Final Report of Major Research Project

(From 5th February 2020 to 28th February 2023)

"Adaptive Speech Recognition for Speech Defects in Malayalam Language"

UGC STRIDE MAJOR PROJECT
(COMPONENT 1)

Submitted To

The Secretary
University Grants Commission
Bahadur Shah Zafar Marg
New Delhi-110002.

Submitted By

Dr. Nobert Thomas Pallath
Associate Professor (Electronics)
Project Co-Ordinator & Principal Investigator
UGC STRIDE Major Research Project
WMO Arts and Science College, Muttil
Wayanad, Kerala
Pin-673122
nobert.pallath@gmail.com

FINAL REPORT

File no: F.2-23/2019 (STRIDE-I), Date: 3rd December 2019

Major Research Project Under Scheme for Trans-Disciplinary Research for India's Developing Economy (STRIDE)

Entitled

Adaptive Speech Recognition for Speech Defects in Malayalam Language

Project Coordinator & Principal Investigator

Dr. Nobert Thomas Pallath
Associate Professor (Electronics)

Submitted To

UNIVERSITY GRANTS COMMISSION
Bahadur Shah Zafar Marg
New Delhi-110 002
Year - 2023

W.M.O. ARTS & SCIENCE COLLEGE
"education and some more"
Affiliated to University of Calicut
(Aided by Govt. of Kerala)
P.O. MUTTIL, WAYANAD, KERALA-673122

Phone : 04936-203382 (Office)
04396- 207532 (Principal)
e-mail : wmocollege@gmail.com
website : www.wmocollege.ac.in

Ref:

Date:..01/06/2023................

To,

The Joint Secretary
University Grants Commission
Bahadur Shah Zafar Marg.
New Delhi-110002

Ref: 1) UGC Letter No. F2-23/2019 (STRIDE -1) Dated 03-12-2019
2) UGC Letter No. F2-23/2019 (STRIDE-1) Dated 21-01-2020
3) UGC Letter No. F2-23/2019 (STRIDE-1) Dated 17-02-2022

Sub: Submission of Final Report

Sir/Madam,

The University Grants Commission had sanctioned Rs 2195000/- (Rupees of Twenty one lakh ninety five thousand only) towards financial assistance for implementing UGC- STRIDE PROJECT COMPONENT-1 (Research Capacity building and Human Resource Development) for a duration of 3 Years as per Ref.1 cited above.

As per the sanction order Ref. 2 cited above, the UGC had released an amount of Rs.1097500/- (Rupees of Ten lakh ninety-seven thousand five hundred only) as the first instalment and the fund was credited to the bank account on 05/02/2020.

The UGC had released an amount of Rs.1097500/- (Rupees of Ten lakh ninety-seven thousand five hundred only) as the subsequent instalment as per the sanction order Ref. 3 cited above. The fund was credited to the bank account in 6 instalments w.e.f. 03/03/2022 to 14/06/2022.

A mentoring and monitoring committee (MMC) including Prof. B.S Sherigara as U.G.C Nominee, Dr. M.P Pillai and Dr. P.S Sathidevi as members was constituted on 19-02- 2020. The first MMC meeting for the year 2019-2020 was held at WMO Arts & Science College, Muttil on 24/02/2020.

During the Covid period, the second annual virtual meeting of the MMC for the year 2020-2021 was held on 28/12/2020, for the year 2021-2022, an annual meeting of the MMC was held on 01/11/2022 and the final MMC meeting was held on 30/03/2023. The committee recommended to forward the final Project Report of the UGC STRIDE Institutional project and utilisation certificate.

Therefore, I am forwarding herewith the final report duly certified by the MMC along with the utilisation certificate and statement of expenditure for your consideration.

Thanking you.

Yours Sincerely

Dr. Nobert Thomas Pallath
Project Coordinator
UGC STRIDE Major Research Project
WMO Arts & Science College, Muttil
Wayanad -673122

WAYANAD MUSLIM ORPHANAGE ARTS & SCIENCE COLLEGE
Muttil P.O.
Wayanad Dist
Date.........

Certificate of Declaration

I, Dr. Nobert Thomas Pallath, hereby certify that the work presented in this report is the original work, conducted as part of the Major Research Project entitled "Adaptive Speech Recognition for Speech Defects in Malayalam Language" under the UGC.

Dr. Nobert Thomas Pallath

Principal Investigator and Project Coordinator

UGC STRIDE (Component 1) Research Project

Department of Electronics

WMO Arts & Science College, Muttil,

Wayanad, Kerala, India

Pincode-673122

Acknowledgement

I am always grateful to the ultimate truth of the universe, the almighty whose grace has been bestowed up on us. I take this opportunity to express my deep sense of gratitude to the University Grants Commission for funding the Institutional Major Research project entitled "Adaptive Speech Recognition for Speech Defects in Malayalam language", under the STRIDE Component – 1 scheme. I would like to express my gratitude to the Mentoring and Monitoring Committee (MMC), for their effective mentoring and monitoring. This endeavour could not have been possible without the generous support of WMO at large and also the Principal, administrative staff and various departments of WMO Arts & Science College, Muttil.

I am also grateful to Block Resource Centres (BRCs) of the various Block Panchayaths, Speech therapy centres and Speech & Hearing Institutes, for their valuable help in data collection. This work could not have been completed without the sincere efforts of the project staff.

Dr. Nobert Thomas Pallath
Associate Professor and Principal Investigator
Department of Electronics
UGC STRIDE (Component 1) Research Project
WMO Arts & Science College, Muttil,
Wayanad, Kerala, India
Pincode-673122

UGC Sanction Letter

डॉ. शकील अहमद
संयुक्त सचिव
Dr. Shakeel Ahmad
Joint Secretary

विश्वविद्यालय अनुदान आयोग
University Grants Commission
(मानव संसाधन विकास मंत्रालय, भारत सरकार)
(Ministry of Human Resource Development, Govt. of India)
बहादुर शाह जफ़र मार्ग, नई दिल्ली-110002
Bahadur Shah Zafar Marg, New Delhi-110002
दूरभाष Phone : 011-23232055
ई-मेल E-mail : shakeel.ugc@nic.in

No. F.2-23/2019 (STRIDE-I)

05 DEC

December, 2019

The Principal
WMO ARTS & SCIENCE College Muttil,
Muttil South, Kerala 673122

Subject: UGC Assistance for Component-1 (Research Capacity Building and Human Resource Development) under STRIDE - (Duration – 3 Years).

Sir/Madam,

1. This is with reference to the proposal submitted by WMO ARTS & SCIENCE College Muttilto support for Component-1 (Research Capacity Building and Human Resource Development) under Scheme for Trans-disciplinary Research for India's Developing Economy (STRIDE) of the University Grants Commission (UGC).

2. On the basis of the recommendations of the Expert Committee at its Interface Meeting held on 26th and 27th November, 2019, I am directed to convey approval of the UGC for the proposal submitted by the College under Component-1 (Research Capacity Building and Human Resource Development) under the STRIDE (duration of 3 years from the date of receipt of first instalment) with the following identified approaches:

   a) Curriculum
   b) Critical Thinking
   c) Faculty Resources
   d) Collaborations and Partnership
   e) Criteria of Selection
   f) Training Strategies
   g) Training Methods
   h) Innovations
   i) Implementation Process (time-frame)

The Coordinator of the Project shall be responsible for implementation of the Project as per the identified approaches under the overall supervision of the Principal. He/shemay continue till the end of the project or till his/her superannuation.

