ISSN 2347 - 3983
Volume 8. No. 2, February 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter20822020.pdf
https://doi.org/10.30534/ijeter/2020/20822020

Mathematical Model-Based Speech Signals Processing Effects on Human Stress

Dr.C.Balakrishnan[1], Dr.RM.Vidhyavathi[2*], Dr.K.Kavitha[3], Dr.J.Joseph Sahayarayan[4], Dr.S.Santhosh Kumar[5]
[1]Assistant Professor, Alagappa Institute of Skill Development, Alagappa University, Karaikudi, India, [1]balasjc@gmail.com
[2,4]Assistant Professor,Department of Bioinformatics, Alagappa University, Karaikudi, Tamil Nadu, India, [2]vidhyamiss@gmail.com, [4]jjsrbioinformatics2016@gmail.com
[3]Assistant Professor,Department of Computer Science, Mother Teresa Women's University, Kodaikanal, Tamil Nadu, India, [3]kavitha.urc@gmail.com
[5]Assistant Professor,Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India, [5]santhoshkumars@alagappauniversity.ac.in

ABSTRACT

The objective of this research paper is to process the typical voice signals which are utterances of ancient verses, chanting of mantras like Gayatri mantra, chanting of 'OM'. Vedic mantras recorded in normal and technically designed acoustic environment. The analysis and processing of these typical voice signals are proposed to carry out time-frequency analysis and to extract speech parameters like spectrum, energy spectral density, power spectral density, cepstrum, pitch, spectrogram etc. The proposed work in this research paper is aimed at establishing an appropriate alternative to medicinal therapies in typical metabolic and behavioral disorders; subsequently reducing the harm to life. This paper reveals the mathematical model for analysis and processing of typical speech signals. This will estimate the various parameters of speech which will be helpful to study human stress response and related stress parameters via. anger, fear, anxiety, bipolar disorders etc.

Key words: Cepstrum, LPC, Pitch, Formants, Vocal tract system.

1. INTRODUCTION

Speech is considered as a non-stationary signal because its frequency and spectral components change concerning time. The generalized example of a speech signal can be explained by a non-stationary multi tone sine wave with the help of following equations [1].

$$(t) = A1 \sin(\omega 1t + \varphi 1)$$
$$0 \leq t \leq t1 = A1 \sin(\omega 1t + \varphi 1) + A2 \sin(\omega 2t + \varphi 2)\ t1 \leq t \leq t2$$

However, the speech signal is completely different from a multi toned signal based on components in a given time interval. The interval considered in context to the above equation is of very short duration which is about 10 to 20 msec. In short, the frequency contents of speech changes continuously with time [2]. Since speech is an acoustic signal; mathematical analysis is carried out for designing a speech production system [2]. The speech production starts when the message is articulated by the speaker and is conveyed to the receiver via the voice signal. This message is converted into a set of phoneme sequences corresponding to the sounds that are made.

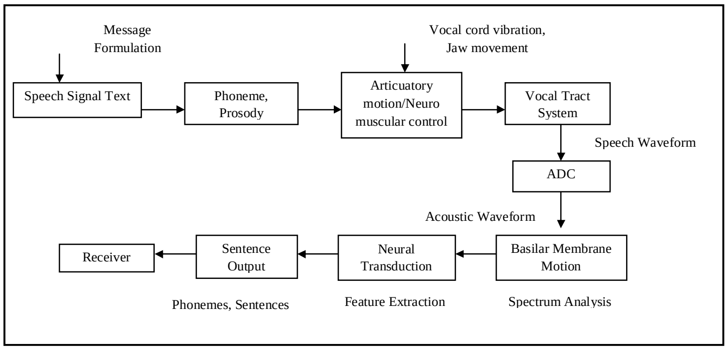

Figure 1: Speech Production Block Diagram

The speech originates mainly by three processes, revolving and twisting of nerves, whipping of membrane or air coming out through lungs [3]. The acoustic signal is produced by executing correct neuromuscular commands i.e. vocal cord vibration, jaw movement, control of lips, tongue and velum. This provides the proper shape to the vocal tract so that the sequence of speech can be obtained. As shown in figure 1, the neuromuscular commands are used for controlling various aspects of articulatory motions. These motions are of continuous inputs. Message formulation, phonemes at the beginning of the block are discrete. Thus vocal tract generates the speech waveform which is analog, hence ADC is provided so that correct acoustic waveform is given to the basilar membrane motion [3]. This block is used for spectrum analysis which analyzes what type of signal frequency is coming out through vocal tract. Neural transduction block is used for feature extraction like duration of sound, intensity, frequency and pitch; these all are continuous output. Hence the proper mathematical model is to be designed so that correct speech sounds can be analyzed for further applications. Then language translation is done to extract proper words and sentences and provided to the receiver. At analog/acoustic level high information rate is required to process the signal. Since the brain has trillions of neurons, information rate is less at message formulation and receiver block [4]. The vocal tract mathematical model is to be developed which describes how speech waveform is produced and analyzed.

