Journal of Engineering Technology
Volume 8, SI (Recent Trends in System Engineering), April 2019, PP. 74-89

RESEARCH ARTICLE -ENGINEERNG TECHNOLOGY

A Neoteric Approach For Dynamic Simulation Based Vocal Tract Modelling

R. Ashok Kumar Reddy[1], M. N. Giri Prasad[2] and G. N. Kodanda Ramaiah[3]

[1] Research Scholar, Department Of ECE, JNT University, Ananthapuramu, India.
[2] Professor, Department Of ECE, JNT University, Ananthapuramu, India.
[3] Professor, Department Of ECE, Kuppam Engineering College, Kuppam, India.

Abstract: Speech production process is an interesting discipline where the practical data is evaluated directly. In this work an exclusive system is used which is based on the motion of the articulators related to speech that comprises of Jaw, tongue , lips and rise and fall of soft palate. The computational techniques used for synthesizing speech based on models of the human vocal tract and the articulation processes occurring there is called Articulatory Synthesis (ASY). But the form of the vocal tract relies upon on the manage and hyperlinks among the articulators stated above. Along with the articulators, input sound source and the timing of movement of links also determines the summary of the vocal tract. Here, we aim to design an approximate phonatory model for acoustic calculation with a fully aerodynamic simulation, explicitly accounting for the propagation of sound along the tract, generating patterns of movement of simulated vocal tract articulators, and specifying the temporal relations among dynamically defined gestures that lead to a time-varying vocal tract filter function, and an acoustic waveform.

Key Words: speech recognition, speech articulators, vocal tract, Dynamic simulation..

1. Introduction

The Speech of a human is an output from various physiological components operating through an acoustic message from the speaker to the listener. The primary speech principle is discussion [1]. The signal which is able to express thoughts or message or data by articulating signals is defined as speech. A wave form which is acoustic in nature that transports the worldly data from a spokesperson to the listener. The transmission and reception of acoustic speech works properly even over a high number degree of distance.

From many works, it is identified that works dealing with vocal tract [2] comprises of the set of collective sounds that one may also produce. The vocal tract incorporates various automatically coupled components along with articulators as the jaw, lips, palate, tongue and velum [3] and in similar to the diverse bones, cartilages[4] and different smooth tissues also have an effect in sound generation. Each of those has its own dynamics and physics traits consists of resonance, compliance, inertia, mass and so on where, these traits imposes restrictions on the fixed arrangements that the vocal tract can reach. Even more on the self-vocal tract’s changing aspects, which confines the group of sequential sounds that someone might additionally go thorough. The fundamental requirements of the speech manufacturing shape are done through the lungs and glottis that carries the deliver characteristics and the vocal tract supports the feature of filtering. The fig.1 represents the mechanism of Speech Production.

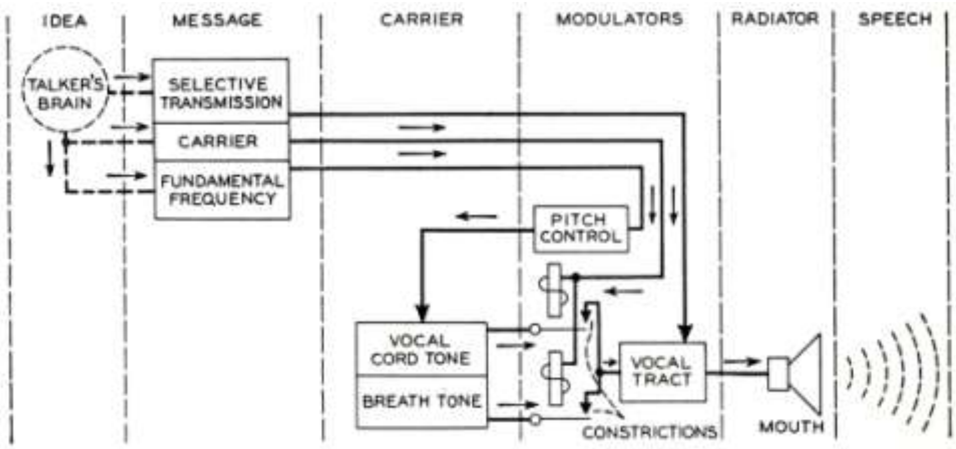

Fig 1: The mechanism of speech production.

Speech is shaped as a series of sound sequences. So, the vocal cords exchange the position form and the length of the articulators is every often to produce diverse forms of sounds. Three additives were seen as the effect of spoken words i.e.
Voice production = voiced sound + resonance + articulation

The purpose of this paper is to revise the differences seen in female and male spokespersons' vocal tract shapes & formants and also to use these differences in vocal tract shapes and formants for better speaker recognition system. As per earlier works, every vowel has its private vibration range with consequential distinct association of formants. For classification of few vowels, the primary four formants were identified. The outline of the vocal hollow discipline modifies the spectrum of the excitation signal to create a recognizable speech acoustics. This promotes conversion and made use of the vocal tract form glottis to lips and with various acoustic capabilities of the sound were seen. Therefore, from the considered vocal tract form account [5], the finalized sound received can be estimated pre-dominantly. The inverse conversion from the speech acoustics to the shapes of vocal tracts shouldn't be fully implicit. The drawback of defining the form of vocal tract from speech vibration is known as Inverse Vocal Tract problem.

This Inverse Vocal Tract problem has a mixture of sensible and theoretical applications. More than a few assessment strategies like intensity of lips acoustic impedance, formant frequencies' capacity and LPC specially placed on evaluation which is been used for the comparison of shapes of vocal tracts. LPC particularly established study of speech which is highest favored method, as it's equipped with delivering actual evaluation of the shape of vocal tract which is in far from the indicators of speech with present software/hardware skills. Further LPC coefficients may be changed to distinctive parametric units for investigating and estimating the intra-spokesperson vocal tract form. It is also suggested for presenting the spared formants for character audio system at particular intervals and investigate the inconsistency function on speakers' identification. This is advantageous in low rated ASR (automatic Speech Recognition) programs for self or financial security, precision detection applications, tele-banking and personal application software designated to entry the data base and information-bases.