3. The financial assistance approved for implementing the present project for a duration of 3 years is given below : -

S. No.	Non-Recurring (Items)	Rs. (In Lakh)
1.	Equipments:	6.0
2.	Upgradation/augmentation extension of existing laboratory for housing and installation of new equipment including air-conditioning	-
	TOTAL	6.0
S. No.	Recurring(Items)	
1.	Core Assistance (Participant Cost and Hiring Services)	-
2.	Human Resources (Subject Experts) Cost	15.0
3.	Contingencies (Utilities, Stationary and Miscellaneous)	0.35
4.	Travel (Domestic and International)	0.60
	TOTAL	15.95
	Grand Total(NR+R)	21.95

Amount in words: Rupees Twenty One Lakh and Ninety Five Thousand only.

4. The College is requested to take immediate steps tosubmit the following documents for necessary action:

    i) Acceptance of the terms and conditions of the grants duly signed by the Principal of the College.
    ii) Name of the competent College Officer with full address and other bank details in (mandate form) the prescribed enclosed proforma so that the fund can be transferred electronically.
    iii) A Certificate from thePrincipal of the College that it is eligible to receive the UGC financial assistance.

5 The first installment of admissible grant will be released only on the receipt of requisite documents requested vide para4 by return of post.

6. The College is required to maintain a separate flexi saving bank account for the grants released for Component-1 (Research Capacity Building and Human Resource Development) under STRIDE. Interest earned against Grant-in-aid (other than reimbursement) released to any grantee institution should be mandatory remitted to the UGC account immediately after finalization of account. Any interest earned out of Grant-in-aid should not be allowed as additional funds over and above the allocation.

7. The College shall constitute immediately a Mentoring and Monitoring Committee (MMC) as per the STRIDE guidelines which can be downloaded from UGC's website ww.ugc.ac.in and follow the terms of reference of the Mentoring and Monitoring Committee (MMC) to ensure effective implementation and monitoring of the Project.

The task of purchasing of equipments and upgradation/augmentation/extensions of existing laboratory for housing and installation of new equipment including air-conditioning, if any, should be completed within a year from the date of receipt of the grant by the College. It may be noted that if it is not done within a year from the date of receipt of the grant by the College, the approval shall be considered as lapsed.

9. No request for any change in the effective date of the implementation of the Project will be considered.

10. The second/subsequentinstallment of grant for any approved items will be considered and sanctioned only on the receipt of the year-wise & item-wise Utilization Certificate along with annual progress report for the earlier installment in the prescribed form duly signed by the Registrar/Principal/Finance Officer as the case may be.

11. The annual progress report shall be submitted as per the Guidelines of STRIDE.

12. The College shall include all existing conditionsalso communicated from time to time by UGC in the Utilization Certificate, in respect of any financial assistance or grants-in-aid to any institution under any of the general or special schemes of the UGC.

13. The College shall take all possible measures to ensure effective implementation of policies and general financial rules of Government of India.

14. This approval is subject to the fulfillment of condition of inclusion of the College under section 2(f) and 12(B) of UGC Act, 1956 or any policy decision in this regard.

15. The College shall strictly follow the STRIDE Guidelines posted on UGC website.

16. The Project will be withdrawn at any point of time if it is found that the institution is ineligible or grant is not utilized for the purpose which it has been sanctioned or UGC rules have not been followed for utilizing the grant.

Yours faithfully,

(Dr. Shakeel Ahmad)
Joint Secretary (STRIDE)

Copy to:

1. The Coordinator of the Project,WMO ARTS & SCIENCE College Muttil, Muttil South, Kerala 673122
2. The Registrar, Calicut University, Mallapuram, Kerala.
3. The Secretary (Higher Education), Thiruvananthapuram,Kerala.
4. The Joint Secretary (Regional Office), Bengaluru - 560009, Karnataka.
5. Guard File.

(Nirmal Kaur)
Under Secretary

UNIVERSITY GRANTS COMMISSION

Utilisation Certificate

It is certified that the amount of Rs. 21,95,000/- (Rupees Twenty-one lakh ninety-five thousand only) out of the grant of Rs. 21,95,000/- (Rupees Twenty-one lakh ninety-five thousand only) received from the University Grants Commission under the scheme of support for STRIDE Project entitled "Adaptive Speech Recognition for Speech Defects in Malayalam Language" vide UGC letter No F.2-23/2019 (STRIDE-1) dated 21 January 2020 has been completely utilised for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission. In addition to the UGC STRIDE funding, an amount of Rs 56,494/- is incurred for the expenditure of the UGC STRIDE project activities. For the above overhead expenditure amount, partial amount of Rs 54,491/- is withdrawn from the interest rate credited for the aforementioned UGC STRIDE grant and the balance amount Rs 2003/- is diverted from the college account.

Signature of the
Project Coordinator

Signature of the
Principal

Principal
W.M.O. Arts & Science College
Muttil P.O., Wayanad-673122

For M/s. FAISAL & CO.
Chartered Accountants
Firm Reg. No.021925S

Faisal T.M
Proprietor
Membership No:236484

Statutory Auditor
(Govt. Internal Auditor / Chartered Accountant)

FAISAL & CO. KALPETTA WAYANAD CHARTERED ACCOUNTANTS

UDIN:- 23236484BGTMOK9394

UNIVERSITY GRANTS COMMISSION

STATEMENT OF EXPENDITURE IN RESPECT OF THE STRIDE PROJECT

1. Coordinator of the project : Dr. Nobert Thomas Pallath
2. University/ College : WMO Arts & Science College, Muttil
3. UGC approval Letter No. and Date : F.2/23/2019 Dated 03-12-2019
4. a. Period of Expenditure : From 05-02-2020 to 28-02-2023

   b. Details of Expenditure

Sl. No	Items	Amount Sanctioned	Amount Approved	Amount Received			Expenditure Incurred	Balance
				First Instalment	Second Instalment	Total Amount Received		
A.	**Non-Recurring**							
1	Equipment	6,00,000/-	6,00,000/-	3,00,000/-	3,00,000/-	6,00,000/-	6,00,000/-	0
B.	**Recurring**							
1	Human Resources	15,00,000/-	15,00,000/-	7,50,000/-	7,50,000/-	15,00,000/-	15,00,000/-	0
2	Contingencies	35,000/-	35,000/-	17,500/-	17,500/-	35,000/-	35,000/-	0
3	Travel	60,000/-	60,000/-	30,000/-	30,000/-	60,000/-	60,000/-	0
Grant total		**21,95,000/-**	**21,95,000/-**	**10,97,500/-**	**10,97,500/-**	**21,95,000/-**	**21,95,000/-**	**0**

Overhead Expenses met from the interest of the UGC grant ensured during the period from February 2020 to February 2023. The details listed below

Item	Received	Expense	Balance
Interest	54,491/-	00.0/-	54,491/-
Human Resource		48744/-	5747/-
Contingencies		5600/-	147/-
Equipment		150/-	-3/-
Travel		2000/-	– 2003/-

Certified that the grant has been utilised for the purpose for which it was sanctioned and in accordance with the terms and conditions attached to the grant. The overspending amount of Rs 56,494/- was utilised from interest rate and the college account.

