Table 1. short and long vowels

Short Vowels			Long Vowels		
Vowel	Vowel sign	Example	Vowel	Vowel sign	Example
അ a **/a/**		ക ka /ka/	**ആ ā** **/a:/**	◌ാ	കാ kā /ka:/
ഇ i **/i/**	◌ി	കി ki /ki/	**ഈ ī** **/i:/**	◌ീ	കീ kī /ki:/
ഉ u **/u/**	◌ു	കു ku /ku/	**ഊ ū** **/u:/**	◌ൂ	കൂ kū /ku:/
എ e **/e/**	െ◌	കെ ke /ke/	**ഏ ē** **/e:/**	േ◌	കേ kē /ke:/
ഒ o **/o/**	െ◌ാ	കൊ ko /ko/	**ഓ ō** **/o:/**	േ◌ാ	കോ kō /ko:/

Table 2. Malayalam vowel phonemes and their positions

	Short			Long		
	Front	**Central**	**Back**	**Front**	**Central**	**Back**
Close	ഇ i /i/		ഉ u /u/	ഈ ī /i:/		ഊ ū /u:/
Mid	എ e /e/		ഒ o /o/	ഏ ē /e:/		ഓ ō /o:/
Open		അ a /a/			ആ ā /a:/	

Front- In mouth, the tongue is in front position
Central-The tongue is further back
Back -The tongue is in the back positionHigh -The tongue is high
Mid -The tongue is lower
Low-The tongue is low position

II. FORMANT FREQUENCY AND ACOUSTIC VOWEL SPACE

The acoustic resonance of the vocal tract refers to the formant frequency in the speech signal. In the power spectrum, the formant frequencies appear as peaks. Though there are several formants, the first two formant frequencies (F1, F2) are mainly used to define the quality of vowels, and it depends on the shape of the vocal tract and position of the tongue. It also depends on the rounding of lips. There is a unique set of tongue position and movement for each language and each voiced sound has a set of formant frequencies. It has a unique physical property associated with each vowel. The first formant frequency F1 has a frequency range of 300-1200Hz, and the second formant frequency F2 ranges between 800-3000Hz.

The first formant value F1 of a vowel is inversely proportional to vowel height or tongue height. The second formant F2 is associated with frontness or backness of the tongue body. If F1 is high, there is a low vowel height and if F2 is high, the tongue is in the front position [4]. The formant frequencies are the most common parameters used to characterise the vocal tract in speech analysis.

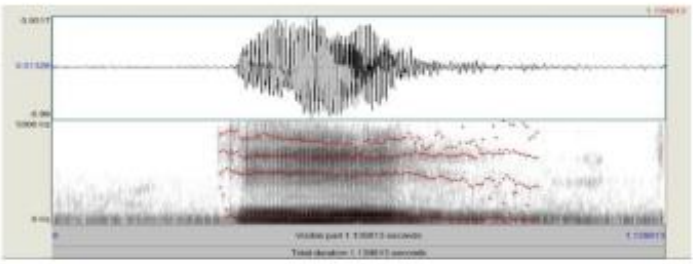

Figure 2. An example of Frequency spectrum and formant frequency plot of a Malayalam vowel phoneme

Acoustic descriptions of speech sounds, especially the formant patterns of vowels give greater accuracy. It can then be used in various areas like phonetics, speech pathology, speech recognition, etc. [5].

Catford (1988) talks about the concept called "vowel space" and "vowel limit". He says that the idea of the Cardinal Vowels by the renowned phonetician Daniel Jones is based on the concept that the vowels are limited by vowel space or limit. In the production of a vowel, there is a certain fixed area or space within oropharyngeal cavity, beyond which the vowel takes space of an approximate type of sound. Thus, theoretically speaking, "any vowel of any language must have its tongue-position either on the vowel limit itself, or within the Vowel Space "(Catford, 1988:130) [6][7].

III. LPC MODEL

In speech processing, Linear Predictive Coding (LPC) is used to represent the spectral envelope of a speech signal in compressed form. LPC method is one of the popular methods in Speech analysis. There are many works available which uses Linear prediction for speech analysis [8]. LPC can be used to determine the recurrence of the vocal cord vibration, vocal tract shape, spectral frequencies, and bandwidths [9]. To extract Formant frequencies, the most commonly used technique is linear predictive analysis because of the ability to provide accurate estimates and relative speed of computation.