1.1 Introduction about Stress

Stress is defined as physical or mental disability considering the emotional factor causing the mental tension. The brain and nerves are the major parts of the body which gets affected by the stress. The other part forms heart, pancreas, reproductive system etc. Stress as related to medical field can include psychological conditions like depression, anxiety, bipolar disorders. Hence, continuous exposure to stress can cause bad effect on human health which results in substantial loss to the society [5]. The stress reduction can be achieved through the chanting of mantras, chanting of OM, Vedic mantras recorded in the technically designed environment. To do this analysis, mathematical modelling of these typical speech/ voice signals is required.

2. BACKGROUND STUDY

2.1 Speech Production Mechanism

The speech signals are classified into voiced and unvoiced sounds. Voiced sounds are characterized by low-frequency components and vocal cords vibrate when these sounds are produced. Vocal cord opens and closes at a particular frequency and this is called a glottal pulse. In the case of unvoiced sounds, vocal cord is inactive and it is characterized by high frequency. Hence, it represents a noise generator. Considering these details, the vocal cords vibrate at a specific frequency which is called a fundamental frequency. The other type of sound is nasal sound in which vocal tract is combined with nasal cavity [6].

The spectrum of vocal tract system has the number of resonant frequencies. These frequencies are called as Formants. In general, every kHz has one resonant frequency. Formants are to be calculated to distinguish between two sounds, whether voiced or unvoiced

2.2 Discrete-Time Models for Speech Production

The digital filter model is to be developed for the human speech production mechanism. This model must precisely signify some characteristics; these are excitation model, vocal tract operation, radiation process of speech and voice and unvoiced speech for short duration i.e. about 10-20 msec as the vocal tract changes shape at every 10-20 msec.

The voiced sounds are represented by impulse train generator and glottis is a type of low pass filter. Discrete-time model for voiced speech production [7] provides the transfer function of the digital filter having a resonant frequency. So it has a transfer function of the form:

Transfer Function of Glottis = $1/1+b1\ z-1+b2\ z-2$

The excitation function will be different depending on the nature of speech. Impulse function is given by δ (n) and e(n) = δ (n-pk) where p is the pitch period, this is done for voiced speech; and for unvoiced speech e(n)=random(n); where e(n) represents input to the vocal tract system. Now taking Z transform; excitation signal is represented as

$(z) = 1/1-z-p$

The transfer function of discrete voiced speech production is given by S(z)/E(z)

$S(z)/E(z) = AvG(z)V(z)R(z)$

Thus all-pole models for vocal tract are developed and the overall transfer function of voiced and unvoiced speech is represented in the equation the following equations.

$(z)/(z) = Av/1+\Sigma\ akz-k\ Pk = 1$

$(z)/(z) = Auv/1+\Sigma P+2L+2\ akz-k\ k=1$

In most of the speech analysis, the vocal tract is modelled by a 12-pole filter. The speech production model is designed by modelling a vocal tract system for short frames of the speech signal of about 10-20 ms.

3. PROPOSED METHODOLOGY

3.1 Linear Predictive Coding

The Due to change in the vocal tract at every 10 ms, a short time interval is to be considered while analyzing the speech signal. Over a short time interval, the vocal tract system has following transfer function; where the equations represent voiced signal transfer function and nasal and unvoiced fricative sound. (p+1) represents the number of poles e(n) is the excitation signal [8]. Considering the frames of 10-20 ms which are found by proper windowing technique parameters of speech processing are calculated. These parameters are further useful in stress analysis in human.

a) Short Time Energy

In this case, every sample in a frame is taken and is squared, then added, which provides energy of that signal. This parameter is used to distinguish between voiced and unvoiced speech for a particular duration. Hamming window or rectangular window is used for forming the frames of proper duration. The energy is calculated by:

$(n) = \Sigma\ ((k).(n-k))2\ \infty k = -\infty$ (10)

where n is the sample rate

b) Pitch Period Estimation

In this vocal cord, vibration takes place, which distinguishes between male/female/child speakers. It is equal to the inverse of the fundamental frequency. The short time autocorrelation function is used to estimate the pitch. The equation to implement this is given by the following equation.