For M/s. FAISAL & CO.
Chartered Accountants
Firm Reg. No.021925S

FAISAL & CO. KALPETTA WAYANAD CHARTERED ACCOUNTANTS

Faisal T.M
Proprietor
Membership No:236484

Signature of the Project Coordinator

Statutory Auditor (Govt. Internal Auditor / Chartered Accountant)

UDIN:- 23236484BGTMOK9394

Signature of the Principal

Principal
W.M.O. Arts & Science College
Muttil P.O., Wayanad-673122

UNIVERSITY GRANTS COMMISSION

BAHADUR SHAH ZAFAR MARG

NEW DELHI – 110002

Final Report of the work done on STRIDE Component -1 Research Project

1.	Project Report No.	: Final
2.	UGC Reference No.	: UGC Letter No. F.2-23/2019 (STRIDE-1)
3.	Period of Report	: 11th January 2020 to 28th February 2023
4.	Title of the Research Project	: Adaptive Speech Recognition for Speech Defects in Malayalam Language
5.	(a) Name of the Project Coordinator	: Dr. Nobert Thomas Pallath
	(b) Department	: Electronics
	(c) Institution	: WMO Arts & Science College, Muttil, Kalpetta, Wayanad, 673122 Calicut University
6.	Effective date of starting of the Project	: 11th January 2020
7.	Grant approved and expenditure incurred during the period of the report:	
	(a) Total Amount Approved	: Rs. 21,95,000/-
	(b) Total amount Received	: Rs. 21,95,000/-
	(c) Total Expenditure	: Rs. 21,95,000/-
	(d) Total Interest	: Rs 54,491/-
	(e) Over head Expenses	: Rs 56,494/-
	(f) Overhead Expense Utilised From	: Interest of UGC Grant and College Account
	(g) Report of the work done	: (Separate Sheet Attached)

Signature of Coordinator

WAYANAD MUSLIM ORPHANAGE ARTS & SCIENCE COLLEGE MUTTIL P.O. WAYANAD DIST. Date:..................

Signature of Principal

Principal
W.M.O. Arts & Science College
Muttil P.O., Wayanad-673122

Report of the work done

i. Objective

A speech disorder is a situation in which a person is aware of what should be spoken but is unable to articulate thoughts. Speech impairment is common among both children and adults. Some of the speech disorders can be corrected by earlier treatment. Some speech disorders are found to be hereditary and are developed over a period of time. Vocal cord damage, brain damage, stroke, respiratory weakness etc. are considered to be some of the causes of speech impairments. Normal speech is usually found to be effortless, but it is actually a complex process that needs precise timing, nerve and muscle control. Language impairments are different from speech impairments. Language impairment is an impairment in understanding and sharing thoughts. Sometimes, an individual may have both speech and language impairments. Repeating sounds which is most often seen in people who stutter, adding extra sounds, elongating words, making jerky movements while talking usually involving the head, blinking several times while talking, visible frustration when trying to communicate, taking frequent pauses when talking, distorting sounds when taking, hoarseness, raspy or gravely sounding voice etc. are considered to be some of the common symptoms of speech impairments.

The proposed study concentrates on speech impairments among children and the modelling of speech defects. Some physiological disorders like orofacial, myofunctional disorders, neurological disorders like aphasia dysarthria and speech sound disorder and functional disorders like preschool language disorder, language-based learning disabilities and selective mutism are the different type of speech desorders . Articulation problems evolve when the patient produces sounds. syllables, or words incorrectly so that the listener does not understand what is being said. The attention of the listener is paid more to the way. the words sound than to what they mean. There are many, types of sound errors but most of the mistakes fall into one of three categories i.e., omissions, substitutions or distortions.

Articulation problems may be a result of a variety of physical handicaps such as cleft lip & palate, cerebral palsy, and hearing loss or they may be related to other problems appearing in the mouth such as dental problems. Stuttering (alalia syllabaries) also known as stammering is a speech defect in which the flow of speech is infringed by involuntary repetitions and extensions of sounds, syllables, words or phrases. Also, a silent pause in which the person suffering from stuttering is unable to produce sounds. Aphasia is a language disorder resulting from brain damage. It affects the parts of the brain responsible for language. The main aim of this project is to develop a system to identify defective speech and to interpret the speech correctly using comparison methods. Analysis of the speech signal is done using features obtained from the extraction of the speech signal and detection can be carried out using decoding of the results obtained from the features.

A. Introduction:

The ability to communicate is crucial to human existence. Humans are capable of communicating with one another through a variety of means, including conversation, gestures, writing, painting and sign language. One of the most natural forms of communication is verbal exchange. Human voice conveys information on two levels that is, it communicates ideas through words on a basic level or on a deeper level, which reveals the speaker's identity, details about their language, emotions, gender and age.

From the physical movement of speech organs such as lips, teeth, tongue, palate, uvula, nasal and oral cavities, and vocal cords, the speech production process has many levels. The mechanism of speech is composed of four processes: language processing, in which the language center of the brain transforms the content of an utterance into phonemic symbols; the brain's motor center produces motor commands for the vocal cords; articulatory movement for the purpose of producing speech via the vocal folds in response to these motor instructions; and the emission of air sent from the lungs in the form of speech.

Speech recognition, is also referred as automatic speech recognition (ASR), computer speech recognition, or speech-to-text. It is a feature that allows a program to convert spoken language into written text. Speech recognition, which is frequently confused with voice recognition, focuses on converting speech from a verbal to a text format, whereas voice recognition merely aims to distinguish the voice of a certain user. There are several speech recognition applications and devices, but the more sophisticated ones make use of AI and machine learning. To interpret and process human speech, they combine the grammar, syntax, structure, composition of audio and voice signals.

The aim of speech is to convey a message in a linguistic form where it consists of articulation, voice and fluency pattern. The problem with human speech is the huge amount of variation that happens while pronouncing a word. This variation occurs based on stress, speech defect, environment, recording equipment, etc. Sometimes speech become unintelligible in some people due to some disorders or speech defect. Interruptions or breaks in normal flow of speech are termed as dysfluencies. It is a type of communication disorder where the normal speech is disrupted. Speech and language disorder is defined as a communication disorder that adversely affects the ability to talk, understand, read and write.

The rhythm of 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. Also, Malayalam is the official language of Kerala, which is one among the southern states of India. There are 51 alphabets 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.

B. Human Speech Production:

Speech, gestures, writing, doodles, facial expressions, body language and sign language are just a few of the ways that people can communicate with one another. Whereas speech is the natural way of communication in Homosapiens. Speech comprises of detailed pathological, acoustic and linguistic background descriptions that include information that is not only verbal but also includes emotional state and intention. The integration of cognitive, linguistic, and motor processes is necessary for fluent speech production. Communication between humans typically takes place through speech. Speech signal is generated from the lips when air is expelled from the lungs and the resulting flow of air is disturbed by a constriction in the vocal tract.

For the purpose of communication, both humans and animals create meaningful sound. The former can produce the sound needed to develop language, that they use to communicate with one another. Although animals can make noises that resemble vowels, they are unable to speak, making it impossible for them to develop their own language.

The ability to articulate sound differentiates humans from other species. The ability to communicate and convey thoughts through speech is a characteristic that is only found in humans. Humans have a well-developed system for the speech production. A speaker with an intention to convey thoughts formulates a sequence of sounds according to the rules of a language. There after the individual's vocal organs, also called as articulators, are controlled through a series of neuromuscular actions which produces an acoustic signal. The acoustic signal propagates as a longitudinal wave travelling with a velocity of 344 m/s through air as pressure variations. At the listening end, frequency component in the range of 20 Hz to 20 kHz produces vibrations in the basilar membrane located in the inner ear and undergoes a transduction to form a neural signal. It is decoded by the listener's brain using the appropriate language code and the message is comprehended.