2. Existing methods:

2.1 Estimation of shape of Vocal Tract- Direct Methods

A geometrical measurements of the vocal tract comprises of the application of X-ray imaging, ultrasonic imaging, magnetic resonance imaging (MRI), and so on [6].
Various X-ray researches signified the usage of static and video images. In cinefluorography, the images are photographed through X-rays after anticipating on a luminous display screen. This radiation risks brutally bound the content of records which is attained and in addition, it is not easy to find the smooth tissue area. But with the help of a radio opaque material, the tongue's level is outlined.
A micro beam's diameter of 1mm of X-Ray is used to minimize the exposure of radiation on the secure tiers and also identifies the movement of gold pellets interlinked to distinct articulators with dental adhesive. There are a few hazards seen for using X-ray micro beam which is expensive and leads to the constrained availability of the truth due to gag reflex of the tongue, that pellets can't be located on the rear portions. There may be complexity of the pellets (3mm diameter) which may additionally allow some degree of normal articulator in particular to the tongue tip.
Magnetic Resonance Imaging (MRI) is a method, used to attain a particular 3-dimensional vocal tract shapes of a spokesperson which are in relation to the particular set of consonants and vowels. For one male spokesperson, a set of 18 shapes are attained whose vocalization is scanned for 12 vowels, 3 nasals and three plosives. The3-D MRI images are segmented to present the airways, which in turn are examined to identify the location of vocal tract cross-section that as a distance attribute from glottis to lips.
The primary problem of MRI method in the study of speech is the long duration of scanning for data acquisition of total image set considered. The dependency of time parameter is in minutes in addition to this the MRI imaging is seen poor at the air tissue interface, teeth and bone.

2.2 Acoustic Measurements Used for the Estimation of Vocal Tract Shape

In general, for shape assessment of vocal tracts, indirect approaches are done with respect to audio measurement of the impedance response or impulse response on the lips [7].
The earlier literature presented the usage of phantom techniques. In this technique, the evaluation of spectrum is done based on short windows (16-32ms) of speech. Investigations that deal only with speech from models instead of human speech requires no longer carry out spectral evaluation; due to the fact the spectral records is generated unswervingly in the speech production version.
From the signal received from speech, the evaluation of speech can be done by guessing the predictor co-efficient slightly, that which explains the shape of a speech wave with spectral properties defining through a digital filter. In LPC analysis, the predominant characteristic seen is the separation of source effects (the glottal impulse or noise) and the clear out (the vocal tract).If the specified models are correct then the source's result and the glottal impulse are said to be unbiased depending on the vocal tract's properties. Also if the predictor coefficients are found to be best are identified, then the modelled source enumerates completely that includes the prediction mistakes and the vocal tract's switch representation were meant to be explained flawless.
Neither of those assumptions kept in exercise changes out the predictor coefficients also create a spectrum that is similar and smoother than log power spectrum. The literature survey enumerates the applicability of LPC spectrum and methods to strengthen the smoothened speech spectrum. Some research reviews highlighted the efficiency of LPC coefficients for accounting the acoustics considered.

3 Proposed method

3.1 Vocal Tract Shape Estimation for Vowels using LPC:

For speech analysis (Durbin's Recursive Algorithm), Auto regression approach is preferred centered on Linear Prediction has been used. This process is recognized as LP Modeling and also considered as AR Modeling. The system's output is prior here for this type of modelling. The simplest method of a vocal tract contains co-axially linked number of tubes of cylindrical in shape provides a transfer function for all-poles. For estimation of Vocal tract form, the reflection co-efficient is used which is attained from speech alerts through LPC evaluation using Wakita's speech analysis model and for effective inverse filtering, Durbin's algorithm is considered. From glottis to lips, the vocal tract dimension may be seen as 17cm long of a male-female. Here the values are measured through vocal tracts, for the average vowels vibration from male and female audio system are obtained.
From the speech production model, it is identified that the speech experiences a spectral angle change of –6 dB/octave. To increase the frequency range and to flatten the spectrum, a pre emphasis filter is used to counter this drop which encountered with a 6 dB per octave rate. This pre-emphasized speech signal is shown in fig.2a, which is windowed using Hamming windows.

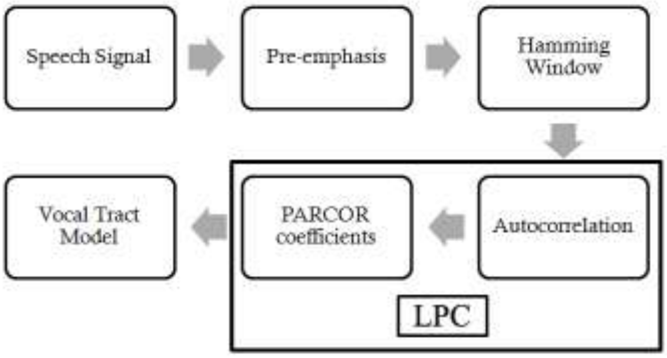

Fig 2a: Block diagram for vocal tract shape calculation.

3.1.1 Pre-emphasis

6 dB per octave rate is applied through Pre-emphasis that results the increase of the amplitude for 6 dB/ octave. The pre-emphasized signal is presented in fig.2b. Here the signal of speech is flattened is marked which resulted in a higher outcome for coefficients calculation using LPC.