The basic steps of LPC process includes the following:

Pre-emphasis: The digitized speech signal, s(n), is put through a low order digital system, to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing. The output of the pre-emphasis network, is related to the input s(n) of the network given by the difference equation:

$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1)$$

Frame Blocking: The output of pre-emphasis step, $\tilde{s}(n)$, is blocked into frames of N samples, with adjacent frames being separated by M samples. If $x_l(n)$ is the l^{th} frame of speech, and there are L frames within the entire speech signal, then

$$x_1(n) = \tilde{s}(Ml + n)$$

where n = 0, 1, …, N – 1 and l = 0, 1, …, L – 1

Windowing: After frame blocking, the next step is to window each individual frame so as to minimize the signal discontinuities at the beginning and at end of each frame. If we define the window as w(n), $0 \le n \le N - 1$, then the result of windowing is the signal:

$$\tilde{x}(n) = x_l(n)w(n)$$

where $0 \le n \le N - 1$

A typical window is the Hamming window, which has the form

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right] \qquad 0 \le n \le N-1$$

Autocorrelation Analysis: The next step is to auto correlate each frame of windowed signal in order to give

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}(n)\tilde{x}(n+m) \qquad m = 0,1,\ldots.p$$

where the highest autocorrelation value p, is the order of the LPC analysis.

LPC Analysis: The next processing step is the LPC analysis, which converts each frame of p+1 autocorrelations into LPC parameter set by using Durbin's method. This can formally be given as the following algorithm:

$$E^{(0)} = r(0)$$

$$k_i = \frac{r(i)\sum_{j=1}^{i-1} \alpha_j^{i-1} r(|i-j|)}{E^{i-1}} \qquad 1 \le i \le p$$

$$\alpha_i^{(i)} = k_i$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i\alpha_{i-j}^{(i-1)} \qquad 1 \le j \le i-1$$

$$E^{(i)} = (1 - k_i^2)E^{i-1}$$

Recursively for i = 1,2, …, p, the LPC coefficient, a_m , is given as

$$a_m = \alpha_m^{(p)}$$

LPC Parameter Conversion to Cepstral Coefficients: LPC cepstral coefficients, is a very important LPC parameter set, which can be derived directly from the LPC coefficient set. The recursion used is

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) . c_k . a_{m-k} \qquad 1 \le m \le p$$

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m}\right) . c_k . a_{m-k} \qquad m > p$$

The LPC cepstral coefficients are the features that are extracted from voice signal and these coefficients are used as the input data of Artificial Neural Network. In this system, voice signal is sampled using sampling frequency of 8 kHz and the signal is sampled within 0.5 seconds, therefore, the sampling process results 4000 sampled data. Because we choose LPC parameter N = 240, m = 80, and LPC order = 12 then there are 576 data of LPC cepstral coefficients. These 576 data are used as the input of artificial neural network [10] [11].

IV. METHODS AND MATERIALS

This study deals with 40 native normal adults (20 males and 20 females) in the age group of 18-45 years. Malayalam consists of 36 consonants and 15 vowel symbols. Though there are 15 vowel symbols in the Malayalam script, only ten vowel phonemes were selected. The vowel phonemes are /a/, /a:/, /i/, /i:/, /u/, /u:/, /e/, /e:/, /o/, and /o:/.

In the analysis, vowel utterances were recorded by using speech sampling at a sampling rate of 16KHz. Noise was filtered in the pre-processing step. The microphone was kept at a distance of 10cm away from the speaker. The analysis was performed by using LPC method to give the values of formant frequencies.

V. RESULTS

The mean value of F1 and F2 were computed for each vowel, and for each speaker. Table 3 shows the average values of F1 and F2 for different Malayalam vowel phonemes.

Table 3. Average Formant Frequencies of Malayalam Vowel Phonemes For Male and Female

Formants	**F1**		**F2**	
Vowels\Gender	**M**	**F**	**M**	**F**
a	730	848	1226	1336
a:	725	813	1220	1320
i	346	368	2345	2822
i:	282	290	2025	2328
u	326	355	862	892
u:	317	323	750	806
e	458	494	2010	2774
e:	439	475	2108	2631
o	457	476	808	983
o:	441	457	798	927

The table shows the comparison of formant frequencies for males and females. In the comparison, it is understood that the formant frequencies of females were comparatively higher than males, and also the short vowel formants have a slightly higher value than long vowels.

