CHAPTER 1-INTRODUCTION

Speech is the basic mode of communication and a unique feature of human beings. Human speech results from the complex interaction between phonation and articulation. Phonation or primary voice signal is generated in the larynx by the vibratory structures like vocal folds, ventricular folds and epilarynx. These primary sound signals are filtered in the respiratory cavities by vocal tract articulators like the tongue, soft palette, lips pharynx and nasal cavity. These articulations are produced by variation of the geometry of the filter cavities and thus the recognizable words are produced. If the vocal tract contains no constrictions then vowels are produced. In general, speech can be considered as a series of voiced and unvoiced sounds in a meaningful temporal sequence [1].

According to their mode of excitation, speech sound can be classified into unvoiced sounds or fricative and plosives [2]. Speech is basically the acoustic wave produced when air is expelled from the subglottal system comprising of lungs, bronchia and trachea. When air expelled from the lungs propagates through the nasal tract and the vocal tract, which are tubes of non-uniform crossectional area, these tubes modify the frequency spectrum by their resonance effect. The geometry and dimension of these tubes decides their frequency selectivity. The resonance frequencies of the vocal tract tube are called the formant frequencies. The uniqueness of speech of any individual is decided by its spectral characteristics. Complex interplay of frequencies produces characteristic features of sound. This emphasizes the importance of acoustic analysis of human voice or vocal disorders. Thus sound spectrograph was the principle tool for speech analysis for many years. In addition several acoustic perturbation methods were also in use for measurement diagnosis and voice treatment. The conventional acoustic analysis tools include fundamental frequency (F0), Harmonic to Noise Ratio (HNR), Pitch Amplitude (PA), jitter, shimmer, Zero Crossing Rate (ZCR). However, all these traditional analysis techniques were based on the assumption that speech production can be modeled as a linear process according to the standard assumptions of linear acoustics and 1D plane wave propagation of sound in the vocal tract [3]. Eventhough this model has been applied to speech coding [2, 4]; recognition and synthesis, there are several experimental and theoretical evidences for nonlinear aerodynamic phenomenon during speech production [5 - 8]. Therefore, recently the application of nonlinear dynamical analysis has received much scientific interest and has successfully characterized laryngeal pathologies [9 - 15].

The sensitiveness of traditional perturbation methods to signal length sampling, noise and its appropriateness to aperiodic voice signal [15 - 18] has added to the significant shift of interest to non-linear methods. Nonlinear analysis methods are generally based on the theory of dynamical systems. Based on this theory, the dynamics of any system can be represented in the state space constructed from its dynamical variables. According to nonlinear time series methods, the dynamics of any system with many degrees of freedom can be investigated using time series of a single scalar observable. From a dynamical system perspective a speech signal can be regarded as an observable of speech production system, which can be used to uncover its dynamics.

Considering the voice signal as a time series data necessary information about the underlying system dynamics can be extracted by reconstruction of phase space behaviour. Recently such method has gained much attention by their successful voice characterization [19, 20]. Although such phase space reconstruction methods are found to be successful to an extent [12, 14,15, 18, 21] they are sensitive to non-stationarity and noise contamination of the signals. Conventional nonlinear method of correlation dimension is found to be affected by non-stationarity, noise and finite signal length in most of the biological data including electroglotagraphic signal [21]. Under such circumstances new method of entropy based on the asssement of the predictability of the system are found to be a better choice [22 - 26].

The aim of this thesis is to compare nonlinear dynamical analysis and perturbation analysis methods for assessing their applicability in real time and online setup. One of the main problems still unsolved in the domain of speech fluency disorders is an objective and an automatic way of judgments of patient performance before and after speech therapy sessions and an assessment of gains made after intervention. For an effective evaluation and comparison of speech quality improvement clinicians and therapist should have quantitative information about the improvement that could be provided by specific methods.

The appropriate application of perturbation analysis and nonlinear dynamic analysis can provide different but complementary information and may potentially improve our ability to objectively assess voice disorders and evaluate the effects of treatment of laryngeal pathologies.

Most of the existing methods of signal analysis give significant results when the time series is simulated from low dimensional dynamical systems and fails or misleads in the presence of noise. Hence real world time series analysis of the data requires preprocessing for noise elimination. Furthermore embedding dimension and time delay are critical parameters in reconstruction of state space and computation is time consuming, which restricts its application on real time basis. Hence it is essential to have a very fast algorithm, which can process the data at the same rate at which it is acquired.
Our goal is primarily to detect dynamical changes using technique from real world data sets where there is no time for preprocessing and fine-tuning of data. Here we aim to make use of nonlinear parameters like maximum Lyapanov exponent (**λ_{max}**), Correlation Dimension (**D_2**), Kolmogorov entropy (**K_2**) and a new entropy measure Permutation Entropy (**PE**) for this purpose by establishing a link with the conventional state space approach.

We also aim to study the effectiveness of the invariant parameters and PE in extracting the change in dynamics caused by abnormalities in the vocal tract. This may be of great advantage in preliminary clinical diagnosis in identifying the vocal disorders before proceeding for further expensive treatment strategy.

