Development of an Effective Method for Evaluation of Vocal Cord Disorders with Speech Analysis

Mr. Satyajit A. Pangaonkar
Assistant Professor, Department of Electronics Engineering
MIT Academy of Engineering, Alandi (D), Pune
sapangaonkar@entc.maepune.ac.in

Dr. Sanjay M. Asutkar
Associate Professor, Department of Electronics and comm. Engg,
Manoharbahi Patel Institute of Engineering & Technology, Gondia
asutkarsanjay@yahoo.com

***Abstract*—**The assessment of vocal quality and vocal efficiency is an issue for discussion. Not much reliable and appropriate methodology for objective and perceptual voice diagnosis has defined. So there is a need for comparing these diagnosis and validate for the improvement. Also the clinicians are relied on oral short diagnosis. This is an initiative for the new upcoming and more user friendly methodologies for the researchers and clinicians. The European Laryngologist Society (ELS)has defined a protocol for commonly accepted and valid voice diagnosis approach. In objective and perceptual auditive evaluation, sometimes there is a need of more than two judgments, evaluation and compensation for the differences in the opinion. In the current literature, a large number of objective methodologies for the diagnosis has been introduced. It is however not clear which combination of one is particularly suitable for the categorization as well as characterization of disordered voice. This paper is a proposed approach which is more reliable and can distinguish the healthy subjects and the pathological disordered group through a voice extraction method using a low power electromagnetic and electromechanical embedded sensor as well as unique feature extraction methodology i. e. a new approach of classification which will yield an accurate and reliable result.

Keywords*-Hoarse,Harsh,puberfonia, cluttering, digital stroboscopy, endoscopy,hyperkeratosis,MFCC,AdaBoost,Purtebation,Nodule, Polyp,Crysts,RASTA-PLP, GAW, TERC Sensor, GMM, HMM,MANOVA.*

*****

I. INTRODUCTION

Vocal fold which is also very commonly known as vocal cards are comprise of twin enfolding of mucous membrane stretched horizontally across the layer of throat structure. Voice is specifically that part of human sound productions in which vocal cards are primary sound source. Vocal fold can be loosen, tighten; its thickness can be changed. Any one action results in change in pitch volume, timbre or tone of the sound production [8]. The presence of vocal fold pathology can cause significant changes for affecting speech production system, resulting in deterioration of voice quality. Initial symptoms of malfunction at the vocal fold are usually breathiness or hoarseness. Research studies suggested early diagnosis and subjective evaluation of the patient voice performed by the physician to detect laryngeal pathology.

Thus the detection and analysis of vocal fold disorder and its management is an important issue for the clinician, voice language therapists and scientists. The focus of this paper is not only to detect or measure all possible pathologies but also to assess quality for case where the probability of pathology is high. This paper explains the methodology for detection and diagnosis of the vocal fold pathology using as low power

ultrasonic sensor for the feature extraction and creating a multimodal function set used to classify more accurately through a simple classifier 'AdaBoost' for observing and classifying healthy and pathological voices.

II. ORIGIN OF THE RESEARCH PROBLEM

Pathological voice assessment has got much attention now days due to modern way of life of professionals like teachers, singers, radio jokies etc. Smoking, alcoholic drinks are some unhealthy regular habits are also causing voice disorders. Unawareness and ignorance sometimes about the severity of the disorders by the people may cause further complications. The symptoms which causes a changed voice quality are unreliable voice ,delayed voice initiation , low gravelly voice ,low pitch , voice break in first passage of sentences , airy or breathy voice ,inability to sing in high soft voice increased efforts to speak or sing, hoarse and rough voice quality, frequent throat cleaning , frequent cough ,extra force needed for voice, feeling of chocking or suffocation, strong emotions and stress , upper respiratory injection as well as vocal fatigue.

The medical science has suggested the categorization of voice disorder [15] in following manner

a) *Organic Disorder*
b) Functional Disorder
c) Psychogenic Disorder

Organic disorder further classified as structural which relates to the physical problem and neurogenic which relates due to problem in nervous system.

In functional voice disorder physical structure is normal but vocal mechanism is being used improperly and/or inefficiently. The psychogenic disorders are mostly due to psychological issues that results into poor voice quality. It has been also observed that up to 40% of the cases of dysphonia referred to multidisciplinary voice clinic [01]. Some disorders are not hypo functional and hyper functional [02] in which

neurological abnormality has contributed to the voice disorders. It has been observed that, when voice quality becomes worse in absence of (but not due to) anatomic and neurological, the disorders shall be functional.

Also as per the literature of survey conducted across the world (table-01) that shows the several disorders cause with their percentage of severity. This suggests the cruelty of the disease and a specific attention to be paid by the clinician to study and rectify it. The voice therapists are likely to have diagnosis of above disorder types but also believed to be unaware to knowledge of it and non attentive behavior towards the detailed diagnosis.

Disorders Cause	Cases (%)
Chronic Laryngitis	4.64
Acute Laryngitis	4.38
Muscle Tension Dysphonia (MTD)	90 (Females)
Vocal Nodules	30
Cysts	14
Cluttering	17
Vocal Bowing	10 (36, 22)
Contact Ulcer	3
Nodules	21 (females)
Vocal Cord Palsy	11.16
Polyps	3.59
Vocal Cord Bowing	3.98
Hoarseness	85
Hemorrhage	13
Spasmodic Disorders	0.01
Contact Granuloma	6.7
Stuttering / Stammering	32
Voice Breaks	65

Table-01 [1, 2, 3, 4, 5]

The above table shows patients have a dysfunctional voice disorders, like conversion dysphonia, puberphonia. In these cases they are capable of producing a normal voice but do not able to use vocal mechanism in a functionally appropriate manner. Some of the voice disorders (e.g. cluttering) often get undiagnosed [07].It causes slowing down the speech rate till the time, person realizes about his cluttering disorders [08]. The disorder due to improper growth of hormones is acromegaly. Approximately 17% of population showing no symptoms [14] when the growth of hormones is a cause of disorders. It has been seen that the growth of hormones in healthy patients may be five times higher than unhealthy patients.

The problem with the voice disorders sometimes diagnosed in subjective techniques may cause errors for clinicians in hearing patients voice which may cause errors. The objective category is based on physical measurement obtained during phonation which includes vocal fold vibratory movement such as digital stroboscopy, laryngoscopy, glotography, electromyography and video endoscopy. These invasive techniques are accurate but require costly equipment and experienced professionals [10]. Sometimes it (especially stroboscopy) causes discomfort to patients which leads to wrong or false diagnosis [9]. HSDE (High Speed Digital Endoscopy) gives poor image quality and high aquisioncost [10].

III. OBJECTIVE

The study also shows that the acoustic analysis sometimes proves an excellent tool for the detection of voice disorders, but false diagnosis may cause the limitations since no formal standards are available. In perceptual analysis of the diagnosis, a Multivariate analysis Manova [11], Residual Spectral Analysis- Perceptual Linear Prediction RASTA- PLP[10] gives the discrimination between normal & pathological voice using HMM (Hidden Markov Model) classifier [9]. The perception & prediction gives the degree of evaluation. Also the temporal derivative by this method doesn't yield accurate results. The GMM (Gaussian Mixture Model) and ANN (Artificial Neural Network) uses methodology of training the samples. More the sample better the result is the phenomenon. Thus in the domain of pathological voice assessment, several methods does not consisting the best possible results. Also the level variation is the big hurdle.

Thus a modern approach is needed to employ more accurate reliable differentiation between healthy and pathological groups. This modern and robust method will yield good result with minimum discomfort to the patients and exact estimation with the value analysis in the form of different parameters estimation such as Voice breaks, Voice Turbulence Index, Tremor frequency, Jitter, Shimmer, Harmonic to Noise Ratio (HNR), fundamental frequency(pitch)analysis likewise.

IV. PROPOSED APPROACH

Identification of voice disorders has been a vital role in our daily life now days. The proposed approach of detection of vocal fold pathology is divided in to two parts as

*a) **Sensorial Voice Detection &Recogntion***

*b) **Transform analysis of the speech extracted.***

Part- A) Sensorial Voice Detection &Recogntion

To have speech estimation, speech signals are extracted from the detection system. This detection system may be a simple microphone but the background noise is the problem to have a best possible approximation of the clean speech to be proposed for the further analysis, under the adverse conditions. As said earlier, acoustic microphone has a limited efficiency in high noise conditions. So primary goal was to develop a sensor which is a non acoustic in characteristics which uses the vocal fold vibrations for the excitation (energy harvesting) as well as uses the change of a dielectric property

 ISSN: 2321-8169
182 – 186

of the neck occurring due to opening and closing of the vocal fold to measure the vibrations for the glottal structure. Following interrelated goals are proposed for this methodology.

a) ***To design and construct a robust low power high noise immune speech acquiring process.***
b) ***Experimentation and testing the data recorded by the developed sensor with human test subjects.***
c) ***Organize, format and distribute the collected data during above experimentation phase of development.***

The earlier approach uses examining the vocal function in scopic as visual form. This is some disadvantageous since the fiber cable inserted in the throat of the patient and examination may be resulted.