Acoustic vowel space is a tool that explains how the formant frequencies help accurately to determine the vowel space in a language. It is also useful for distinguishing individual speakers. The area of the acoustic vowel space was plotted using SigmaPlot software. Figure 3 shows the acoustic vowel space comparison for male and female in Malayalam.

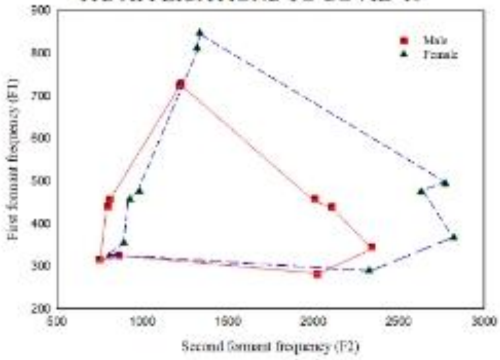

Figure 3. Male and Female acoustic vowel space in Malayalam

It is also showing the vowel space representation of the resultant formant table. Points in the plot indicates the Malayalam vowel phoneme formants for males and females. From Table 3 and Figure 3, it is noted that the formant frequencies F1 and F2 are higher in females, which is attributed to the fact that the vocal tract of females is comparatively smaller than males.

VI. CONCLUSION

From the examination of formant frequencies of various Malayalam vowel phonemes, it is found that there is a definite variation in F1 and F2 for each vowel utterance. Results show that F1 and F2 are greater in short vowels than long vowels. There is no relation between F1 and F2 for a particular vowel phoneme. From the comparison of formant frequencies, it is found that, the formant values of female vowel utterances are comparatively greater than that of a male speaker. It also shows that, the third formant frequency F3 and forth formant frequency F4 do not play a pivotal role in determining the quality of vowels. The acoustic vowel space has an upward shift of F1 and F2 in female utterances compared to male.

VII. REFERENCES

[1] Lekshmi K.R, Jithesh V. S, Elizabeth Sherly, "Malayalam Speech Corpus: Design and Development for Dravidian Language", Language Resources and Evaluation Conference (LREC 2020), Marseille, 11-16 May 2020.

[2] Paul W. Justice, Relevant Linguistics: An Introduction to the Structure and Use of English for Teachers (2nd edition, revised and expanded), California: CSLI Publications, 2004, p. 15.

[3] Gerald Kelly, How to teach Pronunciation, (England: Longman, Pearson Education Limited, 2000) , p.29 Medhi Bhargab, TALUKDAR P. "Different acoustic feature parameters ZCR, STE, LPC and MFCC analysis of Assamese vowel phonemes",ICFM 2015.

[4] Tsz Yin Wong, "Acoustic Analysis of Mandarin Vowels Pronounced by Macao and Hong Kong Cantonese Speakers", Universitat Pompeu Fabra, Màster en Lingüística Teòrica i Aplicada, Barcelona, Spain 2015.

[5] Lee, H., & Narang, V. (2010). A Study in Comparing Acoustic Space: Korean and Hindi Vowels. 2010 International Conference on Asian Language Processing. doi:10.1109/ialp.2010.86.

[6] Tsao, Y.-C., & Iqbal, K. (2005). Can Acoustic Vowel Space Predict the Habitual Speech Rate of the Speaker? 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. doi:10.1109/iembs.2005.1616644.

[7] Vaishna Narang, Deepshikha Misra, "Acoustic Space, Duration and Formant Patterns in vowels of Bangkok Thai", Centre for Linguistics, Jawaharlal Nehru University, New Delhi, International Journal on Asian Language Processing 20 (3):123-140.

[8] Bishnu S Atal, "The history of linear prediction", University of Washington Seattle , IEEE Signal Processing Magazine 23(2):154-161, DOI:10.1109/MSP.2006.1598091, April 2006.

[9] O'Shaughnessy, "Linear predictive coding". IEEE Potentials, 7(1), 29–32. doi:10.1109/45.1890.

[10] Ethnicity Group, "Cepstrum Method". 1998 http://www.owlnet.rice.edu/~elec532/PROJECTS98/speech/cepstrum/ cepstrum.html

[11] Thiang, Suryo Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", Electrical Engineering Department, Petra Christian University Jalan Siwalankerto 121-131, Surabaya 60236, Indonesia, 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore.