International Journal of Engineering Applied Sciences and Technology, 2017
Vol. 2, Issue 6, ISSN No. 2455-2143, Pages 15-19
Published Online in IJEAST (http://www.ijeast.com)

WORD LEVEL LOW TIME LIFTERED CEPSTRUM ANALYSIS OF MALE AND FEMALE SPEAKER

Sakshi Bedi
Department of E.C.E.
S.S.C.E.T, Badhani, Punjab, India

Randhir Singh (Associate Prof.)
Department of E.C.E.
S.S.C.E.T, Badhani, Punjab, India

Abstract— **The capability of speaking the language is one of the most amazing skills human possess. It serves as a very effective way of communication, sharing experiences, feelings, thoughts and ideas among people. In this research work cepstral analysis of word level speech signals of male and female speakers are carried out. Spectrogram analysis of the segmented speech is carried to determine various speech parameters. Cepstral analysis shows a unique trait between male and female speakers in terms of low liftered speech processing.**

Keywords— **Speech signal, word analysis, Hamming, Cepstral, Spectrogram.**

I. INTRODUCTION

Speech is the vocalized form of human communication. It serves as a very effective way of communication sharing experiences, feelings, ideas and thoughts among people. However, such ability has also been taken for granted for long time. The discipline that aims to understand the human speech communication process is speech science, which usually include the study of physiology of speech production, the acoustical characteristics of speech and the process by which listeners perceive speech [1-2]. The function of motor control is to drive the Human brain which originates an idea of what to speak and correspondingly provides the control to signals through sensory nerves to the organs that contribute to speech production. The motor control unit receives the control signals and the respective speech production organs move and take proper shape accordingly to the type of words we speak and the sounds we produced. The whole mechanism referred as Articulatory Motion. The third type of function of the Human speech production is the Speech or sound generation which is made up of air that comes out of the mouth and nasal cavity and propel with force in the open space in the shape of acoustic wave. The acoustic wave reaches the human ear after being generated through mouth and nasal cavity and identified through sensory nerves with the help of attached ear and the Human brain. The speech production starts with the creation of idea in mind about what to speak or what sound to produce. With the help of sensory nerves the idea origination is passed onto the Human vocal apparatus. The whole procedure referred as motor control function which is mainly divided into two parts, the motor commands generation part and the language processing part. Our brain is partitioned into different segments. The function of these segments is to perform various control, think and memory functions. The Auditory region which is often called as Broca's area(region) for the language processing gets input in the form of visual gestures and listening through other sensory organs, helping it to determine what to speak or what sound to produce. The motor control region which is basically known as Wernicke's region produce control signals to progress the apparatus of vocal tract and other speech production organs such as lungs, vocal chords, glottis, tongue, jaw, teeth, lips etc [3-5].

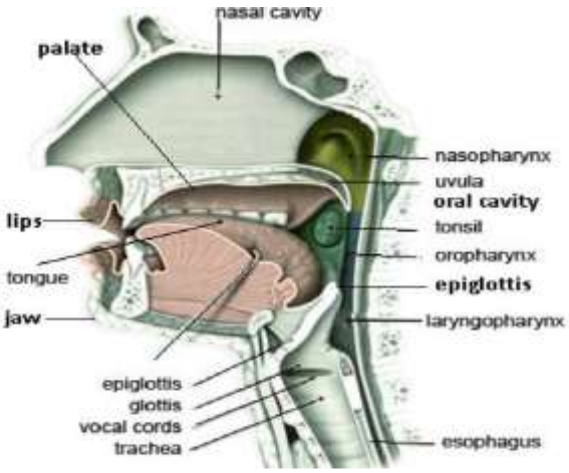

Fig. 1. Human Vocal Apparatus

In the production of speech and sound in the Humans different types of organs are involved. These types of organs are flexible in nature and with the command of motor control signals received from the brain their size alters depending upon the type of the sound and speech to be produced. The function of lungs is to supply necessary air force for the production of sound in the configuration of acoustic wave [6]. The air progress through the pipe linking lungs and the throat, vocal chords, glottis, epiglottis and other organs in the mouth and finally emerge through the nasal cavities and mouth in the type of acoustic wave. Different types of organs through which the air passes during the operation of origination of speech and sound. The air throw out from the lungs moves up through trachea and enters larynx when we speak. Inside the larynx the air is confined by a pair of lip like tissues generally called vocal chords. They decide pitch of the speech produced and are the foremost membranes of the vocal apparatus. For males the length of the vocal folds differs from 17 to 25 mm and it varies from 12.5 to 17.5 mm in the case of females. To produce voiced speech and supply temporarily diminution to produce unvoiced speech there are the vibrations inside the vocal folds [7-10]. However, phonemes are the miscellaneous types of speeches for which carry out the closing and opening of the vocal folds in different fashion with the help of air passage and then direct it to the upper part of the vocal tract shown in Fig. 1.