The development Ultrasonic-based sensor is designed as a Handy Low power Electromagnetic Electromechanically Embedded Sensor (HLEEES). To apply the Doppler effect to speech recognition, we direct a transducer/receiver pair toward a user's mouth. The lower facial region can be modelled as a mesh of infinitely small reflecting surfaces. Each surface reflects a 40 kHz signal toward the receiver, and moves independently of the other surfaces. The result at the receiver is a superposition of sinusoids at different frequencies. In ultrasonic speech applications, the ultrasonic signal entering the vocal tract from the transducer has to travel through the air bounded by VT walls. As the exclusive effects of the medium on ultrasound, attenuation and dispersion are frequency-dependant, we need to have a numerical overview of the significance of these effects on ultrasound propagation in the air. These sensors are very sensitive and consume little power which is based on an ultra-high-frequency acoustic tone (e.g. 40 kHz) is directed at a moving object, causing reflections which are recorded by a receiver which is usually co-located with the emitter. The frequency of the reflected tone will be governed by the Doppler Effect, and can be expressed as

$$f = f0\ (1+ V.C)$$

where f0 is the frequency of the emitted tone, 'f' is the frequency of the reflected tone, V is the velocity of the reflecting surface towards the emitter, and C is the velocity of sound. Thus, if the ultrasonic tone reflects off of a surface moving towards the emitter, the received signal will have a higher frequency. Likewise, a lower frequency tone will be recorded when the reflective surface is moving away from the emitter.

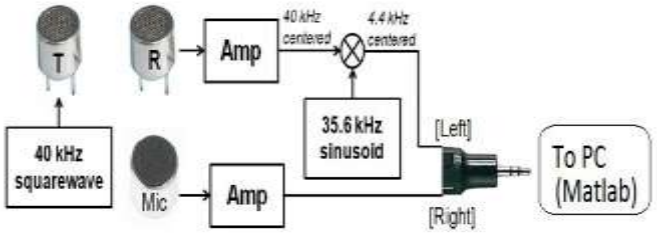

Fig-01*Ultrasonic Sensor circuit*

Part- B) Transform analysis of the speech extrated.

A set of partial differential equation is required to characterize the human speech production system. The Time varying nature of vocal tract shape is an important step for characterization and complete description. This can be done using a detailed mathematical analysis and modeling based on acoustic theory.

Although extensive research has been performed in this area, a universal theory has not been yet emerged. For modeling speech, we need source excitation (which is done in first part of the proposed approach) and the effect of speech radiation i.e. modeling over a finite time which has been done using transfer function analyzed by the DSP processor. As discussed earlier, the acoustic analysis are mainly based on periodicity of vocal fold vibration and air flow from glottis during speed which is nothing but measurement of Jitter(variation between successive fundamental period) so spectral statistical measurement is important.

The proposed approach schematic is as follows.

Here a multimodal feature data set can be created which is extracted and analyzed based on pitch, energy, MFCC (Mel Frequency Cestrum Coefficient), fundamental frequency etc. This feature can be further processed and classified the signal based on AdaBoost, a new classifier for training and recognition of pathological voice samples.

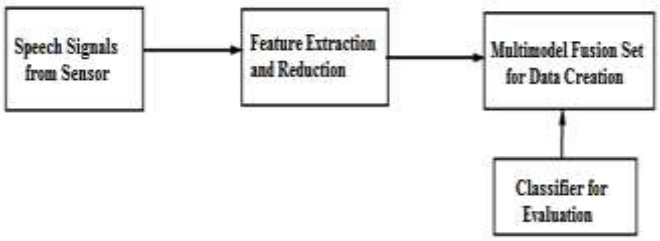

Fig-02*Multimodal Fusion Set with Classifier*

Here first preprocessing as pre emphasis to improve the efficiency will be done. As the speech is a non stationary signal, framing (with length 20-30 ms) can be done. The discontinuity of framing is minimized using each frame multiplied by window function. Further step to analyze the

fundamental (f0) or pitch voice corresponds to number of times per second can be estimated. Pitch duration algorithm is used to estimate this.

As said above, after creating multi model fusion set, the dimensionality reduction will be done to give the signal in training phase. Here supervised training will be adopted. After training the system will be ready to use. The final stage is classifier to classify pathological voice samples from normal ones. The AdaBoost algorithm for the classification also known as a weak learner is the first practical boosting algorithm which will give good results [13, 14].The normal and pathological voice can be characterized and classified in reference to variety of parameters in tabular form and the exact estimation is possible. A MATLAB can be used as a tool to significantly follow the above steps. The transform can be effectively plotted. This involves the Short Time Fourier Transform, Wavelet Transform (CWT, DWT) can be defined and a spectrogram of every pathology can be defined and a parametric estimation (approx.) as bellow.

Parameters (with Units)	Abbreviation	Value (Approx)
Degree of Voice Breaks (%)	**DVB**	0-Normal 10-Severe
Number of Voice Breaks	**NVB**	0-Normal 10-Severe
Degree of Voiceless (%)	**DUV**	0-Normal 10-Severe
Tremor Frequency (Hz)	**Fftr**	0-Bellow Threshold (Normal)
Highest Fundamental Frequency (Hz)	**Fhi**	Bellow 625 Hz-Normal
Lowest Fundamental Frequency (Hz)	**Flo**	Bellow 200 Hz-Normal
Jitter (%)	**Jitt**	1500- 4500 Hz- Inharmonic
Shimmer (%)	**Shim**	Random (based on Hoarse & Breathy Voice)
Noise to Harmonic Ratio (NHR)	**NHR**	Depends upon Jitter & Shimmer
Voice Turbulence Index (no.)	**VTI**	Random (Based on incomplete or loose adduction of the vocal folds)

Table-02

V. DATABASE AND EXPERIMENTATION

The database has been created by the speech language pathologists (using proposed sensor) as well as by recording speech samples from the age group of (40±02 years) men and women with various professions such as school and college teachers, singers, vocalists, radio and music jokies)with pronunciation of

a) ***Sustained phonation of vowels /a/,/e/,/i/,/o/,/u/ etc***
b) ***Specific words***
c) ***Specific texts (sentences)***

The analysis will be made with almost 100 subjects(in all) and a tabular as well as spectrogram analysis can be made. Finally a perception of the severity can be made with degree using GUI (Graphical User Interface) window with a scale of (0-2 Normal, 3-5 abnormal, 6-8 Severe, 8-10 Most Severe)

VI. DISCUSSION

Medical science suggests that the vocal fold disorders have to be diagnosed in the early stage otherwise there is no bypass for the surgical treatment. In most of the cases, a doctor recommends an acoustic analysis. This is a series of tests that measures the quality of your voice, including it's pitch stability, range and intensity. Often these tests are used when vocal cords are paralyzed or if a growth must be removed surgically. Using the test results, doctors & voice therapists can judge the amount of improvement after treatment.

In the same aspect, this paper gives us awareness towards the disfunctionality behavior of the speech organs, a vocal fold. The disorder causes the change in acoustic voice signal. In the literature, there are varieties of algorithm and approach to diagnosis them but having much limitations of accurate estimation of disorders.

VII. CONCLUSION

The main goal of this paper is to propose the unique and effective methodology of voice signal extraction, estimation using a low power electromagnetic electromechanical sensor and further a time domain analysis in the form of spectrogram. This analysis will be proven very accurate and effective method compared with the earlier approach. Thus the clinician can thoroughly and effectively able to categorize healthy and pathological conditions and degree of severity and can guide the patients how to avoid it.

VIII. REFERENCES

[1] Gauton Schlatthaler, Maria Eugenia Tonesm, Maria Jackson Menaldi,"Pattern Recognition approach to Spasmodic Dysphonia and muscle Tension Dysphonia Automatic classification", Journal of Voice, 2008, vol.-23, no.-3, pp. 346-53.

[2] Makhadoom N, Abouloyoun A., "Bakhory H A, Dhafar KO, Azab B A," Prevelance of Gastroesophyringal reflux Disease in Patients with laryngal & Voice Disorders", Saudi Medical Journal , July 207, Vol. 28, issue-7, pp 1068-71

[3] Estes C, Sulica L,"Vocal Fold Pseudocyst: result of 46 cases undergoing a Uniform Treatment Algorithm", Laryngoscope, May 2014.

[4] Danial J. Stein , B A J. Pieter Noordzij MD,"Incidence of All Causes of Chronic Laryngities", Boostan University School of Medicine and Boston University Medical Center

[5] Robert Sataloff,"Laryngopharyngeal Reflux and Voice Disorders-An Overview on Disease Mechanism, Teatment and Research Advances"

[6] A Schindler, F. Mozzanica, D. Ginocchio, P. Maruzzi, M. Atac, F. Ottaviani, "Vocal Improvement after Voice Theorpy in the Treatmentbenign Vocal Fold lesions", Acta Otorhinolaryngologica Italica, Oct. 2012, Vol-32(5), pp. 304-308

[7] Akhito Yamauchi, Hiroshi Imagawa, Hisayuki Yokonishi, Takaharu Nito, Tasuya Yamasoba, Takao Goto, Shingo Takano, Ken'Ichi Sakaibara and Niro Tayama, "Evaluation of Vocal Fold Vibration with an Assesment form for High Speed Digital Imaging: Comparative Healthy Young and Elderly Subjects", Journal of Voice, Vol- 26, Issue-6, pp-742-750.

[8] Jagdish Nayak, P. Subhanna Bhat, "Identification of Voice Disorders Using Speech Samples" , IEEE TENCON-2003 Conference on Convergent Technologies for Asia Pacific Region, Vol.3, 15-17 Oct. 2003, pp. 951-953

[9] Yuling Yan, Kartini Ahmad, Melda Kunduk, and Diane Bless, "Analysis ofVocal-fold Vibrations from High-Speed Laryngeal Images Using a HilbertTransform-Based Methodology", Journal of Voice, Vol. 19, No. 2, April 27, 2004,pp. 161–175

[10] Ali S. M. Saudi, Aliaa A. A. Youssif, Atef Z. Ghalwash, "ComputerAided Recognition of Vocal Folds Disorders by Means of RASTA-PLP", Computer &Information Science, Canadian Center of Science and Education, Vol. 5, No. 2; March 2012 doi:10.5539/cis.v5n2p39, pp. 39-48

[11] Aaron French, Marcelo Macedo John Paulson, Tylar Waterson and Angella Yu, "Multivariate Analysis of Variance (MANOVA)", ResearchArticle published in Dec. 1998

[12] Meisam Khalil Arjmandi, Mohammd Pooyan, Hojal Mohamma Dnejad, Mansour Vali, "Voice Disorder Identification based on Different Feature Reduction Methodologies and support Vector Machine", 18th Iranian Conference on Electrical Engineering (IEEE) 11-13 May 2010, pp. 45-49.

[13] AdaBoost Lecture by Jiri Matas and JanSochman , Centre for Machine PerceptionCzech Technical University, Prague, http://cmp.felk.cvut.czon 11 Feb.2015, 10.00 am.

[14] www.lionsclinic.orgsited on 21 March 2015, 06.00 pm

ACKNOWLEDGEMENT

This work is supported by the Research Grant ID- 13ENG000562 for the duration 2013-2015 under the Board of College and Univeristy Development (BCUD), Savitribai Phule Pune Univeristy, Pune, India

ABOUT AUTHORS

Prof. Satyajit A. Pangaonkar is currently working as an Assistant Professor &a course Inchrge of Master of Engineering in department of Electronics Engineering at MIT Academy of Engineering, Alandi (D), Pune. He has 15 years of teaching & 02 years of industry experience. He has several national & international publication on his credit. His area of interest is Wave and Antennnas,Enginerring Electromagnetics, Digital circuits, Speech Signal Processing etc.

Dr. Sanjay M. Asutkar is working as an Associate Professor in department of Electronics & Communication Engineering Engineering at MIET Gondia. He has 27 Years teaching experience. He has several international & national publications to his credit. He has participated in several international & national conference. His field of interest includes Wireless Sensor Network, FPGA, CPLD,Faultdiagnosis in Digital circuits , Timing issues etc.