A STATISTICAL APPROACH TO STUDY PITCH VARIATION FOR PREDICTION OF VOICE DISORDER

Dr Karabi Ganguly Bhattacharjee
Department of Biomedical Engineering
JIS College of Engineering, Kalyani, Nadia – 741235, India

Nilotpal Manna
Department of Electronics and Instrumentation Engineering
JIS College of Engineering, Kalyani, Nadia – 741235, India

Tina Fiona Sheikh
Department of Biomedical Engineering
JIS College of Engineering, Kalyani, Nadia – 741235, India

ABSTRACT: *The human voice is a magical tool to communicate, express emotions, to create wonderful music through singing. Everyone has a distinct voice, different from all others; almost like a fingerprint, one's voice is unique and can act as an identifier. Unfortunately we never give much attention to the voice problem that can limit our ability to communicate and to complete daily activities. Voice disorders have a higher rate of occurrence among those who are in professions that require speaking, like teachers, aerobatic instructors, lawyers, social workers etc. Symptoms of a voice disorder range from hoarseness or a chronic dry, scratchy throat, a pitch/tone that is not pleasing, limitations in the ability to speak clearly, or periods of voice loss. Voice disorder may be attributed to the abnormality in the structure and/or function of the larynx (vocal folds) as well as the abnormalities of its different components, making each voice different; namely, pitch, tone, and rate. In this paper a statistical approach is investigated on pitch quality to make early prediction of voice disorder for professional talkers.*

KEYWORDS: Phonation, Vocal folds, Kymographic parameters, Electroglottography, Autocorrelation function, Spectral density.

INTRODUCTION

Speech is one of the most fundamental phenomena in human cultures. It is the most developed means of communication in existence. Speech is an extremely complex acoustical carrier of vocal information or voice. The human voice production mechanism is extremely difficult, as humans are capable of varying extensively the functioning of their vocal organs. The foundation for an effective voice is based on the coordination of three factors: Breathing, Phonation and Resonance. Breathing air out of the lungs produces the power supply for the voice. This airflow from the lungs makes the vocal folds (or vocal chords) in the larynx (or voice box) vibrate to make the basic sound of the voice; this process is called phonation. Basic sound produced by the vocal folds is too weak to be heard, which is then modified into the sound and recognised as the human voice as it travels up from the larynx through the throat, mouth and nose; this transformation is known as resonance. Production of a natural, effective voice depends on how well these three fundamental components are balanced or coordinated.

Sometimes voice disorder is observed due to various reasons. Study reports claim that voice disorder is mostly observed to those who are professional talkers like social workers, lawyers, clergymen, telemarketers, aerobics instructors and teachers [1]. Voice disorders have a higher rate of occurrence among university teachers due to Demands of teaching, vocal effort and years of teaching, suggesting investigation toward laryngeal pathology [2]. Reasons for voice disorder are manifolds - viral or bacterial infection, allergic reactions, long standing laryngeal mucosal inflammation, viscous mucus, vocal fold thickening, vocal abuse, sudden onset, chronic condition, airborne irritants (e.g., smoke, toxic fumes), medications, associated with upper gastrointestinal disorders and reflux.

The primary symptoms of voice disorders may be categorized as:

- Hoarseness - reflects irregular vibration of the vocal folds; voice sounds “raspy” or “rough”.
- Vocal Fatigue - feeling tired after prolonged talking; continued talking takes great effort; often co-occurs with hoarseness.
- Breathy Voice - unable to say complete sentences without running out of air; difficulty being heard.
- Reduced Phonation Range - usually associated with singers who report difficulty producing notes that previously presented no problem; tiredness & soreness in throat.
- Aphonia - absence of voice; usually have to speak in whisper; great deal of effort required to speak.
- Pitch Breaks - periodic squeakiness; voice cracks; voice seems out of control; patient reports never knowing what sound will come out.
- Strain/Struggle Voice - difficult to talk; inability to get voicing started or to maintain voice; it is a strain to talk; patients report feeling tension.

LITERATURE UNDERPINNING

When the structure and/or function of the larynx (vocal folds) no longer meets the voicing requirements established for it by its owner, the quality, pitch, and loudness of voice differ from their originality, which is termed as voice disorder. Almost every disorder may present in more than one symptom and one cannot associate one single symptom with one specific voice disorder. As example, hoarseness, increased vocal effort or limitations in pitch and loudness may be a sign of any number of disorders. Hoarseness is one of the most common voice complaints. About 1 in 3 people have their voice affected by hoarseness at some point in their lives. Research has identified that hoarseness costs several billion dollars in lost productivity annually from work absenteeism. Hoarseness affects 31% of telemarketers, 44% of aerobics instructors and 58% of teachers. Some of the most common voice disorders are vocal fold edemia, vocal nodules, vocal polyps, contact ulcers, and laryngeal carcinoma. There exist various mechanical testing methods and constitutive models that are currently in use for the characterization of mechanical properties of the vocal fold tissue [3]. Research

ISSN 2053-5783(Print), ISSN 2053-5791(online)

study has been made on larynx dimensions in respect of the presence of sexual dimorphism [4].

