• The speakers exhibit variability in different acoustical indices from time to time.

• The range of % Voice Regularity is observed to be low and it is 15% to 40%.

• The range of % CQ Mean is observed to be 25% to 49 % with standard deviation of 13% to 20 %.

• The % of regular Tx pulses is very low with respect to all Tx pulses. The glottal pulses exist relatively only for 20 to 50 % of time duration of speech data and are effectively present in particular frequency range in 140 Hz to 300 Hz. The glottal pulses cannot be produced by the disabled speakers in low frequency range and in 400 Hz to 600 Hz spectrum.

• Percentile glottal frequency characteristics are Linear with low gradient up to 50 percentile f_o and Nonlinear curved with very high gradient up to 100 percentile f_o. In some cases Nonlinear with very high gradient.

• The minimum intensity levels attained are in the range 28 dB to 40 dB because of breathing voice segments present consistently.

• A disordered speech after BPF processing shows that f0 graph which was nonlinear curved graph from 55 to 85 percentile range becomes smooth and becomes linear from 5 to 85 percentile

Acknowledgments

The authors would like to thank the participants of research work- all the elderly people with various disabilities, speech disorders and without speech disorders for their extended cooperation and support for the completion of the work. The authors would also like to thank the family members and caretakers of all the participants for their patience and support while acquiring the speech data. The authors would also like to thank Dr. Mrs. Shilpa Hujurbajar Director of "Sanvad Speech and Hearing Therapy Centre, Kolhapur" for authenticating our work.

References

[1] Anandthirtha B. Gudi, H. K. Shreedhar and H. C. Nagaraj "Estimation Of Severity Of Speech Disability Through Speech Envelope" Signal & Image Processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011

[2] Zoran C′ irovic′, Milan Milosavljevic "Multimodal Speaker Verification Based on Electroglottograph Signal and Glottal Activity Detection" Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume 2010, Article ID 930376, 8 pages doi:10.1155/2010/930376

[3] Constantine Kouropoulos and Gonzalo R. Acre "Linear Classifier with Reject Option for the Detection of Vocal Fold Paralysis and Vocal Fold Edema" Hindawi Publishing Corporation EURASIP Journal on Advances in Signal

Processing Volume 2009, Article ID 203790, 13 pages doi:10.1155/2009/203790
[4] Tomoki Toda, Keiichi Tokuda "Statistical Approach to Vocal Tract Transfer Function Estimation Based On Factor Analyzed Trajectory HMM" 1-4244-1484-9/08©2008 IEEE 3925 ICASSP 2008
[5] Kumara Sharma, Anantha Krishna, and Niranjan U. Cholayya "Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology" Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume 2007, Article ID 85286, 9 pages doi:10.1155/2007/85286
[6] Daryush Mehta "Aspiration Noise during Phonation: Synthesis, Analysis, and Pitch-Scale Modification" Massachusetts Institute of Technology February 2006
[7] Arpit Mathur ,Shankar M Reddy ,Rajesh M Hegde "Significance of Parametric Spectral Ratio Methods in Detection and Recognition of Whispered Speech" EURASIP Journal on Advances in Signal Processing doi:10.1186/1687-6180-2012-157
[8] Dolle Deann Pagel "Vocal Fold Perturbation Rates: Comparison between Multiple Sclerosis and Normal Subjects" 1992 M.S. In Speech and Hearing Sciences Texas Tech University
[9] Perry C. Hanavan, Au.D. "Articulation and the Vocal Tract" Audiologist
[10] Pol Mas Soro "A spectral estimator of vocal jitter" Universite Libre De Bruxelles Année académique 2010
[11] D´arcio G. Silva, Lu´ıs C. Oliveira, and M´ario Andrea "Jitter Estimation Algorithms for Detection of Pathological Voices" Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing Volume 2009, Article ID 567875, 9 pages doi:10.1155/2009/567875
[12] R. Čmejla1, J. Rusz1,, L. Vokřál "Laboratory Tasks From Voice Analysis in The Study of Biomedical Engineering Using MATLAB" 2010 Project work supported by grant FRV F3a 316 Audio Signal and Voice Analysis and grants GAČR102/08/H008 Analysis and modeling biomedical and speech signal, SGS10/180/OHK3/2T/13 Voice and Speech Disorders Evaluation
[13] Pascal van Lieshout "PRAAT Short Tutorial" October 2003 University of Toronto, Graduate Department of Speech-Language Pathology, Faculty of Medicine

Biography

[1]Mrs. Manasi Dixit is working as Associate Professor in KIT's College of Engineering, Kolhapur .Her teaching experience is 30 years. Her main fields of interest are Digital Signal Processing, Speech Processing, Image Processing and Microwave Engineering. 28 PG students have completed their research work and have been awarded M.E.(E&TC) degree under her guidance. She has published 21 papers in reputed international journals. She has worked in the capacity of the SENATE member Shivaji University, Kolhapur and BOS-Board of Studies Member for Electronics and Telecommunication Engineering, Shivaji University, Kolhapur.

[2] Prof. Dr. Shaila Dinkar Apte is currently working as a professor on PG side in Rajarshi Shahu College of Engineering, Pune and as reviewer for the International Journal of Speech Technology by Springer Publication, International Journal of Digital Signal Processing, Elsevier Publication. She is currently guiding 8 Ph.D. candidates. Five candidates have completed their Ph.D. under her guidance. About 54 candidates completed their M.E. dissertations under her guidance. Almost all dissertations are in the area of signal processing. She has a vast teaching experience of 37 years in electronics engineering, and enjoys great popularity amongst students. She has been teaching Digital Signal Processing and Advanced Digital Signal Processing since last 18 years. Her previous designations include being an Assistant Professor in Walchand College of Engineering, Sangli, for 27 years; a member of board of studies for Shivaji University and a principle investigator for a research project sponsored by ARDE, New Delhi. She has published 32 papers in reputed international journals, more than 40 papers in international conferences and about 15 papers in national conferences. She has a patent published to her credit related to generation of mother wavelet from speech signal. Her books titled "Digital Signal Processing" and "Speech and Audio Processing" are published by Wiley India. A second reprint of second edition of the first book is in the market. The book titled "Advanced Digital Signal Processing" is published in June 2013 by Wiley publishers.