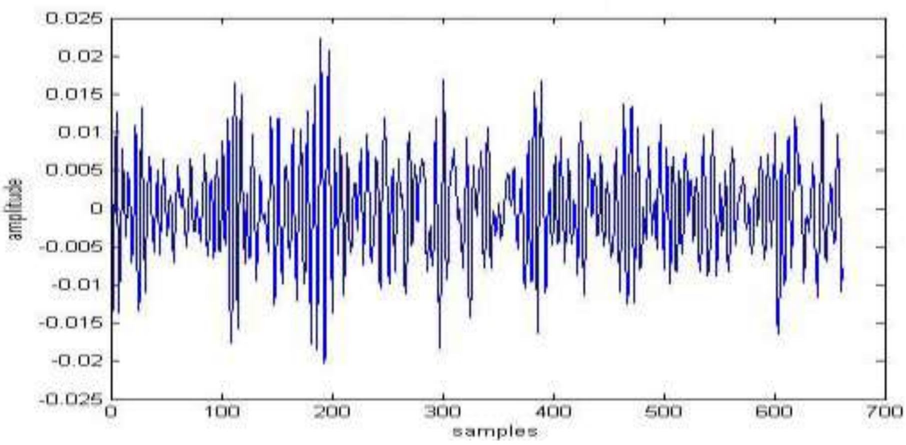

Fig. 2b Pre-emphasized Signal.

3.1.2 Window Analysis

With help of hamming window, the LP analysis is carried on weighed frames where this window, w(n), is selected to produce a better balance in between width of main lobe and attenuation of side lobe.

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) & ;\ 0 \le n \le N-1 \\ 0 & ;\ otherwise \end{cases}$$

The accuracy of vocal tract's transfer function approximation is done with help of the Hamming window which is adequate in nature. The length of Hamming windows is fixed as 30ms with 10ms overlap is applied as to attain smooth estimation. The fig. 2c shows speech signal under windowed speech condition.

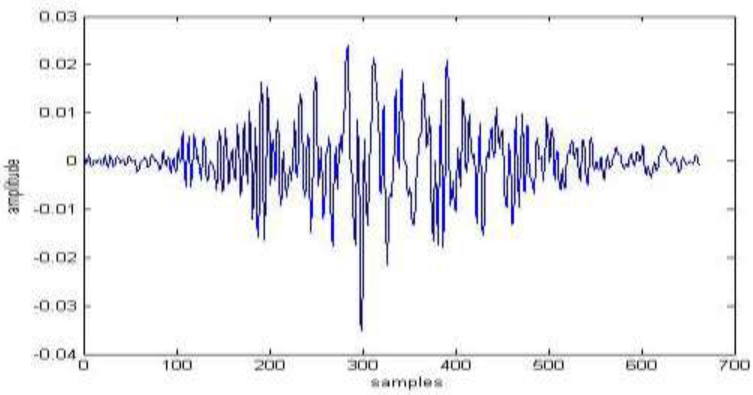

Fig.2c The windowed signal.

3.1.3 Auto-Correlation Analysis

Auto correlation is done after each frame of the signal is said to be windowed signal. Here in this method, $s_n[n]$ is termed as analysis segment which is identically zero outside the interval defined $0 \le n \le N-1$,

$$s_n(n) = s(n+N)\, w(n)$$

This above equation represents the Auto correlation analysis of the windowed signal.

3.1.4 PARCOR (Partial Auto-Correlation Coefficients Extraction)

The extraction of partial Auto correlation coefficients are set as processing components which are calculated through the analysis of Levinson- Durbin recursion strategy. It is found that the intermediate values were quantized with less complexity than the predictor co-efficient straight quantizing the components. This results in small variations that lies in the predictor co-efficient which implies to a considerable large variations in the selected positions of pole. To make certain steadiness of the co-efficient of filter components, the poles and zeros are in need to be inside the unit circle of the plane named as Z-plane. Therefore, the excess 8-10 bits per co-efficient are required to maintain the accuracy. For the PARCOR coefficient's stability maintenance, the necessary and sufficient condition is bounded with +1 or -1. The fig.2d represents the speech signals PARCOR coefficients.

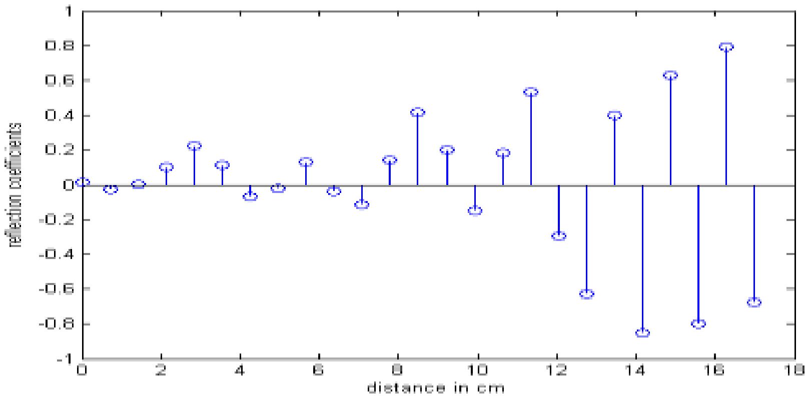

Fig.2d The reflection (PARCOR) co-efficients for a frame.

3.2 Classification of Vowels Based On Formants

The formants are defined as resonant frequencies of the vocal tract which are also named as formants frequencies. Depending on the vocal tract's dimensions and form, these formant frequencies are generated with help of the linked cylindrical tubes and vocal-cavity resonances. This Vocal tract form is defined by the set of formant frequencies, and exceptional sounds that are generated through different vocal tract forms. The main property of spoken speech is nothing but termed as spectral time variance. The Implementation procedure for formant estimation using peak detection method is summarized as follows:

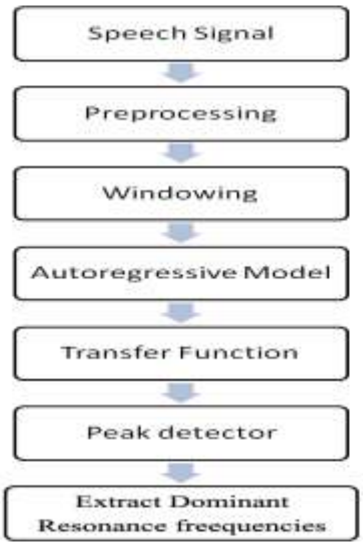

Fig 3a: Block diagram of formant estimation using peak detection.

3.2.1 Preprocessing

The first step of speech signal processing is Pre-processing where the analog speech signal is converted into digital form. It is very crucial step for enabling further processing.
Spoken vowels, the sounds of interest in this study, are recorded with an ensemble of 30 adult male and 30 female native Andhra speakers of Indian English, 30 times at random and once temporarily. Recordings are saved and the analogue data are digitized at 12000 Hz sampling frequency.

3.2.2 Windowing

Windowing is needed to work with short term edges of the speech signal in direction to select a portion of the speech signal that is assumed to be stationary. It is needed to avoid unnatural discontinuities in the speech segment and distortion in the underlying spectrum to ensure that all parts of the speech signal are recovered and possible gaps between frames are eliminated.

3.2.3 Auto-Regressive Model (ARM)

A model which is dependent simplest on the earlier outputs of the process is often called an auto-regressive model (AR), even as a technique which relies only on the inputs of the process is often called a moving average model (MA).The technique founded on each inputs and outputs is an autoregressive- moving-average method (ARMA). The formant or resonant frequencies are evaluated as proven in Fig 3b.

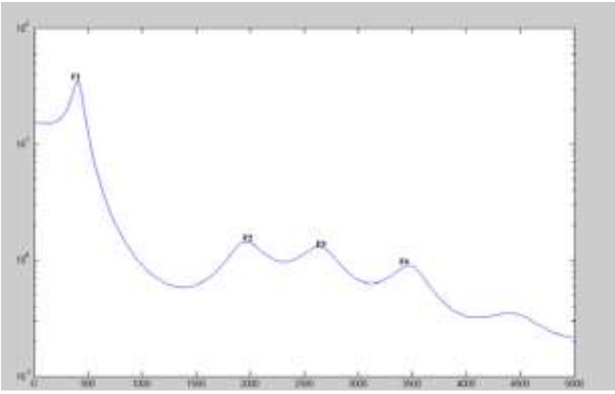

Fig 3b: The frequency response of ARM.

3.2.4 Vocal Tract (VT) Transfer Function

It is seen that the resonance (formants) of speech corresponds to the poles of the transfer function V(z). An all-pole model is a very good representation of vocal tract Frequency (Hz) Energy effects for a majority of speech sounds. The equation of vocal tract transfer function is shown below

$$V(z)=\begin{cases} G/\left(1-\sum r_K Z^{-k}\right) \\ 0 \quad whereK=1 \end{cases}$$

3.2.5 Peak Detector

In direction to detect the resonant points of formants F1, F2, F3 and F4, the peak detection function is implemented. It is observed that positive and negative polarity peaks occur at points of positive to negative and negative to positive slope adjacency respectively.

3.3 Algorithm for Intra Speaker Formants Estimation

In order to quantify the mean of formants, a simple algorithm is shown below:

1. In the first step, for vowel /a/, 30 samples of sound from same speaker at different intervals of time is collected.

2. Next formants F1, F2, F3 and F4 are calculated.
3. The formants F1, F2, F3 and F4 of each samples are calculated.
4. Steps 1-3are repeated for vowels /e/, /i/, /o/, /u/.

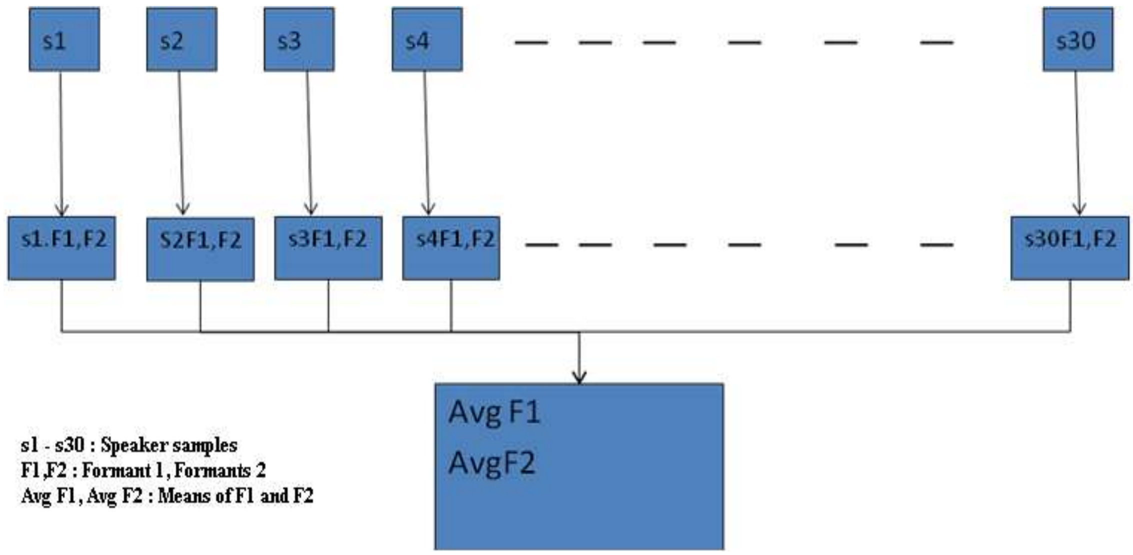

Fig 4a: Block diagram for intra-speaker formants estimation for vowel.

3.4 Algorithm for Inter-Speaker Formants Estimation for Vowels

In this section we have proposed a new model for inter-speaker formants estimation for vowels as shown in the Fig. The algorithm developed is described as follows:

1. In the first step, for vowel /a/, 30 samples of sound from different speakers at different intervals of time are collected.
2. Next formants F1, F2, F3 and F4 are calculated.
3. The mean of formants F1, F2, F3 and F4 of each samples are calculated.
4. Steps 1-3are repeated for vowels /e/, /i/, /o/, /u/.