Aim of the thesis

(1) To study the nonlinear characteristics of stuttering subjects vs. normal Subjects.
(2) To study the nonlinear properties of normal and abnormal speech Processes.
(3) To directly compare the efficiency of traditional perturbation and nonlinear time series analysis methods in characterizing speech signals of vocal disorder and stuttering subjects.
(4) To estimate the efficacy of speech therapy rendered to the stuttering subjects using a fast and robust entropy called Permutation Entropy (PE).
(5) To investigate the effectiveness of different methods in detecting change in dynamics in the above processes.

1.1 THESIS ORGANISATION

Chapter 1: Introduces the problem and defines the aim of the thesis.

It explains that speech production is not a linear phenomenon but it is a complex mechanism and acoustic studies and linear methods are not sufficient to completely study the dynamics. It is better to go for a more general method such as an entropy measure to differentiate between two kinds of signals, the healthy and pathological cases.

Chapter 2: Speech production and evidence of Nonlinear Behaviour.

It deals with the anatomy of speech production and explains the various frequencies that arises due to the oscillations of the vocal tract which appears to be quasi periodic in nature and due to the modulations of muscle parts in the nasal tract like tongue, lips , jaw, teeth etc different resonances of sounds are produced. The evidences of nonlinear dynamics in speech production are also explained.

Chapter 3: Current Techniques in Speech Analysis

It includes various linear and nonlinear methods used in the present work and in general. Various linear measures to characterize speech process are spectrogram, jitter, shimmer, signal to noise ratio, linear predictive coding (LPC), Pitch amplitude, zero crossing. Power spectral analysis, wavelet analysis and statistical methods used to describe the complex nature of the system. Nonlinear measures used for characterizing the dynamics of speech are Lyapunov exponent, Correlation Dimension, cross-correlation sum, pseudo entropy, pseudo correlation dimension, fractal dimension etc.

Chapter 4: Discusses the results of the analysis of stuttering.

Chapter 4 deals with the data recording and methodology of analysis adopted in stuttering. The present study evaluated speech of 10 stuttering subjects of age groups 16 - 40 and 10 age matched control subjects speaking Malayalam –their native language(a south Indian language). The intension of present research is to find the difference in speech parameters between stutterers and fluent speakers. Vocal signals are recorded for Malayalam vowels('അ' 'ഇ' 'ഉ' and a sentence from a children's story "ഇരുവരും വലിയ ചങ്ങാതിമാരാണ്" the meaning in English is "Both are good friends" using a microphone and a multimedia Computer in a sound proof room before a group of audience with a sampling rate of 11khz.

The results achieved were to develop an objective and automatic method for charactersing the stuttered and fluent speech. The present work focuses on the robust and fast processing of the data using Permutation entropy (PE) even in the presence of noise which were not possible by other nonlinear methods.

The results indicate that PE is an efficient measure for characterising the audio signal and establishing quantitatively the improvement in the progress of post treatment rendered by speech therapist and clinicians to Persons With Stammering **(PWS)** time to time. Specifically, the evolution of PE, average phonation number, average PE, maximum Lyapunov exponent (λ_{max}), Correlation Dimension (D_2), Kolmogorov entropy (K_2) calculated proves this. At every level of treatment the PE data can be stored in the database of the patient, which can be compared to the PE data of patients before treatment, and the database of PE of the fluent speaker. This can also give the first hand information in the diagnosis about the treatment and thus help the therapist and clinicians. The result confirms that the stuttering dynamics exhibits lowered complexity when compared to that of normal speech signals. No noise filtering or reconstruction of attractor is needed unlike other nonlinear methods, which requires lots of computation time. PE is robust to noise and allow fast evaluation.

Chapter 5: Discusses the results of the analysis of vocal disorders.

The Permutation Entropy technique allows visualizing the differential dynamics between healthy voices and voices with some vocal fold pathologies. The evolution of PE, Average PE, Correlation dimension, Lyapunov exponent and Kolmogorov entropy calculated differentiates the healthy signals from the pathological signals. The results show that using PE an overall increase is indicated in PE and this can be measured very fast. No noise filtering or reconstruction of attractor is needed unlike other nonlinear methods, which requires lots of computation time. PE is robust to noise and allow fast evaluation.

Two groups of samples were used, one from healthy individuals and the others from people with nodule in the vocal fold, Reinke's edema, paralysis, polyps, cancer and papilloma . Vowels /അ/, /ഇ/ and /ഉ/ from Malayalam language are used for analysis.

Permutation entropy is well suited due to its sensibility to uncertainties, since the pathologies are characterized by an increase in the signal complexity and unpredictability. The results showed that the pathological groups had higher entropy values in accordance with other vocal acoustic parameters presented. This suggests that these techniques may improve and complement the recent voice analysis methods available for clinicians. Effectiveness of PE to identify the vocal pathologies is verified on clinically characterized vocal sound data from 10 patients suffering from six different cases. Time series of sound signal for each vowel from 10 pathological cases are used for PE analysis. PE of order 5 is calculated for sliding window of 512 samples.