Different types of people speak different languages according to the area in which they are born. To speak in their mother tongue they do not need different type of training or knowledge. By understanding both Audio and visual gestures children learn to speak in their respective mother tongue at an early age of one year. With the help of symbols the signs of any language can be pronounced easily called as Phonemes. All the spoken world languages have 20 to 60 phonemes. The contextual outcome, sentiments and characteristics of the speaker need to be pronounced in case of phonemes which are not necessarily required in the written text of any language. The designing of these phonemes is according to the articulatory movement of the vocal tract. The phonetics of any language consists of two types of phonemes: The vowels and the consonants. Vowels are mostly voiced sounds which are produced when the vibration in the vocal chords carried out in periodic manner. On the other hand the unvoiced sounds are totally random in nature. During the production of voiced sounds the vocal chords vibrate frequently almost in the periodic manner when the air passes through it. Similarly for the case of unvoiced sounds the vocal chords are fully open, completely close or partially open. The most acknowledged format of phonemes for the American English language is ASCII symbols which is generally called as ARPAbet. According to the location of the tongue in the mouth cavity, vowel phonemes are further divided into three types Front, Mid and Back [11-12].

II. CEPSTRAL ANALYSIS

Cepstrum analysis is perturbed with the de-convolution of two types of signals: The one is of fundamental (basic) wavelet and the other one which consists of a train of impulses (excitation function). Therefore, the representation of the composite signal is in terms of power, complex and phase spectra [15].

The cepstrum analysis is very much familiar with data which consists of wavelets. However, prior to analysis, this is very true even the shapes of the respective wavelets are not known. It has many uses in military and navy applications in which the power cepstrum was successfully applied in the Radar analysis where the arrival time was determined by reducing the interference of the main wavelet, and in marine exploration where source depth was determined and ocean depth was mapped with the help of cepstrum analysis. The substantial prominence is given to cepstral analysis in this chapter in medicine in the diastolic heart sound analysis for the detection of coronary artery disease, ECG pattern classification and speech signal decomposition for theoretical as well as bandwidth compression application purposes. Bogert et al. developed the cepstrum perspective to find out the echo arrival times in composite signal by decomposing the non-additive constituents. The word cepstrum was coined by reversing the first syllable in the word “Spectrum”. The cepstrum exists in the domain referred to as “quefrency” which means that the reversal of the first syllable in frequency which has units of time [13-15].

Cepstral analysis is based on the observation that:

$$x[n]=x_1[n]*x_2[n]\Leftrightarrow X(z)=X_1(z)\cdot X_2(z)$$

By taking the Log of X(z)

$$\log\{X(z)\}=\log\{X_1(z)\}+\log\{X_2(z)\}=\hat{X}(z)$$

If the complex log is unique and the Z-Transform is valid then, by applying Z^{-1}

$$\hat{x}[n]=\hat{x}_1(n)+\hat{x}_2(n)$$

The two convolved signals are now additive.

The real Cepstrum analysis is defined as:

$$c_x[n]=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|X\left(e^{jw}\right)\right|e^{jwn}dw$$

Its magnitude is real and non-negative.

And the complex cepstrum:

$$\hat{x}[n]=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left[X\left(e^{jw}\right)\right]e^{jwn}dw$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left[\left|X\left(e^{jw}\right)\right|\right]+j\arg\left(X\left(e^{jw}\right)\right)e^{jwn}dw$$

Where, arg() represents the phase. We call it complex because it uses complex logarithm, not due to sequence, which can also be real.