$\emptyset(k) = 1/N\ \Sigma N-1\ s(n).\ (n-k)\ k=0$

c) Linear Predictive Coding

LPC technique is used for estimating the speech parameters like pitch, formants, spectra, vocal tract area function. The basic principle of LPC is that the current speech sample can be predicted from the previous sample, i.e. it is a combination of the past predicted sample. If 12 pole model of speech is used then past 12 samples can be obtained. The pole model 'ak' is calculated using LPC analysis [9]. The speech signal s(n) related to excitation u(n) is calculated using the differential equation which is obtained by taking the inverse Z-transform.

$(n) = \Sigma\ ak\ (n-k) + G(n)\ p\ k=1$

LPC is stated as inverse filtering because it determines all zero filters which are inverse of the vocal tract model. G is the amplitude gain control, u(n) is excitation signal s(n) is speech signal output. The input to the model will be the previous samples of speech s(n-k) and estimated output is the current sample. The error signal will be generated as the difference between the actual sample and the estimated sample. Pitch period estimation is done through LPC analysis [10].

3.2 Cepstral Analysis

There are basically, two components of the speech signal, one is input excitation e(n) and other is vocal tract system. The excitation model and vocal tract model are to be analyzed and modelled individual. Hence these two components are to be separated from speech. Cepstral analysis differentiates the speech into excitation source and system components. According to a speech processing technique, voiced speech is characterized by impulse sequence and unvoiced sound as the random sequence. If e(n) is excitation source and h(n) is impulse response of vocal tract filter then in the frequency domain

$S(\omega) = E(\omega).H(\omega)$

Now the separation between voice speech and excitation is to be done. Therefore IDFT of log spectra of both the components is to be found. IDFT of the log spectrum transforms the speech signal into the frequency domain, related to time domain [11].

The input speech signal s(n) is given to hamming window for proper obtaining frames of the signal at 10-20 ms thus s(n) is converted into x(n). DFT of x(n) is obtained and log of $X(\omega)$ is taken to obtain the magnitude of the speech signal. And then IDFT of this signal is taken to obtain the cepstrum which is basically vocal tract components.

The short term speech has low-frequency components for vocal tract characteristics and high-frequency components for input excitation. Hence to extract these characteristics liftering operation is performed. This provides formant estimation and bandwidth for vocal tract spectrum. The formants are estimated by peak picking procedure from smooth vocal tract spectrum. The phase part of cepstrum is also computed, which is useful for reconstructing the original sequence. [7].

4. RESULTS AND DISCUSSION

More secretion of these hormones results in anger, anxiety, bipolar disorders etc. To control these hormones a measure is a low cost, an alternative approach to medicinal therapy is to be developed. Hence chanting of mantras, Vedic mantras, OM etc. is important to control the flow of these stress hormones. Therefore these typical speech signals are to be processed in a technically designed acoustic environment and correct parameters are to be extracted, in which these biomedical signal are working. The below figures show voiced segment analysis of the typical signal and its cepstrum analysis and slow varying vocal tract characteristics are obtained using low time liftering procedure and hamming window technique. This is useful for formant calculation.

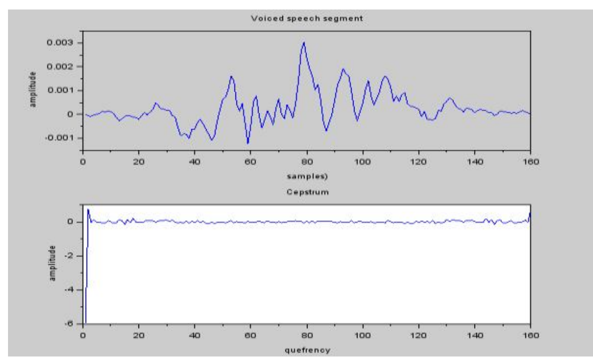

Figure 2: Voice Segment Analysis

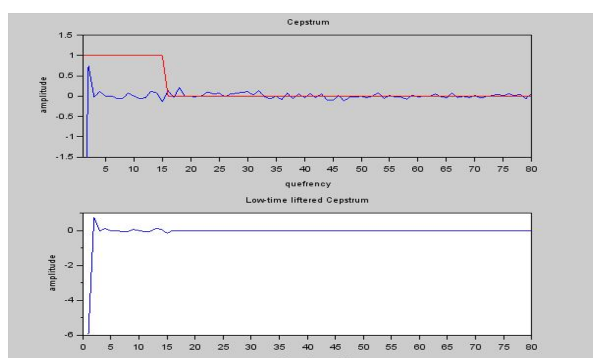

Figure 3: Low Time Liftering Cepstrum Estimation

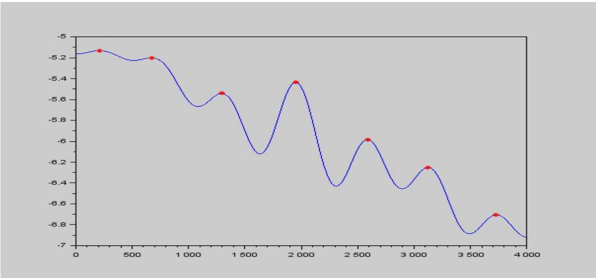

Figure 4: Formant Estimation

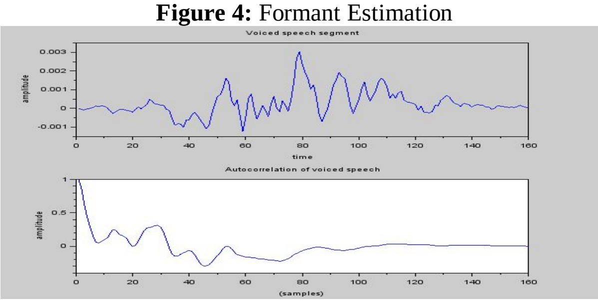

Figure 5: Autocorrelation of voiced speech

5. RESULTS AND DISCUSSION

The mathematical modeling approach is discussed in this paper to analyze and process typical speech signals. The time vs. frequency analysis is developed to show the various components of speech signal like short-time energy, ZCR etc. The vocal tract model, excitation process and transfer function of the discrete-time model of the speech signal is also analyzed. It is reviewed that cepstrum analysis are done to model the excitation and system components of speech and to estimate parameters like pitch in high time and low time liftering and formants. To measure the pitch period and to remove vocal tract components to obtain high pitch period LPC analysis is carried out in this research work.

ACKNOWLEDGEMENTS

This article has been published under RUSA Phase 2.0 grant sanctioned vide letter No. F. 24-51/2014-U, Policy (TN Multi-Gen), Dept. of Edn. Govt. of India, Date 09.10.2018.

REFERENCES

1. Myriam et al., "**The Effect of Music on the Human Stress Response**", *PLOS One Journal*, published DOI:10.1371/journal.pone.0071056, 2013, pp. 1-14. https://doi.org/10.1371/journal.pone.0071056
2. Carlo Magi, "**Mathematical methods for linear predictive spectral modeling of speech**", *Helsinki University of Technology*, Report-9.
3. Punnet Mongia et al., "**Estimation and Statistical Analysis of Human Voice Parametersto Investigate the Influence of Psychological Stress and to Determine the Vocal Tract Transfer Function of an Individual**", *Journal of Computer Networks and Communications*, vol. 2014, pp.-1-18. https://doi.org/10.1155/2014/290147
4. "**How to measure stress in Humans**" Report, Centre for studies on Human stress, 2007, *Fernand-Seguin Research Center*, Quebec, Canada.
5. "**Speech and Audio Processing**", *Lecture Handouts*, 2004.
6. Majed Ismail Hussein, "**Developing a Speech Production model in terms of Z-transform for speech recognition implementation in computer**". *IJRRAS*, vol 14 Issue1, 2013, pp. 136-140
7. Sarah Hawkins, "**Speech and Spectral Analysis**", *Foundation of Speech Communications*.
8. Stephen Levinson, "**Mathematical Models for Speech Technology**", *John Wiley and Sons Publications*.
9. Mark Hall, A.S. Willsky, "**Time varying parametric modeling of Speech**", *Elsevier Science Publications*, 1983, pp.-267-285. https://doi.org/10.1016/0165-1684(83)90074-9
10. S.Santhoshkumar, "**A Study of Data Mining Techniques for Auto Immune Diseases**", *International Journal of Pure and Applied Mathematics*, Volume 118 No. 9, 2018, pp. 1-4.
11. E.S. Gopi, "**Speech Production model", Digital Speech Processing using MATLAB, Signals and Communication Technology"**, DOI:10,1007/978-81-322-1677-3-2, *Springer*, India 2014.