Chapter 6: Contains Summary, conclusions and scope for future work.

This deals with the overall summary, conclusion and scope for future work.

1.2 REFERENCES:

[1] M. Petursson and J. Neppert., Elementarbuch der Phonetik. Buske, Hamburg, 1991.

[2] L.R.Rabiner and R.W.Schafer, Digital Processing of Speech Signals, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[3] Vassilis Pitsikalis and Petros Maragos, Speech Analysis and feature extraction using chaotic models , IEEE, 2002.

[4] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[5] J. F. Kaiser, Some observations on vocal Tract Operation from a Fluid Flow Point of View, in Vocal Fold Physiology: Biomechanics, Acoustic, and Phonatory Control, I. R. Titze and R. C. Scherer(Eds.,), Denver Centre for Performing arts, Denver, CO, pp. 358-386, 1983.

[6] H. M. Teager and S. M. Teager, Evidence of Nonlinear Sound Production Mechanisms in the Vocal Tract in speech Production and Speech Modelling, W. J. Hardcastle and A. Marchal, Eds., NATO Advanced study Institute Series D, vol.55, Bonas, France, July 1989.

[7] T.J.Thomas, A finite element model of fluid flow in the vocal tract, Comput. Speech Lang. vol. 1, 131-151, 1986.

[8] D. J. Tritton, Physical Fluid Dynamics, 2nd edition, Oxford Univ. Press, New York, 1988.

[9] H. Herzel, D. Berry, I. R. Titze, and M. Saleh, Analysis of vocal disorders with methods from nonlinear dynamics, J. Speech Hear. Res. 37, 1008–1019, 1994.

[10] I. R. Titze, R. Baken, and H. Herzel, Evidence of chaos in vocal fold vibration, in Vocal Fold Physiology: New Frontier in Basic Science, edited by I. R. Titze, pp. 143–188, Singular, San Diego, 1993.

[11] A. Behrman and R. J. Baken, Correlation dimension of electroglottographic data from healthy and pathologic subjects, J. Acoust. Soc. Am.102, 2371–2379, 1997.

[12] Y. Zhang and J. J. Jiang, Nonlinear dynamic analysis of signal typing of pathological human voices, Electron. Lett. 39, 1021–1023, 2003.

[13] A. Giovanni, M. Ouaknine, and J.-M. Triglia, Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis, J. Voice 13, 341, 1998.

[14] J. J. Jiang and Y. Zhang, Nonlinear dynamic analysis of speech from pathologic subjects, Electron. Lett. 38, 294–295, 2002.

[15] Y. Zhang, C. McGilligan, L. Zhou, M. Vig, and J. J. Jiang, Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps, J. Acoust. Soc. Am. 115, 2270–2277, 2004.

[16] Kay Elemetrics Corp, Multi-Dimensional Voice Program: Software Instruction Manual, Pine Brook, NJ, 1993.

[17] I. R. Titze, Workshop on acoustic voice analysis; Summary statement, National Center for Voice and Speech, Denver, CO (1995), p. 36. 5M. P. Karnell, A. Chang, A. Smith, and H. Hoffman, Impact of signal type of validity of voice perturbation measures, NCVS Status Prog. Rep. 11, 91–94, 1997.

[18] J. J. Jiang, Y. Zhang, and C. N. Ford, Nonlinear dynamic of phonations in excised larynx experiments, J. Acoust. Soc. Am. 114, 2198–2205, 2003.

[19] I.R. Titze, R.J. Baken, H. Herzel, Vocal fold physiology – frontiers in basic science, Evidence of Chaos in Vocal Fold Vibration, Singular Publishing Group, San Diego, 1993.

[20] Marwan Al-Akaidi, Fractal Speech Processing, Cambridge University Press, Cambridge, 2004.

[21] Behrman and R. J. Baken, Correlation dimension of electroglottographic data from healthy and pathologic subjects, J. Acoust. Soc. Am.102, 2371–2379, 1997.

[22] Paulo Roge´rio Scalassara , Marı´a Eugenia Dajer, Carlos Dias Maciel, Rodrigo Capobianco Guido, Jose´ Carlos Pereira, Relative entropy measures applied to healthy and pathological voice characterization, Applied Mathematics and Computation 207, 95–108, ELSEVIER, 2009.

[23] R. Moddemeijer, On estimation of entropy and mutual information of continuous distributions, Signal Process. 16 (3), 233–248, 1989.

[24] R. Kleeman, Measuring dynamical prediction utility using relative entropy, J. Acoust. Soc. Am. 59 (13), 2057–2072, 2002.

[25] R.E. Hillman, E.B. Holmberg, J.S. Perkell, M. Walsh, C. Vaughan, Phonatory function associated with hyper functionally related vocal fold lesions, J. Voice 4 (1), 52–63, 1990.

[26] B. Hammarberg, Perception and acoustics of voice disorders: a combined approach, in: Proceedings of VOICEDATA98, Symposium on Databases in Voice Quality Research and Education, Utrecht Institute of Linguistics, Utrecht, The Netherlands, pp. 1–6, 1998.