Formant Analysis of Malayalam Phonemes for Cleft lip

Dr. Nobert Thomas Pallath

Associate Professor
Electronics Department
WMO Arts and Science College, Wayanad, Kerala, India

***Abstract*— Speech is the most efficient and widely used form of human communication and it is made up of a series of phonemes. Malayalam is part of the Dravidian language family's Southern branch. Despite its close relationship with Tamil, Sanskrit had a greater influence on Malayalam language than Tamil. This study is based on Malayalam, which is one of the 22 official languages and 14 regional languages in India. Speech deficiency is a condition in which a person's voice or ability to create sound is impaired. It can be tremendously disheartening when a speaker understands exactly what to say but is unable to speak properly due to a speech defect. This study compares the cleft-lip voice of humans in the Malayalam language. Formant frequency is important in speech and speaker recognition, this paper focuses on the formant frequency characteristics of speech signal. In this paper, the LPC model is used for the estimation of the first three formants F1, F2, and F3 in Malayalam phonemes.**

***Index Terms*— Speech, formants, LPC, cleft lip.**

I. INTRODUCTION

As different from other species human can interact with their companions without knowing what to think and talk about. Sending and receiving verbal and nonverbal messages between two or more people is referred to as human communication. A language is a human-created, organised system for communication. Malayalam (IPA: mælə'jɑːləm) is a classical Indian language that serves as the official language of Kerala and Lakshadweep Islands. It is considered challenging when compared with South Indian languages such as Tamil, Telugu, and Kannada, which are all very easy to learn. The majority of Malayalam speakers live in Kerala and the Union Territory of Lakshadweep.

II. HUMAN SPEECH PRODUCTION

Speech Production Mechanism

The natural way for humans to communicate is through speech, which is made up of a series of sounds. Speech is made up of four processes. Fig. 1 shows the block diagram of the human speech production mechanism. The content of an utterance is converted to phonemic symbols in the language center of the brain during language processing. Motor commands to the vocal organs are generated in the motor center of the brain. Based on these motor commands, the vocal organs produce articulatory movement for speech generation and speech is the result of air being ejected from the lungs, touching the vocal cord, throat, tongue, cheek, palate, teeth, and lips [1].

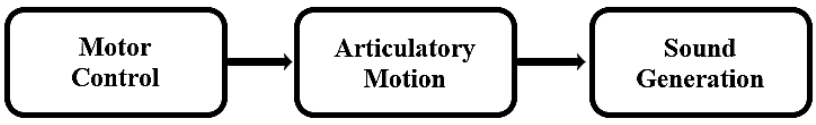

Fig. 1. Block diagram of human speech production mechanism

From the outside, speech production appears to be fairly easy, yet the process behind it is extremely complicated. Humans can produce a wide range of sounds with quickly changing frequency spectrums and volume levels. This is because of very sharp and clear articulatory movement control of the organs [2].

Physiology of Speech Production

Speech organs are the organs that participate in the production of speech sounds. To make sound, the air is released from the lungs and it traverses through vocal cavity. Without the presence of air, no sound can be produced. Fig. 2 shows the anatomy of speech production. In phonation, the tongue plays a significant role. Speech production is impossible without it. The teeth serve as an important articulation point for the tongue and lips. To make sound, they collaborate seamlessly and quickly. We can't speak properly if we don't have teeth. The function of lips is to open and close in accordance with speech. One of the main components of sound production is vocal folds, also known as vocal cords. A voiced or voiceless sound is determined by the vibration of these vocal folds.

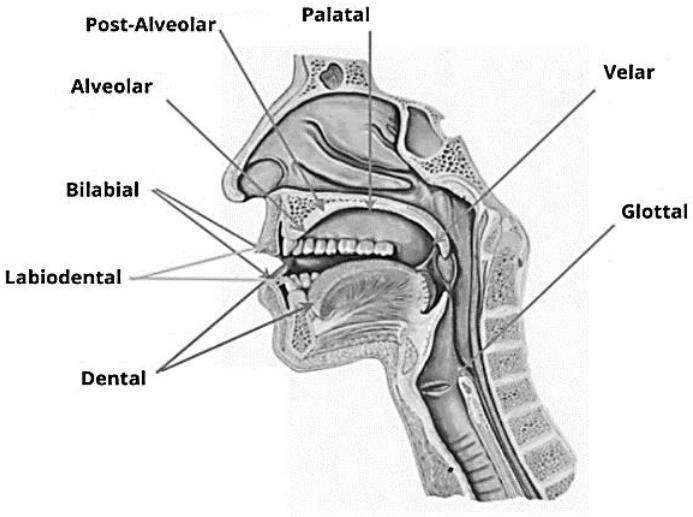

Fig. 2. Anatomy of Speech

The lungs are the source of energy for speech production. It would be impossible to produce sound without the lungs. This is due to their internal pressure system. The vocal folds cannot move or vibrate to produce sound if there is no pressure on them [3].

III. Cleft Lip and Palate

Speech problems, commonly referred to as speech impairments, occur when a person's regular speech is disturbed [5]. Cleft lip and palate are a type of disorder in which the upper lip and roof of the mouth are pruned or split from birth. This happens when the facial features do not fully develop during pregnancy.

Cleft Lip

Failure of the frontonasal and maxillary processes to fuse, results in cleft in the lip, alveolus and nasal floor in variable sizes [4].

The cleft lip can be complete or notched, and it can also involve the cleft alveolus. Because of the wound stress, the severity of the cleft lip can make the healing more challenging. The treatment of more severe cleft lips frequently necessitates a longer preoperative preparation period.

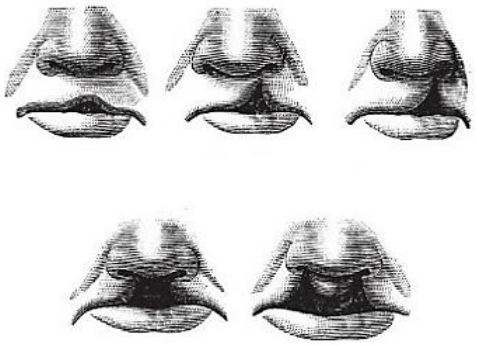

Fig. 3. Cleft Lip

Cleft Palate

Both the primary and secondary palates might be affected by a cleft malformation. Clefts in the primary palate can range in size from an alveolar notch to a cleft that runs through both the hard and soft palates [5]. There are three types of cleft palate. A soft palate cleft in the back of the mouth is referred to as an incomplete cleft palate. This form begins in the soft palate at the back of the mouth and extends forward and may not always reach the front. This form affects both the hard and soft regions of the palate and affects in full length. The cavities of the mouth and nose are connected. A full cleft palate can occur on one side (unilateral) or both sides (bilateral) [6].

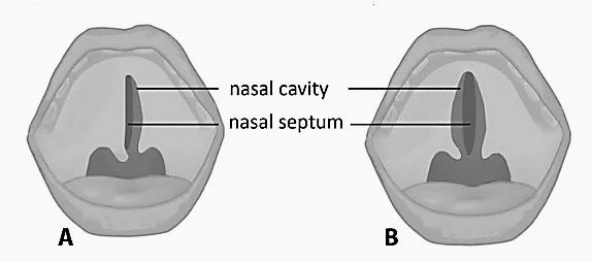

Fig. 4. (A) Unilateral cleft palate. (B) Bilateral cleft palate

IV. MALAYALAM PHONEMES

Speech and language are two distinct from each other. Speech is the way we express ourselves through sounds and words. Speech includes articulation, voice, and fluency. The way pronounce words with our mouth, lips, and tongue is called articulation. The way we make sounds with our vocal folds and breath is called voice. The rhythm of our speech is referred to as fluency. Language is essentially a set of traditional, spoken or written symbols that humans use to communicate with one another.

A phoneme /ˈfoʊniːm/ is a unit of sound in phonology and linguistics that distinguishes one word from another in a given language. In spoken language, phonemes are the essential building units. A phoneme can be classified into two types, vowels and consonants. Consonants can be voiced or unvoiced, whereas vowels are always voiced.

Malayalam is a Dravidian language that comes from the southern branch. Malayalam is the official language of Kerala, laying in the southern part of India. There are 51 letters in the character set, with 15 vowels and 36 consonants. Vowel phonemes are further classified into three categories based on the position of the tongue in the oral cavity. They are Front (ഇ i /i/, ഈ ī /iː/, എ e /e/, ഏ ē /eː/), Central (അ a /a/, ആ ā /aː/), and Back (ഉ u /u/, ഊ ū /uː/, ഒ o /o/, ഓ ō /oː/). Malayalam has 2 diphthongs (ഐ ai /ai̯/, ഔ au /au̯/). A diphthong, often called a gliding vowel, is formed when the tongue, lips, and jaw move from one pure vowel sound to another. Consonants can be classified into Nasal, Plosives, Fricative, and Affricate. Nasal sounds such as മ /ma/, ങ /ŋa/, ഩ /ɲa/, ൺ /ɳa/, ൻ /na/are produced when air resonates and leaves through the nasal cavity as the velum is lowered. When the airflow is paused at a certain point and then released out of the mouth, plosive sounds are created. Some examples are ക /ka/, ഖ /kʰa/, ഗ /ga/, ഘ /gʱa/, ച /t͡ʃa/, ഛ /t͡ʃʰa/, ജ /ɟa/. The vocal tract is constricted in fricative sounds to create a turbulent airflow (air hisses or buzzes). Examples are സ /sa/, ഹ /ɦa/. The dual action of plosives followed by a fricative produces affricate sounds.

V. FORMANT FREQUENCY AND LPC

Formants are vocal tract resonances in speech processing. Many approaches depend on the estimation of their locations and bandwidths (especially during the generation of spoken speech). Formant frequencies can be obtained using a variety of approaches. A spectrogram can be used to calculate it from the frequency spectrum of the sound [7].

The first four resonant frequencies can roughly describe the primary resonances of the vocal tract. The first (F1), second (F2), third (F3), and fourth (F4) formants are the resonant frequencies. The fundamental frequency F0 and the formant frequencies are correlated. The relation between the nth formant frequency Fn and the fundamental frequency F0 can be approximated as [8]:

$$F_n = a_n(F_0 + b_n) \quad (1)$$

where an and bn are vowel dependent constants

LPC method is used to obtain the formant frequencies by finding the roots of the prediction polynomial [9].

Signals generated by linear filtering mechanism that changes slowly are best for LPC, especially if the filter is triggered by rare and brief pulses.

The vocal tract parameters (LP coefficients) and glottal excitation are decomposed into two highly independent components in LP analysis (LP residual). For unvoiced speech segments, a linear time-varying filter (the vocal tract) is excited by random noise, while for voiced speech segments, a train of pulses is used.

A linear predictor, which produces values based on a linear combination of prior signal values, may accurately forecast the future values of similar signals. The Fourier transform can also be used to represent a signal. A Fourier transform, often known as a frequency representation, can be used to highlight key characteristics of a signal [10].

Figure 5 shows a model of speech production for LP analysis. It is made up of an H(z) time-varying filter that is activated by either a quasi-periodic or random noise source. The speech sample S(n) is represented as a linear mixture of previous outputs, current input and previous inputs. It can be expressed mathematically as S(n).

$$S(n) = -\sum_{k=1}^{p} a_k s(n-k) + G\sum_{l=0}^{q} b_l u(n-1), \qquad b_0 = 1 \quad (2)$$

where ak,1 ≤ k ≤ p, bl ,1 ≤l≤ q and gain G are the filter's parameters. Alternatively, the linear prediction speech model's transfer function in frequency domain is

$$H(z) = \frac{1+\sum_{l=1}^{q} b_l z^{-1}}{1+\sum_{k=1}^{p} a_k\ z^{-k}} \quad (3)$$

The model H(z) is known as a pole-zero model. The zeros indicate the nasals, while the poles represent the vocal-tract resonances (formants),

where a_k=0 for 1≤k≤p

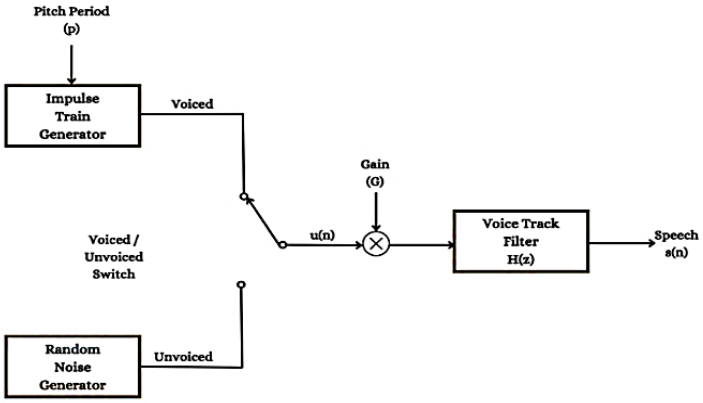

Fig. 5. Model of speech production for LP analysis

Because of its ability to generate accurate estimates and relative speed of computation, linear predictive analysis is the most used technique to extract formant frequencies. The basic steps of the LPC process include the following [11-13]:

1. Pre-emphasis: The digitised language (voice) signal s(n), is passed through a low-order digital system to spectrally flatten it and make it less susceptible to finite precision effects later in the signal processing. The following equation connects the output of the pre-emphasiser network to the network's input s(n):

$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1) \quad (4)$$

2. Framing and Blocking: The previous step's output is blocked into N-sample frames, with M-sample separation between nearby frames. If the lth frame of a language is xl(n), and there are L frames in the total language signal, then x1(n) is equal to

$$x_l(n) = \tilde{s}(Ml + n) \quad (5)$$

where n=0, 1……. N-1; l=0,1,….L-1

3. Windowing: Each frame is windowed in this stage to decrease signal discontinuities at the start and end of each frame. The resultant signal is obtained by windowing if the window is defined as w(n), in which $0 \le n \le N-1$.

$$\tilde{x_l}(n) = x_l(n)w(n) \quad (6)$$

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

The Hamming window is a common type of window that comes in the form.

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right] \quad (7)$$

in which $0 \le n \le N-1$

4. Autocorrelation Analysis: The next stage is to auto correlate every frame of the windowed signal in order to provide

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x_l}(n)\,\tilde{x_l}(n+m); \quad m = 0,1,\ldots p \quad (8)$$

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

5. Linear predictive coding (LPC) Analysis: It is the next processing step that translates every frame of p+1 autocorrelations into linear predictive coding (LPC) parameter set by using Durbin's method. The algorithm is as follows:

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

$$k_i = \frac{r(i) = \sum_{j=1}^{i-1} \propto_j^{i-1} r(|i-j|)}{E^{i-1}} \qquad 1 \le i \le p \quad (9)$$

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

$$\propto_j^{(i)} = \propto_j^{(i-1)} - k_i \propto_{i-j}^{(i-1)} \qquad 1 \le j \le i-1 \quad (10)$$

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

The LPC coefficient am, is calculated by recursively solving E(0) to E(i) equations for i = 1,2,...p.

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

6. Linear predictive coding (LPC) parameter conversion to cepstral coefficients: LPC cepstral coefficients, which may be calculated directly from the LPC coefficient set, are significant LPC parameter. The recursion utilised is as follows:

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right). c_k. a_{m-k} \; ; \quad 1 \le m \le p \qquad (11)$$

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m}\right). c_k. a_{m-k} \; ; \quad m > p \qquad (12)$$

The features retrieved from voice signals are called LPC cepstral coefficients, and these coefficients are used as input data for Artificial Neural Networks. The estimation of resonance peaks from the filter coefficients acquired by LPC analysis of segments of the speech waveform is a commonly used technique for formant placement

VI. METHOD AND MATERIALS

Voice samples collected from healthy public are compared with cleft-lip candidates who had undergone surgery in childhood. The study includes 20 healthy adults and 15 people with cleft lips who speak Malayalam language natively.

The speech samples were recorded in a quiet environment using a high-quality microphone. For phonemic character investigations, test words were created to place nasal vowels and consonants in various phonologic conditions. Malayalam Vowels (Swaraksharam) and a few short words are among the test letters and words used. When pronouncing most of the words, the lips are more involved and therefore these words are used for evaluation.

The samples are recorded from healthy as well as cleft-lip subjects (with 16Khz sampling frequency). Table 1 lists the vowels, while table 2 lists the consonants obtained from recorded samples.

Table 1. Malayalam vowel phonemes (Swaraksharam)

Short Vowel	**IPA**	**Long Vowel**	**IPA**
അ	a	ാ, ആ	a:
ി, ഇ	i	ീ, ഈ	i:
ു, ഉ	u	ൂ, ഊ	u:
െ, എ	e	േ, ഏ	e:
ൊ, ഒ	o	ോ, ഓ	o:
Vocalic Consonant	***IPA***	***Anusvaram***	***IPA***
ൃ, ഋ	ṛ	അം	aṁ
Diphthongs	***IPA***	***Visargam***	***IPA***
ൈ, ഐ	ai	അഃ	aḥ
ൗ, ഔ	au		

The documented Malayalam vowels (swaraksharam) are categorised into Short Vowels, Long Vowels, Diphthongs, Vocalic Consonant, Ansawara, and Visarga [14,15] as seen in table 1. Table 2 shows some selected consonants which have the higher role of lip and nasal components in the pronunciation of the letter.

The recorded samples are organised for analysis by segmenting unwanted signals. The noise is removed in the pre-processing step and each word is broken down into its consonant syllables. The microphone is kept at a distance of 10cms away from the speaker for recording and the test words were recorded a multiple number of times so that the best one is selected. The analysis is performed by using LPC method using MATLAB to give the values of formant frequencies.

Table 2. Malayalam consonant syllables

Syllables		**IPA**
Dental	ത	t̪a

	ല	la
Labial	മ	ma
	പ	pa
	വ	va
Retroflex	ള	ɭa
	ഴ	ɻa
	ര	ɾa

VII. RESULTS

In this paper, Tables 3 and 4 show the values of normal and cleft lip speech for Malayalam vowels and consonants respectively.

Table 3. Formant frequencies of Malayalam vowels

Vowels	**F1**		**F2**		**F3**	
	Normal	***Cleft***	***Normal***	***Cleft***	***Normal***	***Cleft***
അ a	826	429	1326	1558	3464	3895
ആ ā	789	409	1293	1825	3617	4783
ഇ i	307	229	2993	4467	4403	5497
ഈ ī	301	231	3087	4695	4376	6323
ഉ u	364	234	707	2417	3935	5102
ഊ ū	381	234	698	5347	3749	6537
ഋ ṛ	407	223	1920	1809	3696	4647
എ e	456	258	2881	2952	4444	4676
ഏ ē	422	268	2883	2847	4351	4913
ഐ ai	409	246	2632	2277	4368	4812
ഒ o	494	263	840	1516	3940	4627
ഓ ō	473	251	846	1490	3733	5210
ഔ au	448	240	826	1559	3895	4702
അം aṁ	366	227	1142	1402	3638	4574
അഃ aḥ	780	407	1503	1945	3777	4889

Table 4. Formant frequencies of Malayalam consonants

	F1		**F2**		**F3**	
	Normal	***Cleft***	***Normal***	***Cleft***	***Normal***	***Cleft***
ത ţa	753	242	1603	2234	2928	4777
ല la	818	276	1488	2109	2954	4859
പ pa	837	211	1711	2179	4083	4961
വ va	787	180	1538	1163	3324	2515

മ ma	574	235	1370	1364	2895	5727
ര ra	769	266	1700	2096	3235	4746
ള ḷa	826	256	1361	1848	3090	2398
ഴ ɻa	765	272	1446	2231	2906	6454

The severity of a cleft lip is evaluated by comparing the frequencies of normal and cleft speech. From the results, it is found that the first formant frequency for cleft lip F1 is lower while F2 and F3 values are higher than that of normal values. Changes in F3 are more noticeable in cleft lip speech than changes in F1 and F2. Some utterances like ഛ and ള seem to be different in F3 variation with cleft lip. For candidates with cleft lip, substitution is a common articulatory mistake.

The three plots shown in figures 6, 7 and 8 are the graphs for F1, F2 and F3 frequencies versus different Malayalam vowels respectively. Each plot shows the difference between normal speech and cleft lip.

F1 values for normal speech and cleft lip are represented graphically in figure 6. For all Malayalam phonemes, the initial formant (F1) for cleft speech is lower than that of normal speech. The difference is especially noticeable for the phonemes അ (a), ആ (ā) and അഃ (aḥ) due to high vowel height. The graphical representation of F2 versus different Malayalam vowels is shown in figure 7, The frequency F2 for cleft lip patients is higher than that of normal speech because of the more frontal positioning of the tongue, however, the difference is rather slight in some letters. In figure 8 the third formant seems to be higher for the cleft lip subject and the difference is greater. In some utterances, the F3 difference is greater and here the lip has a greater involvement on how that letter is articulated. F3 difference is higher for long vowels than short vowels.

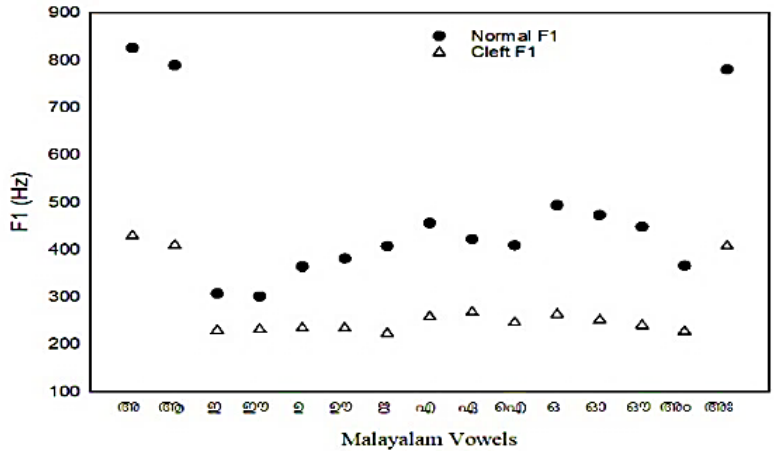

Fig. 6. F1 values for normal and cleft lip

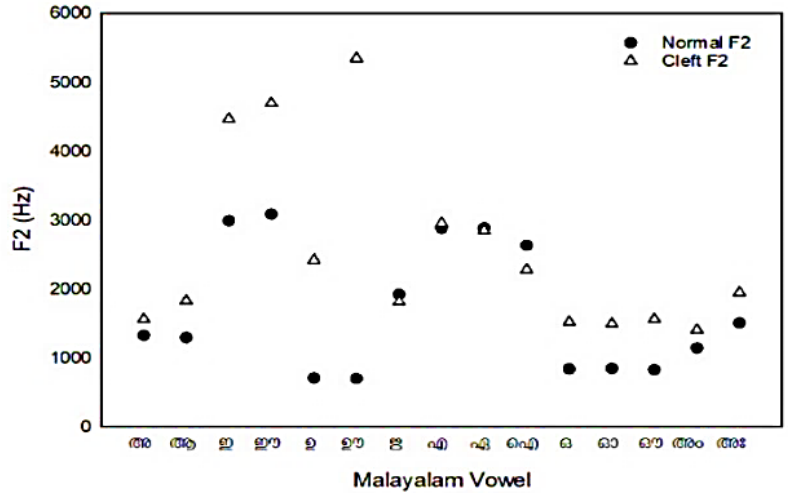

Fig. 7. F2 values for normal and cleft lip

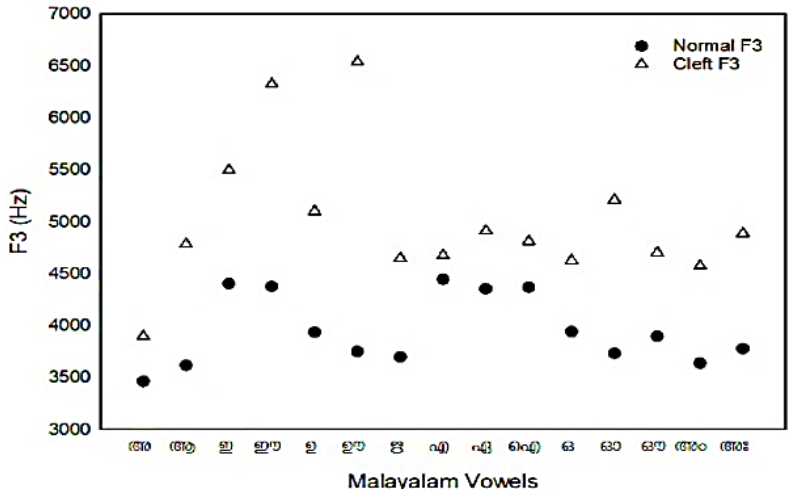

Fig. 8. F3 values for normal and cleft lip

VIII. Conclusion

This study shows a strong relationship between formant frequencies obtained from different phonemes. Here the Malayalam vowel phoneme sound produced by cleft lip patients whose cleft lip surgery where already performed have been studied. In this study the first formant F1 is found to be lower than that of normal speech, while F2 and F3 were higher. F2 and F3 formant frequencies play a key role in estimating the severity of cleft lip. The pronunciation difficulty will vary in accordance with F3 variations, i.e., the articulation difficulty will be higher for higher F3 values. For utterances that rely more on lip and nasal, F3 will be greater. As a result, fluctuations in F3 could be used to track the severity of cleft lip and indicate the requirement for more speech training. In the case of consonants, F3 varies depending on how the articulators are oriented at the time of pronunciation. Because of articulation difficulties, several consonants were misarticulated.

IX. Acknowledgment

The author would like to acknowledge the authors of various research papers used in this work. The author would also like to acknowledge the developers of MATLAB. The author would like to thank all departments of WMO Arts and Science College as well as WMO Deaf and Dumb School, Muttil, Wayanad, Kerala and various Block Resources Centers (BRC) of Wayanad, Kannur and Kasaragod districts of Kerala.

References

[1] Dr V.Geethakumary and Dr. B.A.Sharada, “Malayalam Speech Sounds and Their Mapping to Unicode Symbols: A Case Study,” Central Institute of Indian Languages, Manasagangotri, Mysore, Proceedings of the Language Engineering Conference (LEC’02), IEEE, 2002.

[2] M. Honda, “Speech synthesis technology based on speech production mechanism, How to observe and mimic speech production by human,” NTT CS Laboratories, Journal of the Acoustical Society of Japan, Vol. 55, No. 11, 1999, pp. 777-782.

[3] Harish Chander Mahendru, “Quick Review of Human Speech Production Mechanism,” Department of Electronics & Communication Engg., Haryana Engineering College, Jagadhri, Haryana, International Journal of Engineering Research and Development, Volume 9, Issue 10, 2014, PP. 48-54.

[4] Liv Baker, “Anatomy and Physiology of the Speech Mechanism and Features of the Human Body,” Grand Valley State University, Fishladder: A Student Journal of Art and Writing, Volume 16, Article 32, 2018.

[5] Semer N, “Practical plastic surgery for non-surgeons,” Philadelphia: Hanley&Belfus, Inc; 2001. pp. 235–43.

[6] David Shaye, Travis T Tollefson, and C. Carrie Liu, “Cleft Lip and Palate: An Evidence-Based Review,” Article in Facial plastic surgery clinics of North America, 2015

[7] Deekshitha G, and Leena Mary, “Broad Phoneme Classification Using Signal Based Features,” In Advanced Digital Signal Processing Research Laboratory, Department of Electronics and Communication, Rajiv Gandhi Institute of Technology, Kottayam, Kerala, International Journal on Soft Computing (IJSC) Vol. 5, No. 3, 2014.

[8] Roy C. Snell, and Fausto Milinazzo, “Formant Location from LPC Analysis Data,” In IEEE Transactions on Speech and Audio Processing, Vol. 1, NO\o. 2, 1993.

[9] A.R. Jayan, “Speech and Audio Signal Processing,” Department of Electronics and Communication Engineering, Government Engineering College Sreekrishnapuram, Palakkad, Kerala, 2017.

[10] J. D. Markel and A. H. Gray, “Linear Prediction of Speech,” New York: Springer-Verlag, (Issss976), 1977.

[11] G Sawant, A Singh, K Kadam, R Mazarello, and P Dumane, “Speech Synthesis using LPC,” Don Bosco Institute of Technology, Mumbai, India, International Conference and Workshop on Emerging Trends in Technology

(ICWET 2010) – TCET, Mumbai, India, 2010.
[12] Thiang and Suryo Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot," In Electrical Engineering Department, Petra Christian University, Jalan Siwalankerto 121-131, Surabaya 60236, Indonesia, International Conference on Information and Electronics EngineeringIPCSIT vol.6, IACSIT Press, Singapore, 2011.
[13] Lawrence Rabiner, Biing Hwang Juang, "Fundamentals of Speech Recognition," Prentice Hall, New Jersey, 1993.
[14] Bhuvaneshwari Jolad, and Dr. Rajashri Khanai, "Different Feature Extraction Techniques for Automatic Speech Recognition: A Review," In International Journal of Engineering Sciences and Research Technology, 2018.
[15] Neo-Brahmi Generation Panel [NBGP], "Proposal for a Malayalam Script Root Zone Label Generation Ruleset (LGR)," version 4.0, 2020.