III. EXPERIMENT AND RESULT

Speech material was collected from four speakers (2 male and 2 female, ages 23-27 years). The male speakers were represented as M1 and M2. On the other hand, female speakers are referred as F1 and F2. The speakers we use in our experiment were university students of different age groups and had English as their first language. In the first part of the speaker selection, speech recording and segmentation is done. In this experiment the speech of two male and female speakers was recorded in the form of words. In this experiment the segmentation of speech is carried out since each speaker takes different time while speaking words. This process is also carried out with the use of

MATLAB program which uses Hamming Window. For signals which are changing with time we must take a sample or window of the signal over some definite interval. If we simply cut out some portion of signal which introduces distortions in the signal, we reduce these distortions by multiplying our portion with smoothing function which reduces the size of signal at edges. We need a shape of that signal in such a way which has a spectrum with narrow central lobe and small side lobes. The window based on raised cosine shape is called Hamming Window.

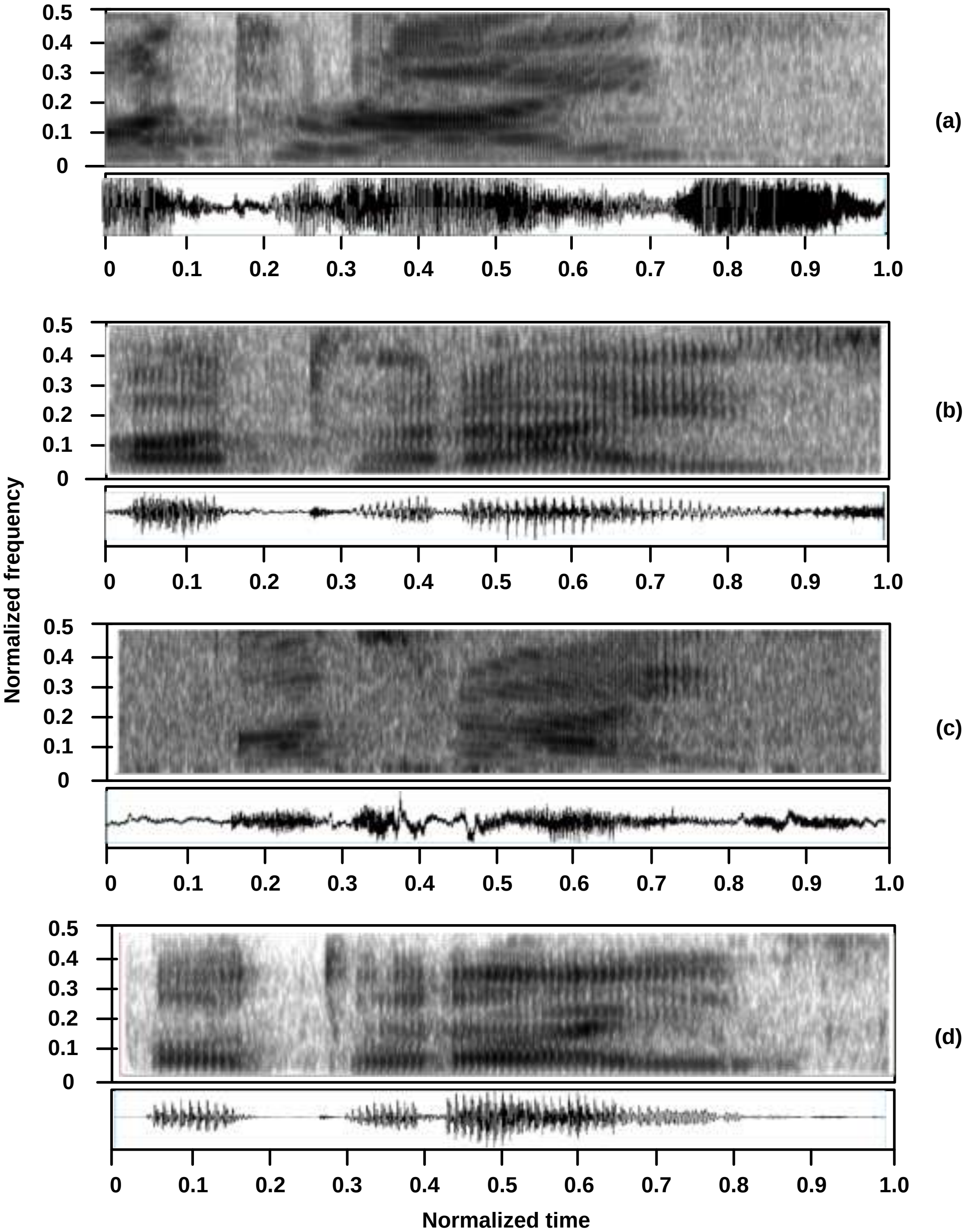

Fig. 2. Human Vocal Apparatus Normalize signal and spectrograms of word “Authorize” (a)-(b) male speaker and (c)-(d) female speaker.