A string of events called 'communication', enables both the speaker and the listener to convey thoughts and feelings. Voices of individuals are complex in nature and it generates multidimensional acoustic waves that contain information about the speaker's identity, the language being spoken, the speaker's physical and mental health as well as the person's race, age, sex, education level and religious background. Experts should have a thorough understanding of the mechanisms involved in speech generation before beginning any language analysis because they are quite complex. The organs of speech are typically referred to as the components of the human body that directly contribute to the creation of speech. Respiratory, phonatory and articulatory organs are the three primary speech organs.

Respiratory system consists of lungs, muscle of chest and trachea. Respiration, which is responsible for the movement of air, is the most important function of the lungs that is significant to speech production. Lungs act as the source of energy for human speech production. Lungs are controlled by a set of muscles which make them expand and contract alternately, so that the air from outside is drawn in and pushed out alternatively. When the air is pushed out of the lungs it passes through the windpipe or trachea, which has at its top the larynx.

The vocal fold's primary physiological role is to protect the body by closing the airways when eating or drinking so that, no solid food or liquid particles can reach the lungs through the windpipe. Normally the pharyngeal and the oral cavity are grouped into one unit called the oral tract. The nasal cavity is normally called the nasal tract. A timed exhale from the lungs releases air while speaking. Larynx is a system of cartilages, muscles, and ligaments which controls the air flow. Vocal cord or vocal folds are masses of flesh, ligaments and muscles, stretched between the front and back of larynx. Vocal cord has an average length of 1.3 cm for women and 1.5 cm for men. A slit-like orifice between the vocal folds is called glottis.

Vocal tract is a tubular passage that begins at the glottis and ends at the lips. The vocal tract is composed of muscles and bony tissues. It consists of the pharynx and the mouth or oral cavity. For an average male/female vocal tract has an average length of 14cm to 17cm. The vocal tract controls the spectral distribution of energy and helps in the generation of different sounds of a spoken language.

The nasal tract begins at the velum also known as soft palate and it ends at the nostrils or external openings of the nose. It acts as a valve that allows selective flow of air through the nose and/or through the mouth. On the lowering of velum, the nasal tract gets acoustically coupled to the vocal tract, producing nasalised sounds of speech. Structures in the vocal tract that move during the production of different sounds are called articulators. The movement of tongue, lips, jaw, etc. to make speech sounds is called articulation. Articulation can be characterised by its (i) manner and (ii) place. Manner of articulation refers to the way in which speech organs make a sound. It deals with the configuration and interaction of articulators during the production of a sound. Place of articulation refers to the position of obstruction in the vocal tract produced by speech organs, while producing different sounds. The study of production of different sound units by the interaction of physiological structures is called articulatory phonetics. Articulators can be active or passive depending on their mobility during speech production. Passive articulators remain static during speech production for example, upper lips, teeth, alveolar ridge, hard palate or roof of the mouth, soft palate, pharynx wall, etc. Active articulators are movable relative to passive articulators. The most important active articulators are tongue and the lower lips.

Here muscle force is used to press air from the lungs through the larynx, more specifically the epiglottis. The vocal cord vibrates and interrupts the air and produces a quasi-periodic pressure wave. The pressure impulse is called pitch impulse. The frequency of the pressure signal is the pitch frequency or fundamental frequency. The pressure impulse stimulates the air in the oral tract and for certain sounds the nasal tract also. When the cavities resonate, they radiate a sound wave which is the speech signal. Both tracts, that is vocal and nasal act as resonators with characteristic resonant frequencies called formant frequencies. It is possible to change the cavities by moving the jaw, tongue, velum, lips and mouth. Because of this it is possible to pronounce different sounds.

Speech sounds are produced during articulation process. The articulators, a group of speech organs, performs a sequence of action during the process. Active articulators are those which move as part of the articulation process. Passive articulators are the speech organs that remain relatively motionless. The points at which the articulators move towards or come into contact with certain other organs are the place of articulation. The type or the nature of movement made by the articulator is called the manner of articulation. The majority of the articulators are located on the bottom side or on the floor of the mouth, because they are connected to the lower, movable jaw. The points of articulation, or the majority of articulators are located on the roof of the mouth and connected to the fixed upper jaw. The alveolar ridge, the top lip, the upper teeth, the hard palate, the soft palate and the velum are the primary points of articulation. The upper lip and the upper teeth are easily identifiable parts in the mouth. The rough, uneven and raised area directly behind the upper teeth is known as the alveolar ridge. The hard palate is the hard bony structure with a membranous covering that follows the alveolar ridge. The hard palate is immediately followed by the soft palate or the velum. It resembles a soft muscular sheet that is connected to the hard palate at one end and ends in the uvula, a soft muscular protrusion that resembles a pendulum. The pharyngeal wall may also be regarded as a point of articulation.

Tongue and lower lip are the two most significant articulators. The tongue is the most adaptable articulator due to its mobility. The surface of the tongue is relatively large and the different points of the tongue are capable of moving towards different places or point of articulation. It may be conveniently divided into different parts, viz; front, center, blade, back and root of the tongue. When the tongue is at rest behind the lower teeth, the part of the tongue, which lies below to the hard palate towards the incisor teeth, is called the front of the tongue. The area where the front and back meet is referred to as the front, while the area that confronts the soft palate is referred to as the back. Some literature refers to the entire upper surface of the tongue, or the area below the hard and soft palates, as the dorsum.

The muscles in the throat that are connected to the vocal cords are stretched and relaxed while talking. But for speech to occur, the brain must coordinate the actions of lungs, larynx, vocal cords, tongue, lips, mouth and facial muscles. The shape and size of the tongue and lips, the position of the teeth, the tissue density, the elasticity and density of the vocal cords are all physical characteristics that affect the sounding of voice. However, due to individual variances in the vocal system, every person has a distinctive voice quality. Every person individually develops a distinct and individual method of learning to speak and is quite unlikely that any two people would have the exact same vocal tract form and size, vocal cavity or articulator control. This is where the uniqueness of voice rests. This demonstrates that each person has his own distinct voice.

The study of speech sound production and perception is known as phonetics. It is the study of speech sounds found in all languages and provides a physical description of how the sounds are produced. Phonation is the term used to describe how the larynx works and the various ways in which vocal fold vibrations are used to produce speech. It is a mechanism in which vocal cord vibrations cause air molecules to vibrate.

Three broad categories can be used to categorize phonetics. Acoustic Phonetics is the study of how speech sounds are physically produced and transmitted. Auditory Phonetics deals with the perception of spoken sounds. Articulatory Phonetics is the study of how speech sounds are produced.

Voiced sounds are produced by forcing air through the glottis with the vibration of vocal folds. This results in quasi-periodic air puffs that excite the vocal tract. Examples: /a/, /i/, /u/. Quasiperiodicity can be observed in the waveform of the speech signal for these sound units. Trying to utter sounds like /a/, /i/, and /u/ carefully, keeping fingers close to the glottis region of the neck, can clearly exhibit the feeling of vibration of the vocal folds during the production of these sounds. Unvoiced sounds are produced by forming a constriction in the vocal tract and forcing air through this constriction at a high velocity to produce turbulence. Examples: /sh/, /s/, /ph/. No periodicity is observed in the signal and the signal is similar to noise. While trying to utter /sh/ and /s/ sounds, keeping fingers close to the glottis region of the neck, it can be easily verified that the vocal fold vibration will be absent during the production of unvoiced sounds. The fundamental frequency can be used to detect the voiced and unvoiced parts of speech. Another way is to calculate the energy in the signal frame and there is more energy in a voiced sound than in an unvoiced sound.