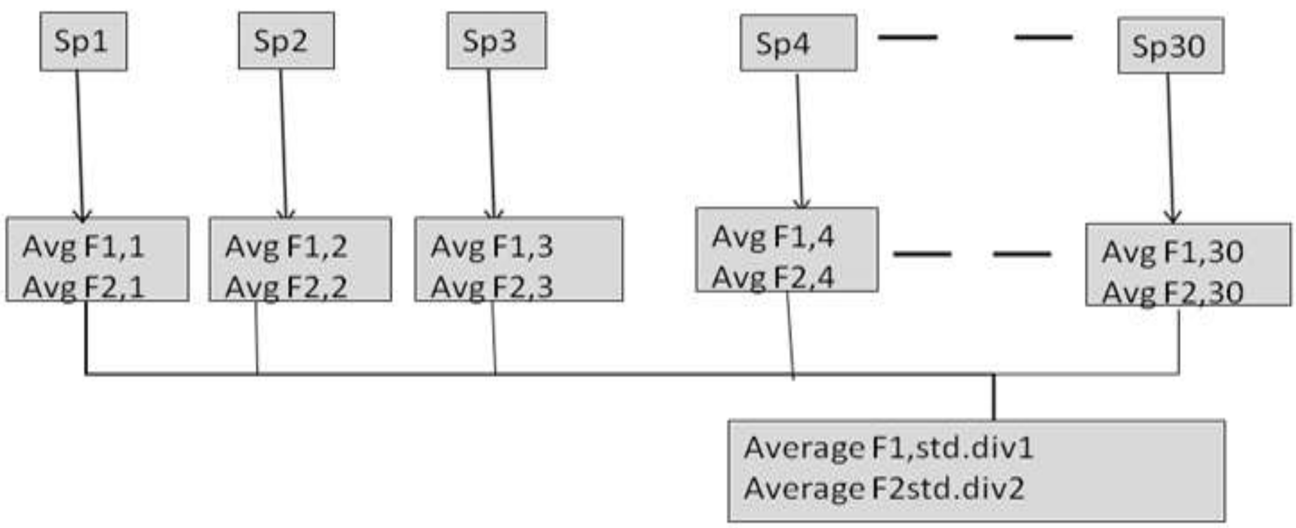

Fig 4b: Block diagram for inter-speaker formants estimation

3.5 Independent Speaker Recognition for Vowels Using Euclidean Distance:

Euclidean distance measure is applied in an effort to measure the similarity or the dissimilarity between two spoken vowels, which take place after quantizing a spoken vowel into its code book. The matching of an unknown vowel is performed by using measuring the Euclidean distance between the facets vector (formants) of the unknown vowel to the reference model (codebook) of the known vowel formants F1, F2 within the database.

$$d(x,y) = \sqrt{\left[\sum_{i=1}^{D}\{wi(xi - yi)2/\sigma i\}\right]}.$$

Based on the above consideration and by using decision rule the vowel can be recognized.

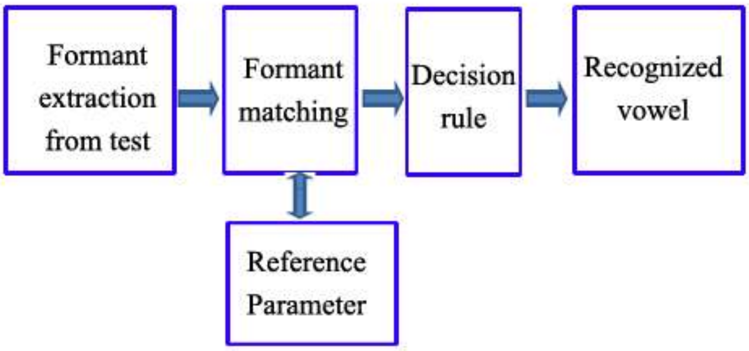

Fig 4c: Block diagram of vowel recognition

4. Results and discussions

4.1 Formants Estimation of Vowels for Male Speakers

The vocal tract shapes which are estimated from vowels gross vocal tract shapes arrived at from F1-F2 are located to create the largest spread among the vowels, among individual speakers. We have regarded on the spread of formants for man or woman audio system at one of the kind occurrences and investigated the variety position on speaker identification & speaker specific-recognition applications.

4.1.1 For Vowel /a/

Time domain representation and spectrum of vowel /a/ is shown in fig 5a and fig 5b. The Table 1 shows formants observed for vowel/a/.