In this experiment the speech of two male and female speakers was recorded in the form of words. This process is also carried out with the use of MATLAB program which uses Hamming Window. The outputs of various stages are carried out during the computation of cepstrum. Fig. 3 which is extracted from the low-time liftered cepstrum the maxima at 3.3 w.r.t quefrency at 600Hz and the minima at 2.6 of log magnitude spectrum w.r.t quefrency at 1300 Hz. The graph of Fig. 4 which is extracted from the low-time liftered cepstrum the maxima at 1.4 w.r.t quefrency at 1500 Hz and the minima at -0.4 w.r.t quefrency at 3000Hz. Fig. 5 which is extracted from the low-time liftered cepstrum the maxima at 1.8 w.r.t quefrency at 200 Hz and the minima at 0 w.r.t quefrency at 3500Hz. Fig. 6 which is extracted from the low-time liftered cepstrum the maxima at 1 w.r.t quefrency lies in the frequency range of 300 Hz and the minima at -0.5 w.r.t quefrency at 2500 Hz.

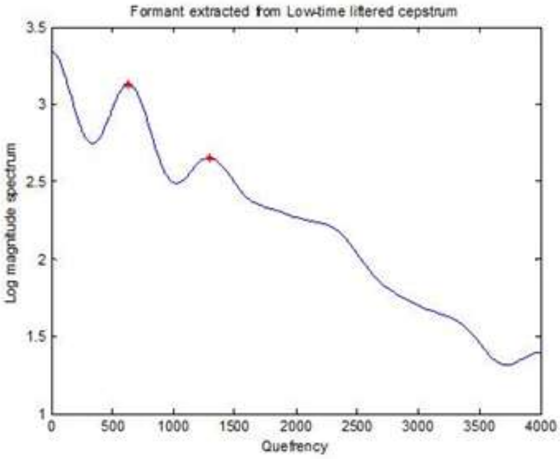

Fig. 3. Low time liftered cepstrum of Sp1 for word “Authorize”.

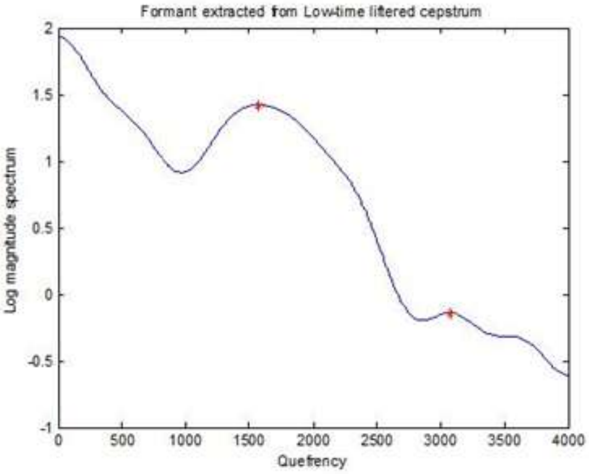

Fig. 4. Low time liftered cepstrum of Sp2 for word “Authorize”.

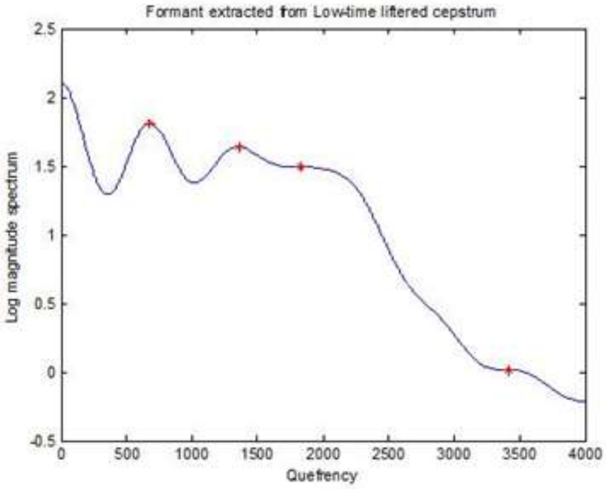

Fig. 5. Low time liftered cepstrum of Sp3 for word “Authorize”.

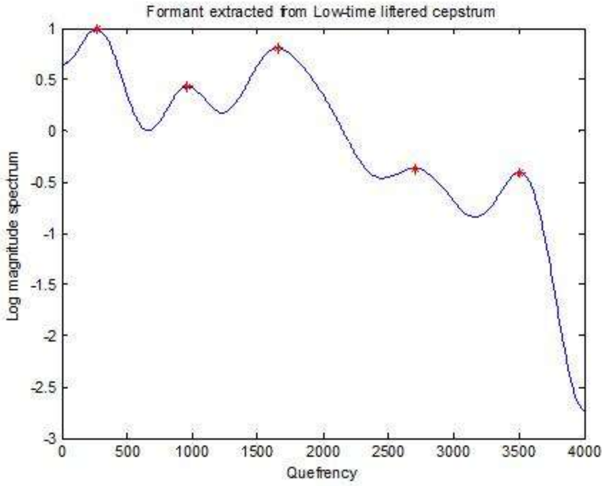

Fig. 6. Low time liftered cepstrum of Sp4 for word “Authorize”.

Various parameters such as pitch, intensity, and formant frequency of male and female are given Table 2 and Table 2.

Table 1 Value of Speech paramters of word “Authorize” for female speaker.

No of **Parameters**	**Female**	
	Sp1	**Sp2**
Pitch	303.84 Hz	247.05 Hz
Intensity	72.30 dB	69.5 dB
Formant Freq.	894.56 Hz	987.08 Hz

Table 2 Value of Speech paramters of word "Authorize" for male speaker.

No of Parameters	**Male**	
	Sp3	**Sp4**
Pitch	134.55 Hz	132.93 Hz
Intensity	73.02 dB	77.65 dB
Formant Freq.	651.35 Hz	812.03 Hz

IV.CONCLUSION

In this research work cepstral analysis of word for male and female speaker is carried out. From speech analysis it is observed male speaker has high intensity as compared to female speaker whereas pitch and formant frequency are high for female speaker. It is also observed from the results that male speaker have more maxima and minima in the cepstral curve as compared to female speaker in word level analysis.

V. REFERENCE

[1] B. Corona, M. Nakano, H. Pérez, "Adaptive Watermarking Algorithm for Binary Image Watermarks", *Lecture Notes in Computer Science, Springer, pp. 207-215, 2004.*

[2] K. Honda, "Physiological Processes of Speech Production," Springer Handbook of speech processing.

[3] C. Darwin, " The Expression of Emotions in Man and Animals," 1872.

[4] "Species–Specific Formation of Human Vocal Apparatus," Language in the Brain : Critical Assessments.

[5] M. Akay, "PREFACE" Biomedical Signal Processing. Pp. xiii-xiv, 1994.

[6] W.H. Goodenough, "Language Origin Philip Lieberman, Uniquely Human: The Evolution of Speech, Thoughts and Selfless behaviour." Cambridge, MA: Harvard University Press, 1991. Pp. 210.

[7] W.T. Fitch, "The Evolution of Speech: a comparative Review," Trends in Cognitive Sciences, vol. 4, No. 7, Pp. 258-267, July. 2007.

[8] A.C. Cohn, J. Clark and C. Yallop, "An Introduction to Phonetics and Phonology Language," vol. 68, No. 1, p. 156, March 1992.

[9] O' Saughnessy, "The Speech Communication- Human and Machine (Addison-Wesley,1987).

[10] Rossing. T. The Science of Sound (Addison-Wesley, 1990).

[11] R. Carlson, G. Fant and B. Granstrom, "Two- Formant Models, Pitch and Vowel Perception," Auditory Analysis and Perception of Speech. Pp. 55-82, 1975.

[12] H.M. Teager and S.M. Teager, "Evidence of Non-linear Sound Production Mechanisms in Vocal – Tract," Speech Production and Speech Modelling, Pp. 241-261, 1990.

[13] B.H. Story and I.R. Titze, "Voice Simulation with the body-cover model of the vocal folds," The Journal of Acoustical Society of America, vol.97, no.2, pp. 1249-1260, Feb. 1995.

[14] R.L. Whitehead, D.E. Metz, and B.H. Whitehead, "Vibratory Patterns of vocal folds during pulse register phonation," The Journal of Acoustical Society of America, vol. 75, No. 4, pp. 1293-1297, Apr. 1984.

[15] Barne, C.H. Shadle and P.O.A.L. Davies, "Fluid Flow in a dynamical mechanical model of vocal folds and tract.I. Measurements and theory," The Journal of Acoustical Society of America, vol. 105, No. 1, Pp. 343-356, Jul. 2000.