Whispering is a sort of sound produced by holding the vocal folds together while maintaining an opening between the arytenoids. This modifies the characteristics of spoken sounds while having no effect on unvoiced sounds. Low frequency sounds called creaks are made when the glottis closes, but only a small portion of the vocal folds can vibrate, because of settings in the vocal cords. Falsetto sounds are comparable to voiced sounds where, the vibrations run through the entire length of the vocal folds, but they are able to reach distinctively high frequencies because, the vocal folds are thinned out to create an edge. Murmur is a combination of voiced sounds without full closure of the glottis. Its breathy characteristics is caused by the constant airflow allowed through the glottis.

C. Malayalam Language:

The proposed study concentrates on speech impairments among children and adults in Malayalam language. The human voice communicates ideas through words on a basic level, and on a deeper level, it reveals the speaker's identity and details about their language, emotions, gender, and age. 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 one of the 22 languages and 14 regional languages of India. Malayalam has 10 vowel phonemes. 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. A vowel is a sound produced by the voiced excitement of the open vocal tract. In speech and speaker recognition, the Formant frequency plays a significant role. The study first focused on the first four formants F1, F2, F3, and F4 which are analysed using the LPC model in Malayalam vowel phonemes. 40 normal adults (20 males and 20 females) in the age group of 18-45 years participated in this study.

Any language is composed of a number of elemental sounds, called phonemes that combine to produce the different words of that language. Phonemes are classified into vowels, glides, semivowels and consonants. Malayalam has 11 monothongs and 2 diphthongs and 52 consonant phonemes. A vowel is defined as a continuant sound (it can be produced in isolation without changing the position of articulators), voiced (using the glottis as a primary source of sounds), with no friction (noise) of air against the vocal tract passage. Malayalam consists of 36 consonants and 15 vowel symbols. Though there are 15 vowel symbolsin the Malayalam script, the number of vowel phonemes are only. A phoneme is the smallest unit of sound which makes a difference in meaning when used a language. A phoneme can be classified into two types, vowels and consonants (Lekshmi K.R et al, 2020). Sounds is classified as voiced or unvoiced. When produced a sound causes the vocal cords in the larynx to vibrate to produce a voiced sound and, if it is unvoiced sound, there is no vibration. In contrast, it can be found that although consonants may be voiced or unvoiced, all vowel sounds are voiced. It may be single (Monophthongs) or a combination (Diphthongs).

Vowels can be produced by controlling the airflow through mouth. It can be adjusted by the position of tongue, rounding of lips, and degree of opened mouth. The vowels can be classified based on the position of tongue in terms of how far it is forward and how high it is. Malayalam vowel consists of 5 short vowels (അ a /a/, ഇ i /i/, ഉ u /u/, എ e /e/, ഒ o /o/) and 5 long vowels (ആ ā /aː/, ഈ ī /iː/, ഊ ū /uː/, ഏ ē /eː/, ഓ ō /oː/). ഋ (/rĭ/, r̥) is a vocalic consonant, which is not officially a vowel. The vowel അം (aṁ /am/) is an Anusvaram and അഃ (aḥ /ah/) is a Visargam. Malayalam has 2 diphthongs (ഐ)ai /ai̯/,ഔ au /au̯ /). A diphthong, also known as a gliding vowel, forms by the movement of tongue, lips, and jaw from one pure vowel sound to another. It is a combination of two adjacent vowel sounds within the same syllable.

D. Speech Defects:

The aim of speech is to convey message in a linguistic manner and it consists of articulation, voice and fluency patterns. There are many children with temporary delays in their speech and language development and they will catch up with it eventually. For some people, improving their communication skills won't be easy. Speech and language disorders are included in the category of communication disorders. A person with a speech impairment may struggle with fluency, resonance or the production of speech sounds (the flow of speech). They are typically diagnosed and treated in accordance with their most prominent impairment, which is typically either speech or language (Jaya Bharti *et al*, 2022).

1. Speech Disorder

Speech disorders affect the ability of a person to form sounds which allow them to interact with others and they differ from language disorders. Language disorders interfere with a person's capacity to learn words or understand what other people are saying, whereas speech disorders prevent people from generating actual speech . A person may find it more difficult to communicate their ideas and emotions if they have either a speech disorder or a language disorder. There are three basic types of speech disorders, and they are speech sound disorders, fluency disorders and voice disorders.

i. Speech Sound Disorder

Speech sound disorder in which some speech sounds (called phonemes) are either not produced correctly or not used correctly. People with speech sound disorders are unable to speak with all of the speech sounds which make their speech hard to understand. In everyday situations people may not understand them. The causes of speech sound disorders are unknown in the majority of cases. Other speech sound difficulties may be related to conditions like a cleft lip and palate, dental issues, hearing loss, or trouble in managing mouth motions.

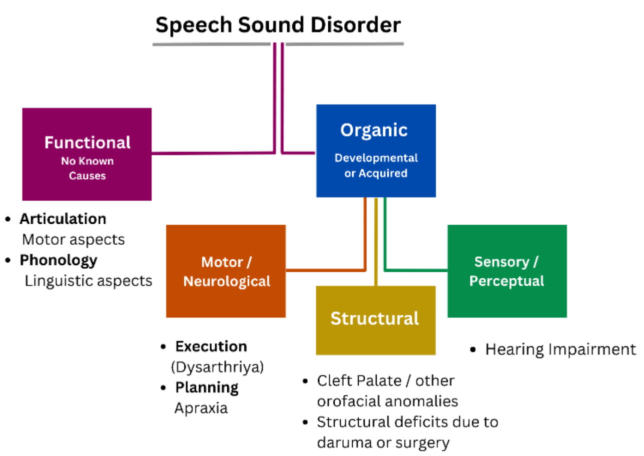

Fig. 1: Speech Sound Disorder

At least half of what the child says should be properly understood by the time he or she turns three. With the exception of a few, most sounds should be spoken correctly by the ages of four and five. Harder sounds should be correctly uttered by the age of seven or eight. Even after reaching the usual age, a kid with a speech sound impairment will still have difficulty pronouncing words. One of the following speech sound problems may be indicated by difficulty with speech patterns:

a. Functional Speech Sound Disorder

Disorders affecting the motor and linguistic components of speech production are both considered to be functional speech sound disorders. These conditions are known as articulation and phonological problems, respectively. Articulation disorder focuses on the mistakes such as distortion and substitutions in speech production, which is the most common type of functional speech sound disorder. Problems with word pronunciation can result from phonological disorders. These complications could be brought on by common grammatical errors, like saying grapheme rather than speech sound. As a subset of phonological disorders, phonemic disorders have an impact on how words are spelled.