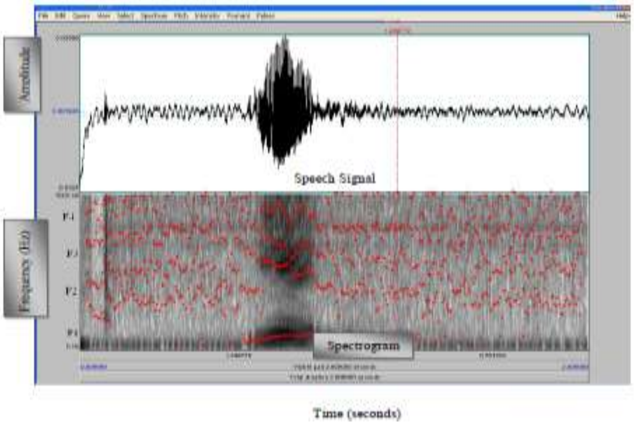

Fig. 5a. Time domain representation of vowel /a/

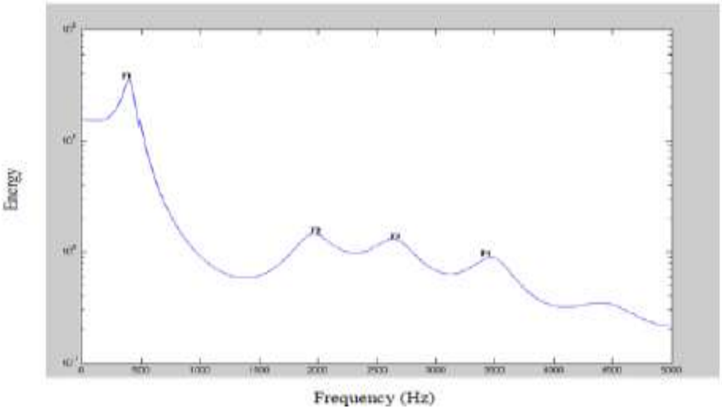

Fig. 5b. Spectrum for vowel /a/

Table 1 formant frequencies for Vowel /a/ in Hz

Formants	Frequency(Hz)
F1	412.6
F2	1947.0
F3	2654.7
F4	3489.2

Table 2 Overall results of formant frequencies of vowels

Vowels Male	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/a/	412	1947	2654	3489
/e/	325	2166	2477	3440
/i/	523	1533	2337	3414
/o/	500	1016	2311	3918
/u/	349	1127	2196	3366

4.2 Formant Estimation of Vowels for Female Speakers

4.2.1 For Vowel /a/

Time domain representation and Spectrum of vowel /a/ is shown in fig 5c and fig 5d. The Table 3 shows formants observed for vowel/a/.

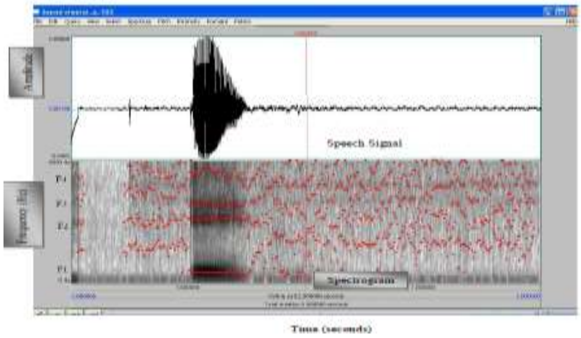

Fig. 5c. Time domain representation for vowel /a/

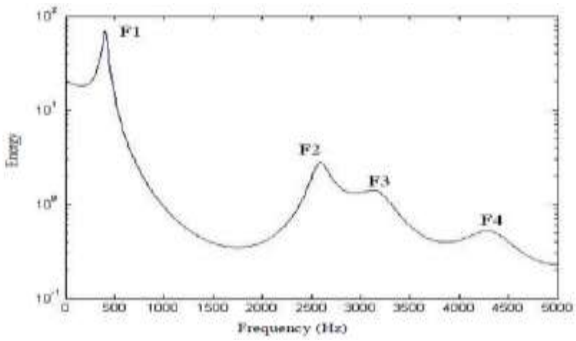

Fig. 5d. Spectrum for vowel /a/

Table 3 Formant frequencies for vowel /a/ in Hz

Formants	Frequency(Hz)
F1	410.2
F2	2587.9
F3	3125.0
F4	4287.1

Table 4 The results of formant frequencies of vowels

Vowels female	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/a/	410	2588	3125	4287
/e/	234	2851	3593	4385
/i/	244	781	1787	3095
/o/	429	859	2128	3252
/u/	322	1728	2832	3964

Table 5 Comparison of formant frequencies F1, F2, for vowel /a/, using peak detection of spectrum using ARM with Praat software

For Vowel /a/	F1 (Hz)		F2 (Hz)	
	Proposed Method	Praat Software	Proposed Method	Praat Software
Sample 1	368	348	1975	2085
Sample 2	401	385	2061	2088
Sample 3	410	424	2570	2558
Sample 4	431	410	2572	2462
Sample 5	380	388	2343	2395

Further investigation is carried out to show how closely the plots for different speakers match up for any of the vowels. These plots are demonstrative of the similarities in the vowel speech patterns of various speakers. The fig 5e, show all the speakers frequency response plots.

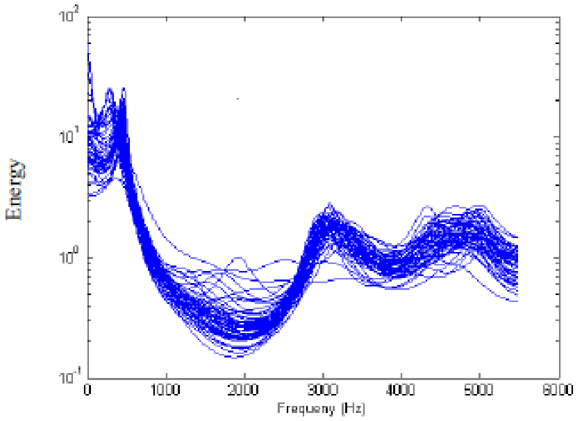

Fig.5e Auto-Regressive model of vocal tract for vowel /a/

Table 6 Percentage recognition for vowel of male speakers

Vowels	Predicted					
Actual	**/a/**	**/e/**	**/i/**	**/o/**	**/u/**	**% Correct**
/a/	48	1	1	0	0	96
/e/	3	47	0	0	0	94
/i/	5	2	43	0	0	86
/o/	0	3	1	46	0	92
/u/	2	1	1	0	46	92

Table 7 Percentage recognition for vowel of female speakers

Vowels	**Predicted**					
Actual	**/a/**	**/e/**	**/i/**	**/o/**	**/u/**	**% Correct**
/a/	46	4	0	0	0	92
/e/	2	48	0	0	0	96
/i/	2	0	47	0	1	94
/o/	0	1	0	49	0	98
/u/	0	3	0	0	47	94

The results are shown in fig 6a and 6b for male and female speakers respectively. It is concluded that overall recognition rate and vowel classification results obtained by Euclidean distance method are better than that are obtained by MLR method.

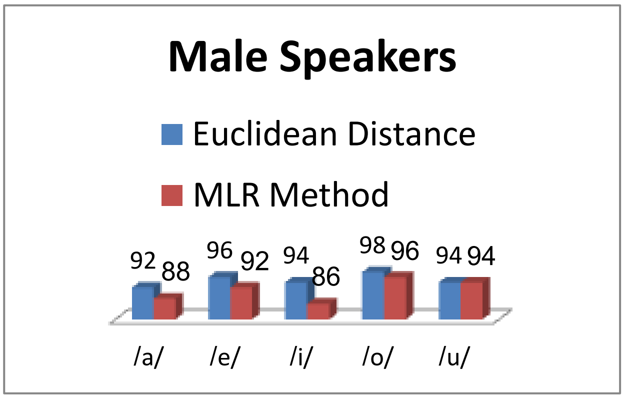

Fig.6a. Comparison of Euclidean distance technique with MLR using Standard deviation for male speakers

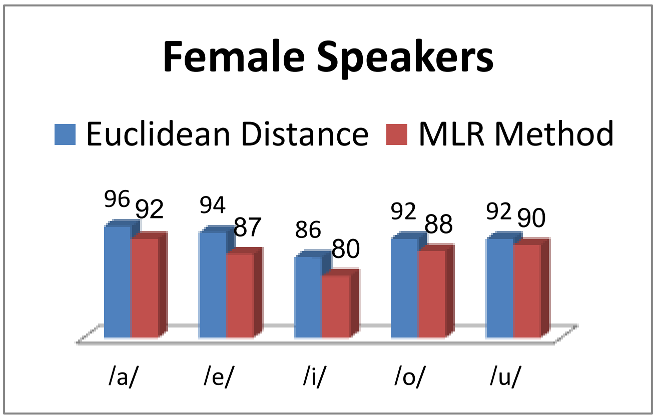

Fig. 6b. Comparison of Euclidean distance technique with MLR using Standard deviation for female speakers

5. CONCLUSION

Wakita's Speech Analysis Model is used to investigate the reflection image coefficients which are collected through intra and inter-speaker estimation of vocal tract shapes of speech signals obtained from LPC. In this research, by selection of 24 reflection coefficients with a 22,100Hz sampling frequency, the dynamic vocal tract modeling is done on vowels and the dynamic models for vowels are derived. And from the research we can conclude that the reflection image coefficients on the lips and place values for vowels specifically /a/, /e/, /i/, /o/ and /u/ values range is attained respectively. "When compared to male speakers the female speakers have better rate for vowel recognition" is also identified in this work. The following table table8 help us in comparing the values.

Table 8 Comparison of Percentage recognition for vowel of female speakers using Euclidean and MLR Methods

Male		Female	
% Of recognition using Euclidean Method	**% Of recognition using MLR Method**	**% Of recognition using Euclidean Method**	**% Of recognition using MLR Method**
92	88	96	92
96	92	94	87
94	86	86	80
98	96	92	88
94	94	92	90

From this research work another conclusion can also be drawn that the usual reputation rate and vowel classification results through Euclidean distance technique provides better results than MLR approach and also identified as the vowel recognition charge is better for female spoke persons when compared to the male spoke persons. Table 8 provides the numerical data also which represents that the Euclidean distance technique has the higher percentage of recognition when compare to recognition rate in MLR method.

6. References

[1] L.R.Rabiner and R.W.Schafer, Digital processing of Speech alerts, Droling Kindersly(india)pvt.Ltd., licensees of pearson education in south Asia, 1978, PP. Fifty 4-a hundred and one, 412-460.
[2] Thomas F. Quatieri, Discrete time speech sign processing thoughts and practice, 2002, pp 56-59.
[3] Philip Rubin and Louis Goldstein, Articulatory Synthesis, Haskin laboratories, 1995, pp 3-eight.
[4] Nandaprasetiyo Koestoer B.Eng, Robust Linear prediction analysis for low bit rate coding, 1998.
[5] J.F.Curtis, (Ed.), Processes and Disorders of Human Communication. New York: Harper and Row, 1978.
[6] R. S. Nikerson, Characteristics of the speech of deaf males and females, Volta Rev., vol. Seventy seven, 1975, pp. 342–362; reprinted in: H. Levitt, J. M. Pickett, and R. A. Houde Sensory Aids for the Hearing Impaired, (Eds.). New York: IEEE Press, 1980, pp. 540–545.
[7] M.R. Schroeder, Determination of the geometry of the human vocal tract thru acoustic measurements, J. Acoust. Soc. Am., vol. Forty one, no. Four, pt. 2, 1967, pp. 1002–1010.

[8] J. Schroeter and M. M. Sondhi, “Techniques for estimating vocal tract shapes from the speech signal,” IEEE Trans. Speech Audio Process., vol. 2, no. 1, pt. 2, , 1994, pp. 133–one hundred fifty.

[9] B.S. Atal and S. L. Hanauer, “Speech evaluation and synthesis by manner of the use of linear prediction of the speech wave,” J. Acoust. Soc. Am., vol. 50, no. 2, 1971, pp. 637–655.

[10] H. Wakita, “Direct estimation of the vocal tract shape with the useful resource of inverse filtering of acoustic speech waveforms,” IEEE Trans. Audio Electro Acoust., vol. 21, no. Five, 1973, pp. 417–427.

[11] G. Fant, Acoustic Theory of Speech Production. The Hague: Mouton, 1960.

[12] M.M. Sondhi and B. Gopinath, “Determination of vocal-tract shape from impulse response on the lips,” J. Acoust. Soc. Am., vol. Forty nine, no. 6, pt. 2, 1971, pp. 1867–1873.

[13] M. M. Sondhi and J. R. Resnick, “The inverse trouble for the vocal tract: numerical techniques, acoustical experiments, and speech synthesis,” J. Acoust. Soc. Am., vol. Seventy three, no. Three, 1983, pp. 985–1002.

[14] H. Wakita, “Normalization of vowels with the aid of vocal-tract length and its software program to vowel identity,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-25, 1977, pp. 183–192.

[15] P. Ladefoged, R. Harshman, L. Goldstein, and L. Rice, “Generating vocal tract shapes from formant frequencies,” J. Acoust. Soc. Am., vol. Sixty 4, no. Four, 1978, pp. 1027–1035.

[16] C. H. Coker, “A model of articulatory dynamics and manage,” Proc. IEEE, vol. 64, no. 4, 1976, pp. 452–460.

[17] Z. Yu and P. C. Ching, “Determination of vocal-tract shapes from formant frequencies based totally on perturbation principle and interpolation approach,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1996, pp. 369–372.

[18] Y. Laprie and B. Mathieu, “A variational technique for estimating vocal tract shapes from the speech signal,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, 1998, pp. 929–932.

[19] A. E. Mahdi, “Visualization of the vocal tract primarily based mostly on estimation of vocal vicinity talents and formant frequencies,” in Proc. Eighth European Conference on Speech Communication and Technology (Geneva, Switzerland), 2003, pp. 2381–2384.

[20] D. Rossiter, D. M. Howard, and M. Downes, “A actual-time LPC-based definitely vocal tract vicinity display for voice improvement,” J. Of Voice, vol. Eight, no. Four, 1994, pp. 314–319.

[21] K. Livescu, Feature-Based Pronunciation Modeling for Automatic Speech Recognition, Ph.D. Thesis, MIT Department of Electrical Engineering and Computer Science, 2005.

[22] C. S. Blackburn, Articulatory Methods for Speech Production and Recognition, Ph.D. Thesis, University of Cambridge, 1996.

[23] L. Deng, DYNAMIC SPEECH MODELS — Theory, Algorithms, and Applications, Morgan and Claypool Publishers, May 2006.

[24]Gully, A. J., Yoshimura, T., Murphy, D. T., Hashimoto, K., Nankaku, Y. & Tokuda, K., “Articulatory Text-to-Speech Synthesis Using the Digital Waveguide Mesh Driven with the resource of a Deep Neural Network”, Proc. Of InterSpeech 2017, pp. 234-238, Stockholm, Sweden, Aug. 20-24, 2017.

[25] J.M.Pardo “vocal tract form evaluation for kids”, in proc IEEE int.Conf.Acoust, speech sign processing, 1982, pp 763-766.

[26] S.H.Prak, D.J.KIM.Et al., “Integrated speech schooling gadget for paying attention to impaired”, IEEE prous rehab engg, vol 2, no.Four, 1994,pp 189-196.

[27] S. Hiroya and M. Honda, “Estimation of articulatory actions from speech acoustics the usage of an hmm-primarily based speech production version,” IEEE Transactions on Speech and Audio Processing, vol. 12(2), pp. One hundred seventy five–185, 2004.

[28] V. Mitra, H. Nam, C. Espy-Wilson, and E. Saltzman L. Goldstein, “Noise robustness of tract variables and their software program software to speech popularity,” in Interspeech, Brighton, UK, September 2009.

[29] Z. Al Bawab, L. Turicchia, R. M. Stern, and B. Raj, “Deriving vocal tract shapes from electromagnetic articulograph records thru geometric model and matching,” in Interspeech, Brighton, UK, September 2009.

[30] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum threat from incomplete facts via the em algorithm,” Journal of the Royal Statistical Society, vol. 39(1), pp. 1–38, 1977.

[31] H. Hermansky, D. Ellis, and S. Sharma, “Tandem connectionist feature extraction for classic hmm systems,” in ICASSP, Istanbul, Turkey, June 2000.

[32] R. Singh, M.L. Seltzer, B. Raj, and R.M. Stern, "Speech in noisy environments: Robust computerized segmentation, characteristic extraction, and hypothesis mixture," in ICASSP, Salt Lake City, Utah. USA, May 2001.
[33] X. Li, Combination and Generation of Parallel Feature Streams for Improved Speech Recognition, Ph.D. Thesis, Carnegie Mellon University ECE Department, Pittsburgh, PA., February 2005.
[34] M. Brand, "Structure gaining knowledge of in conditional hazard fashions via an entropic in advance and parameter extinction," Neural Computation, vol. 11, pp. 1155–1182, 1999.
[35] Paulraj M.P, Sazaliyaacob, ShahrulAzmi M.Y. "Vowel magnificence based on frequency Response of Vocal Tract", intending of the global conference on computer and verbal exchange engineering 2008, pp.1125-1129.
[36] MILAND Shah Investigation of VT shape estimation for the duration of Stop closures, 2009 pp 10-22.
[37] Sami Lemmetty evaluate of speech synthesis era. The draw close's thesis, 1999, pp four-6.
[38] Ahmed AliSa fiullah Bhatti, dr.Munammad Sleem Miam, formants based Analysis for speech recognition, IEEE 2006.
[39] Jonatham Harrington Acoustics Phonetics.
[40] H.Wakita, "Estimation of vocal-tract shapes from acoustical evaluation of the speech wave: the country of the paintings," IEEE Trans.Acous., Speech, Signal processing, vol.AU-21, no.5, 1973, pp.417-427.
[41] Analysis of Vocal Tract Shape Variability based on Formant Frequency Ratio at Various Conditions of Vowels for Indian English Speakers, Anil Kumar Chandrashekar [1*], M. B. Manjunatha. [2].
[42] Pitch, Formant Frequencies, Intensities as Speech Cue to perceived age.
[43] A Probabilistic Feature Based SVM Model for English Speech Recognition Priyanka Bansal and Syed Akhtar Imam.
[44] On a Transition Detection Method by Flattering the Spectrum of Speech Signal Seong-Geon Bae and Myung-Jin Bae.
[45] Social arithmetic robot for childhood education using voice-recognition and face detection based interaction Eric Kleist and Yong Zhu.