The pitch of speech, or speaking fundamental frequency, is often used as an indicator of voice development and, indirectly, the hormone activity. It has been observed that the kymographic parameters showed a negative correlation with fundamental voice frequency, whereas the open quotient displayed a positive correlation with the fundamental frequency. There were significant intergroup differences in the fundamental frequency, amplitude and lateral peak [5]. In this paper an investigation has been made on variation of pitch frequency due to voice disorder and to establish an algorithm for early prediction of it to take preventative measure.

Everyone has a distinct voice, different from all others; almost like a fingerprint, one's voice is unique and can act as an identifier. The human voice is composed of a multitude of different components, making each voice different; namely, pitch, tone, and rate. The human voice has many components and is created through a myriad of muscle movements [6]. Pitch is an integral part of the human voice. The pitch of the voice is defined as the rate of vibration of the vocal folds. The vibrations, and the speed at which they vibrate, are dependent on the length and thickness of the vocal cords, as well as the tightening and relaxation of the muscles surrounding them. The length and thickness of the vocal cords, however, are not the only factors that affect one's pitch. The pitch of someone's voice can also be affected by emotions, moods and inflection. Interestingly, our emotions can also affect the pitch of our voices. When people become frightened or excited, the muscles around the voice box (or larynx) unconsciously contract, putting strain on the vocal cords, making the pitch higher. A change in pitch is known as inflection and humans exercise this naturally all the time. The pitch of our voices is created through vibrations of the vocal folds. The rate at which these folds vibrate changes the way our voices sound, with faster rates equating higher pitches. The voice and the way it is used are unique to every individual. Voice is a periodical signal that presents perturbations in the glottal cycles along the frequency (jitter) and intensity (shimmer) [7]. These two parameters have been described as robust measures of the biomechanical properties of the vocal folds in situations of vocal assessment, which compares the results of therapeutic interventions [8]. For this study, the algorithms developed to calculate the jitter and shimmer values were based on the proposal of Davis [9, 10]. Normality values were validated in 0.18% (standard deviation of 0.1%) for jitter and 1.08% (standard deviation of 0.37%) for shimmer [11]. Based on these statistical data, a scheme is proposed to make prediction on voice disorder.

METHODOLOGY

Voice problems most commonly originate from the vocal folds or the laryngeal musculature that controls them. As the vocal folds are subjected to collision forces during each vibratory cycle and dried while the air being forced through the small gap between them, and the laryngeal musculature is intensely active during speech or singing and is bound to be tired by excessive use. However, it is physically difficult to perform dynamic analysis of the vocal folds and their movements. Direct invasive measurement of movements is prohibited due to location of the vocal folds. The vocal cords are surrounded by cartilage preventing the imaging methods such as x-rays or ultrasounds that distorts image quality. Positioning of fibre optic probe or camera in the throat for Stroboscopic or high-speed video measurement method is ruled out as these trigger a gag reflex closing larynx and stopping voice.

ISSN 2053-5783(Print), ISSN 2053-5791(online)

Indirect methods are based on inverse filtering of either microphone or oral airflow recordings, where the speech sound or the oral airflow waveform from a circumferentially vented (CV) mask is recorded outside the mouth and then mathematical analysis is applied to estimate of the waveform of the glottal airflow pulses, which in turn reflect the movements of the vocal folds. Electroglottography (EGG) is the other kind of noninvasive indirect analysis of vocal fold motion is the, in which electrodes are placed on either side of the subject's throat at the level of the vocal folds to record the changes in the conductivity of the throat according to how large a portion of the vocal folds are touching each other. Though this process is not sufficient to describe the vocal folds movement, but can provide useful information.

In this paper a noninvasive method for voice analysis is investigated. The proposed work is described as schematic at figure 1. Voice is captured by microphone converting to electrical signal. It is pre-processed at the preamplifier stage for signal conditioning followed by filters to remove the unwanted noise and artifacts. It has been observed that fundamental frequency or pitch is restricted to within the frequency range of 80 Hz to 300 Hz. Female speakers posses higher pitch compared to male speakers. Therefore the filters upper cutoff frequency is selected at 900 Hz. Microcontroller is used at next stage for sampling, data acquisition, pitch extraction, statistical analysis and data storage. Finally this is interfaced to a display device and desk computer to observe the analysed data.

Figure 1: Schematic Diagram of propose work

PITCH EXTRACTION SCHEME

Pitch is the fundamental frequency of vocal cord vibration. The average pitch of the speaker varies considerably from individual to another. There are several methods of of estimating the pitch as mentioned below.

- Autocorrelation function
- Zero-crossing interval sequence
- Average magnitude difference function
- Cepstral method
- Simplified inverse filter tracking algorithm

Here autocorrelation function is adopted for pitch estimation. The autocorrelation function of a discrete time signal may be expressed as:

$$\varnothing(k) = \sum_{n=-\infty}^{\infty} x(n)x(n+k)$$

For a random signal the appropriate expression is:

$$\varnothing(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-\infty}^{\infty} x(n)x(n+k)$$

The autocorrelation function has the properties that if the signal is periodic of period P, then its autocorrelation is also periodic of P, i.e., $\varnothing(k) = \varnothing(k+P)$ and $\varnothing(0)$ attains maximum value.

ISSN 2053-5783(Print), ISSN 2053-5791(online)

This means that the maximum value of autocorrelation is certain interval of P. The inverse of P is fundamental frequency of signal.

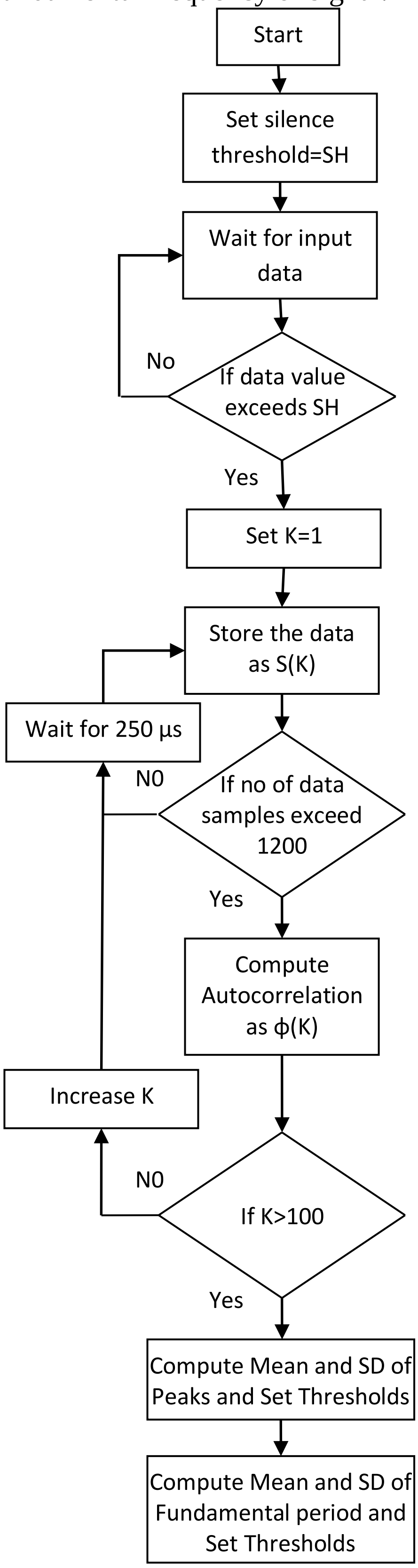


Figure 2: Computation of Autocorrelation

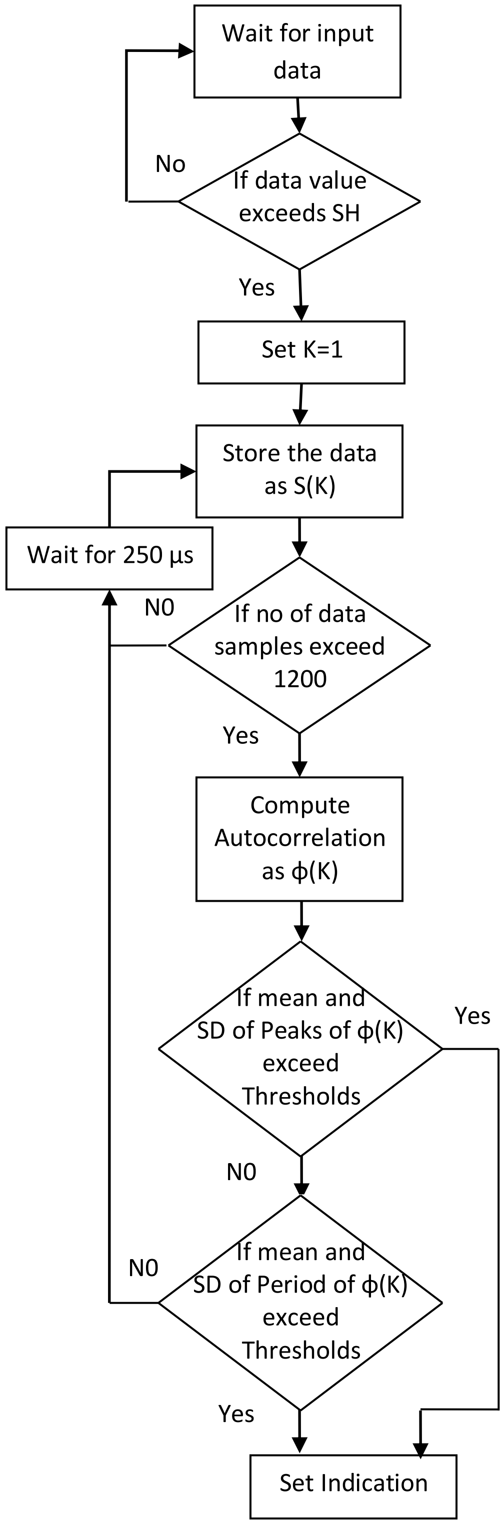


Figure 3: Prediction of Voice Disorder

ISSN 2053-5783(Print), ISSN 2053-5791(online)

The computation of autocorrelation function is not practical as the function sequence is spread over minus infinity to plus infinity. Therefore short time representation of autocorrelation is considered. The equation for computation is modified as:

$$\emptyset(k) = \sum_{n=-m}^{m} x(n)x(n+k)$$

Short time autocorrelation can retain all the information of speech signal and estimate accurately the vocal tract transfer function. Here, for this investigation, segments lengths of 30ms are considered at 10ms interval with the overlapping period of 20ms.

Pitch characterisation is of two folds types, - analysis of pitch intensity (amplitude of peaks) and analysis of jitter (variation of frequency). This scheme of pitch extraction is a dynamic process, where initial few seconds are analysis and calculation of pitch frequency establishing the limits of threshold limits of speech intensity and variation of pitch frequency. At the later stage is monitoring and indication if these parameters are crossing the limits of thresholds. The speech is recorded through a microphone, the received signal is preconditioned and then passed through an active lowpass filter. The processed signal is fed to microcontroller. The microcontroller is programmed to receive the signal at every 0.25 ms ensuring the sample rate of 4.0 KHz which is sufficient to process for pitch. Sets of 1200 samples, equivalent to 30 ms duration are stored and used for short time autocorrelation calculation. The amplitude of peaks of autocorrelation and its repetition times are averaged and noted. .

This process is repeated for 100 times considering each set of 30 ms of signal. Now mean and standard deviation of amplitude of autocorrelation peaks and repetition times are calculated. This follows the establishing of threshold limits for both the parameters. Following sets of signal is now compared with these threshold limits. If those exceed the limit, alarm indication is activated. The algorithm flow charts are demonstrated at figure 2 and figure 3.

IMPLEMENTATION

Simple circuits of operational amplifiers with few resistors and capacitors are used in pre-amplifier section as well as in filter section. Output of the filter is interfaced with microcontroller type ATMEGA 8 that has the advantage of inbuilt analog to digital converter. Therefore use of external analog to digital converter is avoided, yielding circuit simplicity and reliability. Microcontroller also performs all the computation for autocorrelation function, power spectral density, threshold settlement and indication.

RESULT AND DATA ANALYSIS

Autocorrelation function describes complete features of voice signal even if is contaminated with noise and artifacts. Fundamental frequency can be extracted from this. Some of the laboratory analyses for autocorrelation as well as power spectral density are shown at figure 4 to 10. It may be observed little variation in frequency can be clearly identifiable by autocorrelation analysis. Above investigation of autocorrelation analysis has been applied for vocal fold feature extraction with 10 different people of different age group and gender. The result is promising as small deviation of fundamental frequency is detectable. The statistical analysis on variation of frequency establishes the maximum limit that may cause the voice disorder.

ISSN 2053-5783(Print), ISSN 2053-5791(online)

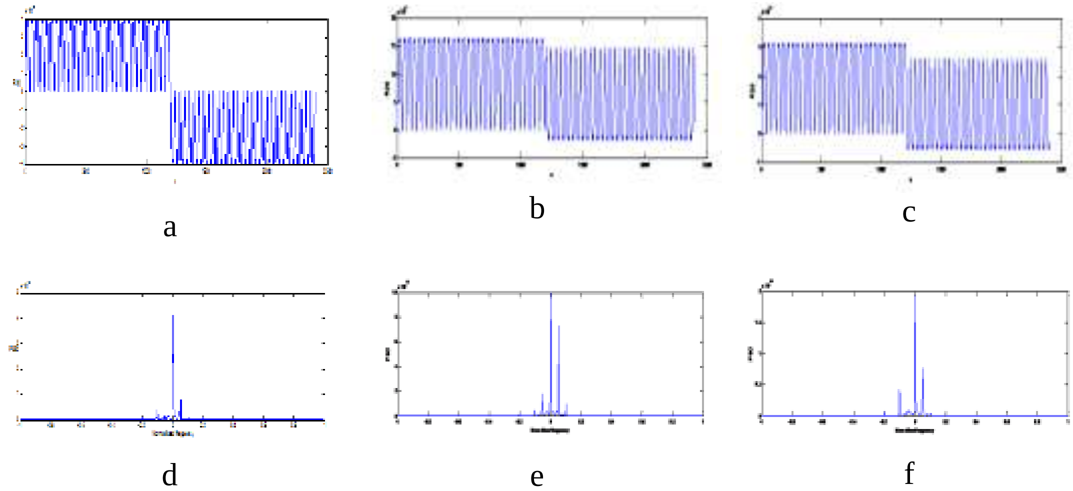


Figure 4: a, b, c are the Autocorrelation Functions at pitch frequency of 100 Hz with 1 Hz variation; d, e, f are Power Spectral Density

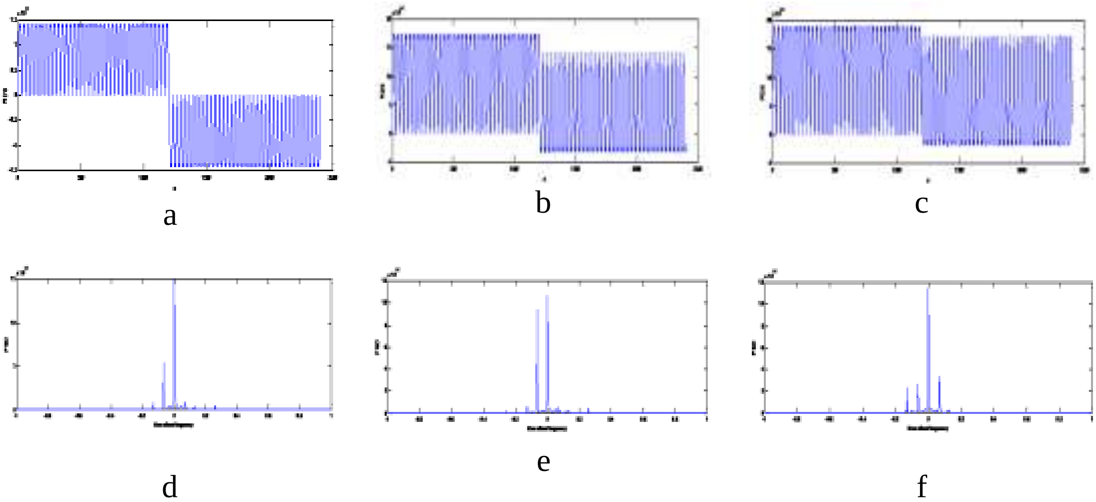
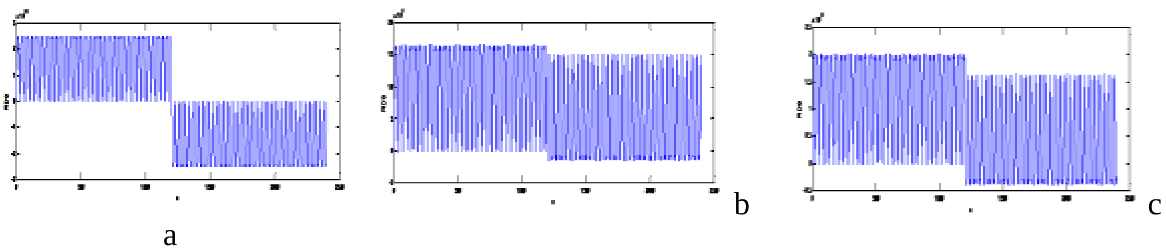
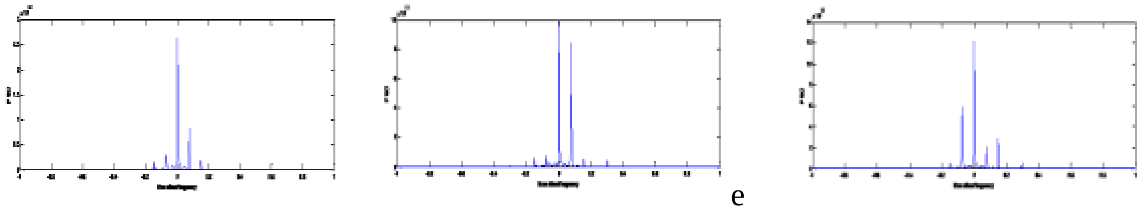


Figure 5 :a ,b ,c are the Autocorrelation Functions at pitch frequency of 130 Hz with 1 Hz variation; d, e, f are Power Spectral Density

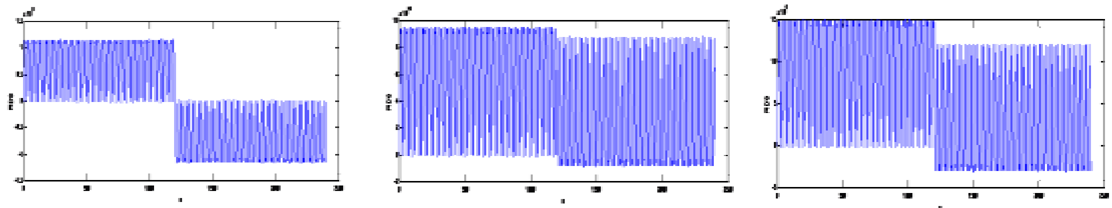


ISSN 2053-5783(Print), ISSN 2053-5791(online)

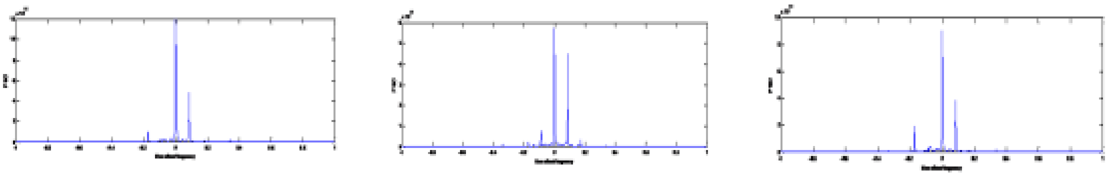


d e f

Figure 6 :a ,b ,c are the Autocorrelation Functions at pitch frequency of 150 Hz with 1 Hz variation; d, e, f are Power Spectral Density

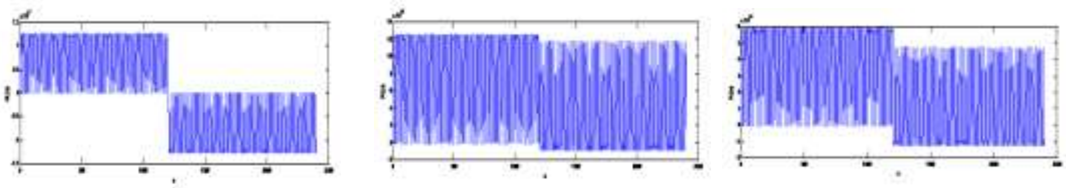


a b c

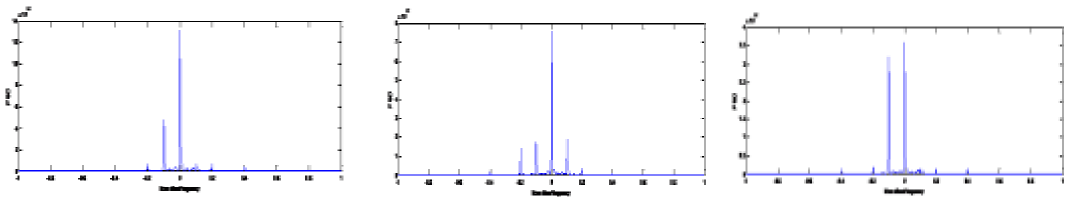


d e f

Figure 7: a, b, c are the Autocorrelation Functions at pitch frequency of 170 Hz with 1 Hz variation; d, e, f are Power Spectral Density



a b c



d e f

Figure 8: a, b, c are the Autocorrelation Functions at pitch frequency of 200 Hz with 1 Hz variation; d, e, f are Power Spectral Density

ISSN 2053-5783(Print), ISSN 2053-5791(online)

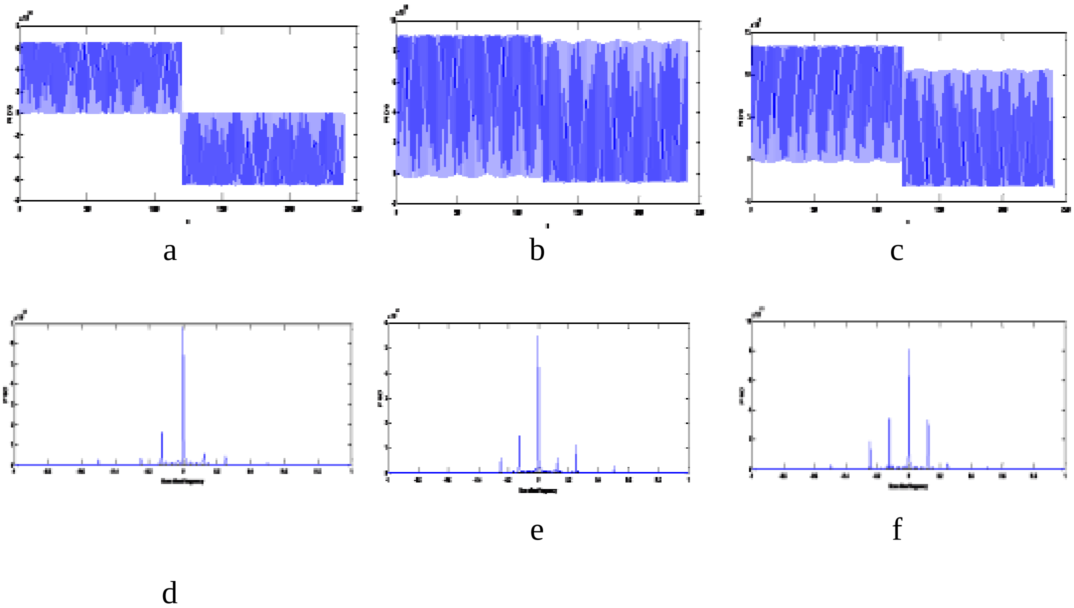


Figure 9: a, b, c are the Autocorrelation Functions at pitch frequency of 250 Hz with 1 Hz variation; d, e, f are Power Spectral Density

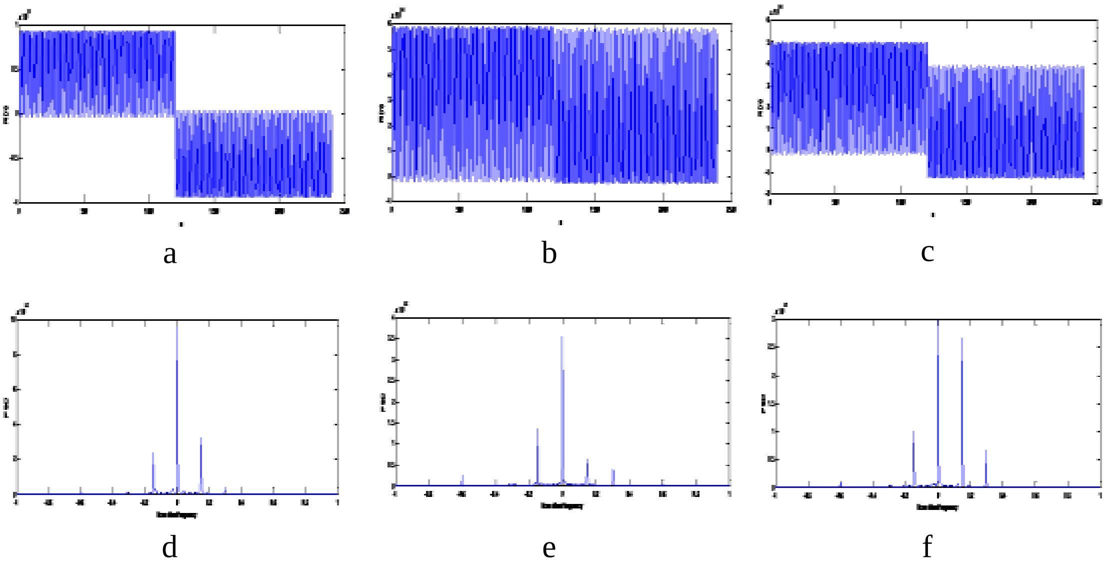


Figure 10: a, b, c are the Autocorrelation Functions at pitch frequency of 300 Hz with 1 Hz variation; d, e, f are Power Spectral Density

DISCUSSION

A simple method of determining the voice disorder or its prediction has been demonstrated in this investigation. One important point of consideration is to establish the threshold. The speech volume varies person to person as well as time to time and hence the amplitude of autocorrelation. Therefore the threshold limit is not unique and presetable for analysis of

ISSN 2053-5783(Print), ISSN 2053-5791(online)

speech quality of all the persons. Here the threshold selection is dynamic type and knowledge based. It has been created from initial few samples every time.

IMPLICATION TO RESEARCH AND PRACTICE

Vocal fold disorder cannot be diagnosed by invasive method as it is difficult to speak with the placement of sensor instrument in vocal fold. There exist various indirect methods of detection of vocal fold disorder. The method demonstrated in this investigation one of the simple and low cost types. The methodology may be applied for investigation of physiological parameters related to vocal fold which is helpful for treatment of larynx related diseases.

CONCLUSION

The method to detect voice disorder as described here gave satisfactory result in the laboratory. However the results are not consulted with pathological laboratories and certain limitations exist that may be overcome at further investigations. Firstly single parameter of pitch frequency may not be sufficient to predict the voice disorder. Other parameters related to vocal characteristics such as Short Term Power Spectra, Spectral Correlation, Ceptral Analysis may be considered. Secondly sampling frequency is selected as 4 KHz, which may be increased yielding better accuracy. But that requires more working memory; memory expansion may overcome this limitation. Lastly more number of people may be involved as subject to establish the algorithm.

The system demonstrated here is very simple and of low cost. A formal gadget may be developed for continuous monitoring of speech quality of professional talkers.

FUTURE RESEARCH

Though a good amount of investigations has been reported voice disorder, still there is enormous scope of research in this area. As mentioned, there are several types and causes of voice disorder; hence research scope exists for extraction of speech parameters in simpler way to provide useful information to a person. A professional talker may be benefitted from these information.

REFERENCES

1. BÖR Fritzell (1996). Voice disorders and occupation. *Logopedics Phoniatrics Vocology*. doi:10.3109/14015439609099197. 21(1):7-12.
2. Lilia Brinca, Patrícia Nogueira, Ana Inês Tavares, Ana Paula Batista, Ilidio C Gonçalves, Maria L Moreno (2015). The prevalence of laryngeal pathologies in an academic population. *J Voice.* 29(1):130.e1-9.
3. Amir K. Miri (2014). Mechanical Characterization of Vocal Fold Tissue: A Review Study. *J Voice.* DOI: http://dx.doi.org/10.1016/j.jvoice.2014.03.001. 28(6):657-667.

4. Geraldo Pereira Jotz, Marco Antônio Stefani, Omero Pereira da Costa Filho, Tais Malysz, Paula Rigon Soster, Henrique Zaquia Leão (2014). A Morphometric Study of the Larynx. *J Voice*. DOI: 8. 28(6):668-672.
5. Koishi HU, Tsuji DH, Imamura R, Sennes LU (2003). Vocal intensity variation: a study of vocal folds vibration in humans with videokymography. *Rev Bras Otorrinolaringol*. 69(4):464-70
6. Titze IR (2006). Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. *J Speech Lang Hear Res*. 49(2):448-59.
7. Elaine Lara Mendes Tavares, Roberto Badra de Labio, Regina Helena Garcia Martins (2010). Normative study of vocal acoustic parameters from children from 4 to 12 years of age without vocal symptoms. *Brazilian Journal of Otorhinolaryngology*. Print version ISSN 1808-8694. http://dx.doi.org/ 10.1590/ S1808-86942010000400013. 76(4):485-490.
8. Brockmann-Bauser M, Drinnan MJ (2011). Routine acoustic voice analysis: time to think again? *Curr Opin Otolaryngol Head Neck Surg*. DOI:10.1097/MOO.0b013e32834575fe. 19(3):165-70.
9. Davis SB (1979). Acoustic characteristics of normal and pathological voices. *ASHA*.11:97-115.
10. Cooley JW, Tukey JW (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*. 19(90):297-301.
11. Regina Aparecida Pimenta, María Eugenia Dájer, Adriana Hachiya, Domingos Hiroshi Tsuji, Arlindo Neto Montagnoli (2013). Parameters Acoustic and High-speed kymography identified effects of voiced vibration and vocal fry exercises. *CoDAS* 25(6):577-83.

ISSN 2053-5783(Print), ISSN 2053-5791(online)