Articulation Disorders

By the age of four or five, most kids can make all speaking sounds. After that age, children who struggle with speech clarity may suffer from a speech sound disorder. Articulation disorder, which affects the ability to produce certain speech sounds, is one type of speech sound disorder. Humans make sounds, syllables, and words through certain process called articulation. It may make the child's speech difficult to comprehend and have an impact on the child's socialising and academic performance. There is no recognised reason for articulation problem. It doesn't imply that the child's brain, nerves, lips, tongue, teeth, jaw, lungs, or hearing are damaged in any way. Understanding a child with articulation issues could be challenging. They exhibit one or more of four types of articulation errors:

- Addition: the insertion of unnecessary sounds or syllables into words
- Distortion: altering a sound that might seem like a lisp
- Omission: omitting certain consonants from speech
- Substitution: replacing one sound with another

Phonological Disorders

Articulation disorder and phonological disorder are similar and often confused because these two have different treatments and is difficult to distinguish. In children with phonological issue, it may produce correct pronunciation of sounds, but the sounds to be combined appropriately becomes difficult. A child can have both types of disorders at the same time.

ii. Organic Speech Sound Disorder

Organic speech sound disorders include those resulting from:

- Motor or neurological disorders (E.g. childhood apraxia of speech and dysarthria)
- Structural abnormalities (E.g. cleft lip or palate and other structural deficits or anomalies)
- Sensory or perceptual disorders (E.g. hearing loss)

Motor Speech Disorders: Apraxia and Dysarthria

In terms of motor speech disorders, apraxia and dysarthria are the two main types. Everyone must coordinate a range of muscles and muscle groups, including those regulating the larynx with the vocal cords, the lips, the tongue, the jaw, and the respiratory system, in order to produce speech. Running speech is the result of coordinated motions that are planned by the brain, executed by the corresponding muscles in milliseconds with the proper timing and force. Apraxia or dysarthria may be the result of difficulties in this process, whether they happen in children as a result of development and inborn neurological differences or in adults as a result of injuries, diseases or neurological changes.

Apraxia

The term "apraxia of speech" (AOS) refers to a central nervous system (CNS)-damaged neurogenic speech impairment. AOS is linked to a dysfunction in motor planning and/or programming. The CNS injury makes it challenging to train the precise motions required for clear and expressive speaking, even when there is no weakening or paralysis of the muscles.

Children with childhood apraxia of speech experience difficulties with the accuracy and consistency of the motions that underlie speaking. For the purpose of positioning and sequentially moving muscles to produce speech, these people have sensorimotor problems. These people struggle to plan and programme their movement patterns and because of their inconsistent and numerous articulation faults, they usually exhibit groping behaviours and poor intelligibility.

Dysarthria

The word "dysarthria" refers to a collection of speech problems caused by damage to the central nervous system (CNS) and/or the peripheral nervous system, which impairs motor speech control and execution (PNS). The type of dysarthria, the related traits and the treatment objectives are categorised by the area of damage. Usually, the speaking muscles become weak or uncoordinated as a result of this damage. Cerebral palsy, brain injury, degenerative condition, tumor, and stroke can all result in dysarthria. All of the speech systems must be taken into account in the diagnosis and therapy of this condition because it affects all of them.

Structural abnormalities: cleft lip and palate

Cleft lip and cleft palate, also known as orofacial cleft, is a group of conditions that includes cleft lip (CL), cleft palate (CP) and both together (CLP). A cleft lip contains an opening in the upper lip that may extend into the nose. The opening may be on one side, both sides or in the middle. A cleft palate is when the roof of the mouth contains an opening into the nose. These disorders can result in feeding problems, speech problems, hearing problems, and frequent ear infections. Less than half the time the condition is associated with other disorders. Cleft lip and palate are the result of tissues of the face not joining properly during development resulting in birth defect. The cause is unknown in most cases. Risk factors include smoking during pregnancy, diabetes, obesity, an older mother and certain medications (such as some used to treat seizures). Cleft lip and cleft palate can often be diagnosed during pregnancy with an ultrasound examination. A cleft lip or palate can be successfully treated with surgery and is often done in the first few months of birth for cleft lip and before eighteen months for cleft palate. Speech therapy and dental care may also be needed.

Speech problems, commonly referred to as speech impairments, occur when a person's regular speech is disrupted. 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. Failure of the frontonasal and maxillary processes to fuse, results in cleft in the lip, alveolus and nasal floor in variable sizes. 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. 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.

iii. Fluency Disorder

These are difficulties with the rhythm and timing of speech characterized by hesitations, repetitions or prolongations of sounds, syllables, words or phrases. Common fluency disorders include:

Stuttering

The stuttering also known as dysphemia and stammering is a disorder that affects the fluency of speech. It is evident from the past and concurrent literature survey that, stuttering can be an appraised genetic disorder. It is the rapid-fire repetitions of consonant or vowel sounds especially at the beginning of words, prolongations, hesitations, interjections and complete verbal blocks. Based on previous studies stuttering can be divided into different types.

Developmental stuttering

Developmental stuttering is very common among children, they are unable to get command on verbal; skill as their speech and language processes are underdeveloped phase.

Neurologic stuttering

Neurogenic stuttering is caused by the impairment between motor control, nerves and muscle contradiction.

Psychogenic stuttering

Psychogenic stuttering is directly connected with patient’s mental stress and speaking behaviours. Stuttering is one of the serious problems focused in speech pathology and it is commonly assumed that stuttering results from cohesion of biological, psychological and even social reactions. Voice carriers various acoustic and linguistic characteristics like basic pitch, subsequent formant frequency etc. Every individual involves different noticeable flow of a speech, rhythms determined by structure and arrangement of larynx, pharynx, oral and nasal activity, paranasal sinuses, and thorax. It occurs in about 1% of the population and has found to affect four times as many males as female. Research carried out shows that stuttering cannot be thoroughly cured but, it may go into remission of time. People Who Stutter (PWS) can overcome from speech dysfluency by shaping the tempo, loudness or duration of their utterance and learn to control the speech fluency under the supervision of appropriate speech pathologist. In some research motor measures have achieved good result, but on the other hand audio-visual indication are used to better classify occurrence of stuttering. In earlier stuttering, repetition and prolongation are the ubiquitous dysfluency, as opposed to the other type of dysfluency. Hence, they are commonly used in stuttering assessment process analysis of the performance of stutter before and after therapy. Conventionally, a Speech Language Pathologist (SLP) is employed to count the dysfluencies, measurement of severity and classify the episode of stuttering manually, to keep track of the enhancement of treatment. These types of stuttering assessment are subjective, inconsistent, time consuming and prone to error. Therefore, it might be better if stuttering assessment can be done automatically and get the maximum time for the treatment session. One of the important aspects of dysfluencies detection in speech technology is to accumulate the Automatic Speech Recognition (ASR) system to decrease the recognition error.

iv. Voice Disorder

These are problems with the quality or use of one's voice resulting from disorders in the larynx. Voice disorders are characterized by abnormal production and/or absences of vocal quality, pitch, loudness, resonance, and/or duration. It includes: Hypernasality, Hyponsality etc.

2. Language Disorder

The ability of a child to understand or use language is affected by a language disorder, which is a communication disorder. This is not the same as a speech disorder, which affects how a child makes sounds. Between the ages of 3 and 5, language problems are detected in about 5% of young children. All forms of communication are often affected by language problems, which can have an impact on a child's performance at home, in the classroom and in social settings. Learning any language will be difficult for a child with a language disorder.

Receptive language disorder

Understanding language is challenging for kids with receptive language impairment. They have trouble understanding both what they hear or read and what other people are saying. They frequently give responses that are illogical.

Expressive language disorder

Children with expressive language impairment have problems using words. They might be able to understand what others are saying, but they could find it difficult to communicate with others or to articulate their needs and ideas. The condition can have an impact on verbal, written and sign language.

Mixed receptive-expressive language disorder

Children with both disorders at the same time have difficulty understanding what others say as well as being understood by others.

E. Method and Materials:

This study of speech defects in Malayalam language is based on speech signal analysis. The steps involve both time domain and frequency domain analysis of human voice features. The analysis is done by extraction of features by signal processing algorithms. The steps involve pre-processing such as noise removal using filtering and identifying prominent features in the signal. After identification of the features, they are separated using segmentation and feature extraction.

For the research, we require a speech corpus database in Malayalam language with different types of speech defects. The majority of studies on speech impairments has concentrated on single words rather than taking into account people's abilities at the sentence level. The objective here is to develop a model based on speech defects in Malayalam language. The database will also be helpful for conducting a study based on learning using Artificial Neural Network (ANN). For this we require speech samples of length up to 30 seconds for different types of speech defects such as Tongue-tie, Stuttering/stammering, Aphasia, Cleft lip & palate, Speech sound disorder, Dysarthria and Apraxia. We require samples for the speech defects in different age groups: Child (5-12 years), Adolescence (13-19 years) and Adult (20-59 years).

Data collection and analysis is an important part in this study, The speech samples should be collected in a quiet environment, such as a voice recording booth or a place having minimal external noise and echo problems. High quality dynamic and unidirectional microphone, condenser studio grade microphone is used for recording. An analog to digital interface with 24-bit resolution is used to convert analog speech signal to digital samples with low distortion and ultra-low noise. Depending on the environment of recording, vocal microphone having uniform cardioid pattern with a frequency response of 50 to 15000 Hz or Condenser studio microphone having open circular sensitivity and frequency response 20-20000 Hz is used. The samples are recorded using workstations.

The recorded samples are processed using various speech processing methods, such as noise removal to remove unwanted signals from the recorded data and isolate the desired data alone. This noise removed sample then undergoes segmentation process, in which the sample is divided into different speech segments. In this work, we use several features such as energy, pitch, formant and spectral features such as Linear Prediction Coefficients (LPC), Mel-Frequency Cepstral Coefficients (MFCC) to extract the required information. The analysis of speech signal and programming is done using MATLAB and Python. Classification of speech defects are done based on the analysis of speech samples.

F. Results:

1. Classification of Speech Defects in Speech Impaired Children

Analysis of Speech Samples

Speech is a quasi-periodic signal which lies in the band range between 300 Hz and 3500 Hz. In earlier studies, dysfluency analysis were done by identifying single instances and also the dysfluencies to join or cluster together on same or adjacent words. Some studies have thrown light on the clustering phenomenon in preschooler's speech. A study has been conducted on dysfluencies in the speech of 10–15-year-old children.

Comparison of Defective Speech with Normal Speech

Four children (3 girls and 1 boy) in the age group 10-15 years participated in the study, the children are students of special school, WMO Orphanage, Muttil and they are partially or completely deaf and dumb and also, they have some speech problems Speech samples were elicited using the sample phrases developed during the study. Simple questions were asked to involve the child in conversation and for word utterances, pictures, flash cards and cartoons were used. Different speech samples of same words from selected students were collected. Five words from each child was considered for the study. The speech samples from the children were audio recorded using a digital wave recording. For each speech sample, plots of time domain spectrum were obtained and converted them to frequency domain spectrum using Short Time Fourier Transform (STFT).

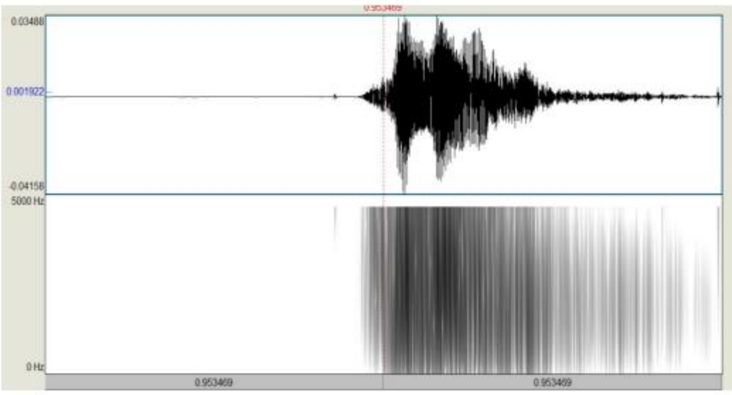

Fig. 2: Spectrogram of normal speech "hello"

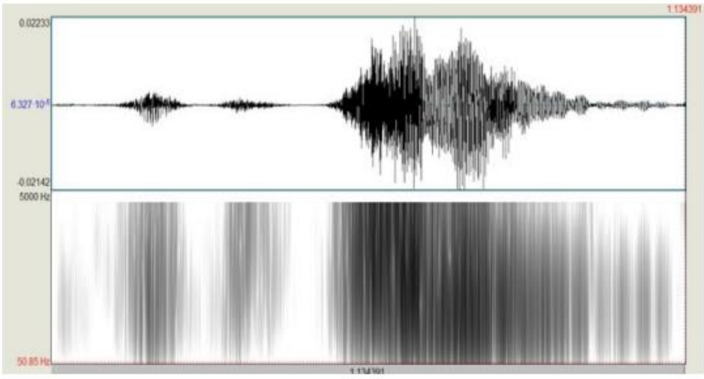

Fig. 3: Spectrogram of defective speech "hello"

The figures 2 & 3 show spectrograms of normal and defective speech "hello" using STFT. In the plot above amplitude versus time is shown, which indicates the signal strength or the amplitude of voice with respect to time. The plot below shows the STFT which has been obtained using hamming window, which divides the signal into sections of fixed length. Frequency analysis of normal "hello" has been obtained in figure 4, which shows the amplitude gain versus frequency plot.

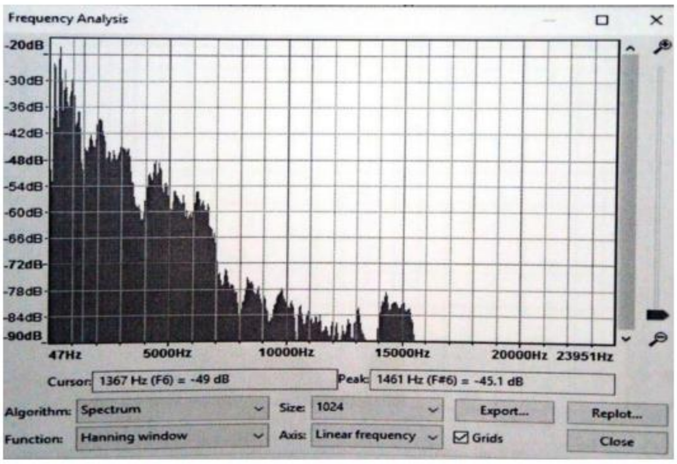

Fig. 4: Frequency analysis of normal "hello"

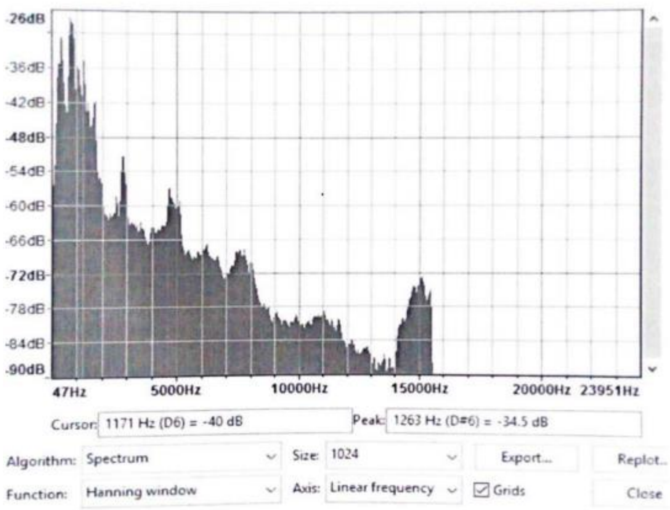

Fig. 5: Frequency analysis of defective "hello"

The plot above is the amplitude versus time plot and the plot below shows the STFT which has been obtained using hamming window.

In figure 5 frequency analysis of defective "hello" has been obtained. Comparison between normal speech and defective speech has been done based on time domain and frequency domain spectral analysis. As can be seen spreading of the continuous speech signal in the defective speech spectrum. Also, from the sample obtained after applying STFT reveals that the samples are spread. For a specified interval between 12x104 and zero. Here it can be inferred that the defective speech having the characteristics of tongue tideness has a spread spectrum when compared with normal speech.

In this work, a study based on different defective speech sample were studied. It has been found that each kind of defect produces a continuous speech spectrum in time domain as well as in frequency domain. From the shape of the envelope and the density of samples, the particular speech defect can be identified. An attempt has been made to classify the various speech defects which help in the further study defective speech. From the analysis carried out, it is evident that, there are different speech production models in the production of defective speech which are modification of the speech production model in normal speech. The work can be extended for the determination of specific speech defects using the defective production model. The expert system designed will be able to identify the factors responsible for various speech disorders. The work is beneficial to know the area of improvement in. of childhood abnormal speech at the early stages of childhood.

2. Gender Based Formant Analysis of Malayalam Vowel Phoneme

Malayalam belongs to the Southern branch of the Dravidian language family. Although closely related to Tamil, Sanskrit influenced Malayalam than Tamil. Malayalam is one of the 22 languages and 14 regional languages of India. Malayalam has 10 vowel phonemes. A vowel is a sound produced by the voiced excitement of the open vocal tract. In speech and speaker recognition, the Formant frequency plays a significant role. In this paper, the first four formants F1, F2, F3, and F4 are analysed using the LPC model in Malayalam vowel phonemes. 40 normal adults (20 males and 20 females) in the age group of 18-45 years participated in this study.

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.

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 1: 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.

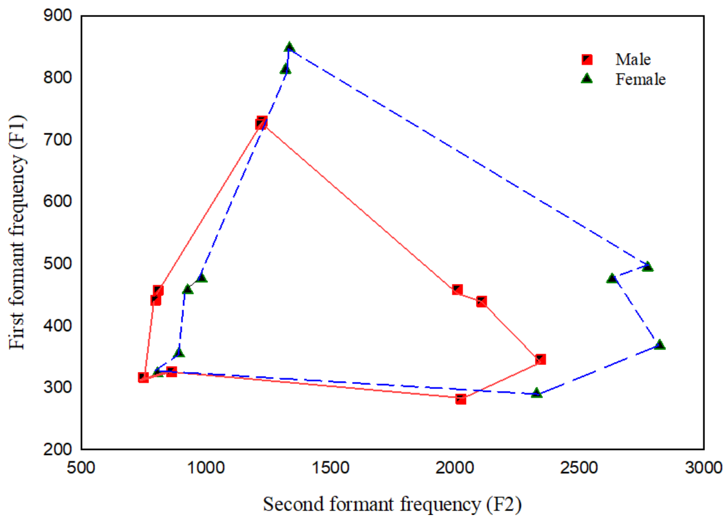

Fig 2: Male and Female acoustic vowel space in Malayalam

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 1 and Figure 6, 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.

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.

3. Formant Analysis of Malayalam Phonemes for Cleft lip

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.

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 2: 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 as seen in table 2. Table 3 shows some selected consonants because of the higher role of lip and nasal components in the formation of these letters.

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

Table 3: Malayalam consonant syllables

Syllables		IPA
Dental	ത	t̪a
	ല	la
Labial	മ	ma
	പ	pa
	വ	ʋa
Retroflex	ള	ɭa
	ഴ	ɻa
	റ	ɾa

In this study, Tables 4 and 5 show the values of normal and cleft lip speech for Malayalam vowels and consonants respectively.

Table 4: 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 r̥	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 5: Formant frequencies of some Malayalam consonants

	F1		F2		F3	
	Normal	*Cleft*	*Normal*	*Cleft*	*Normal*	*Cleft*
ത ṯa	753	242	1603	2234	2928	4777
ല la	818	276	1488	2109	2954	4859
പ pa	837	211	1711	2179	4083	4961
വ ʋa	787	180	1538	1163	3324	2515
മ ma	574	235	1370	1364	2895	5727
ര ɾa	769	266	1700	2096	3235	4746
ള ɭa	826	256	1361	1848	3090	2398
ഴ ɻa	765	272	1446	2231	2906	6454

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

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

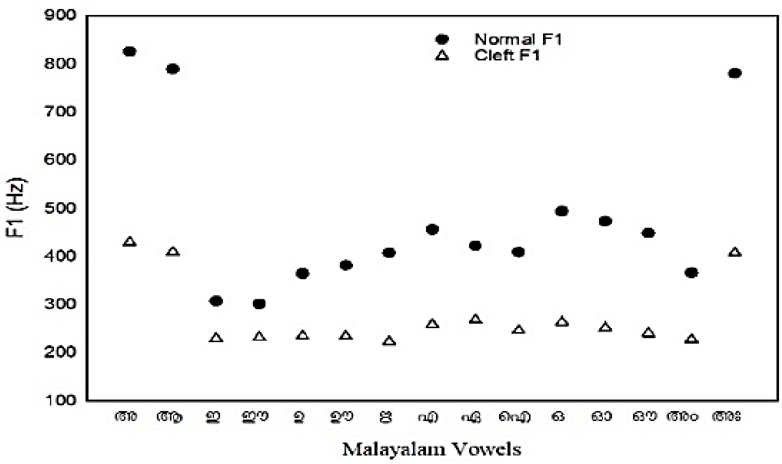

Fig. 3: F1 values for normal and cleft lip

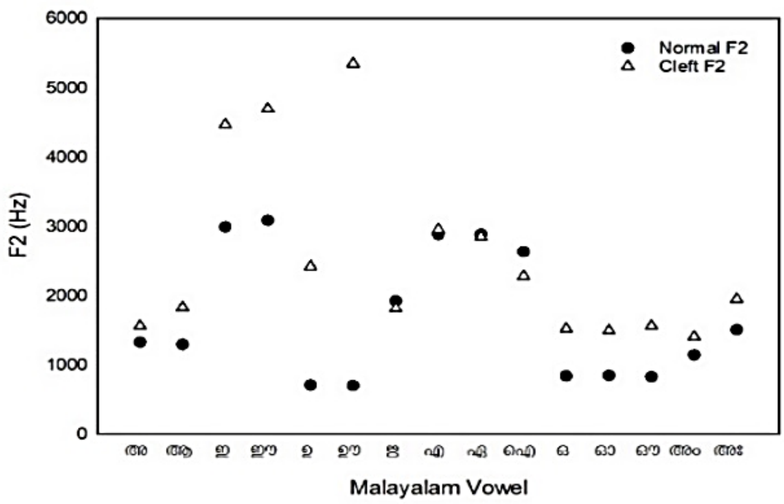

Fig. 4: F2 values for normal and cleft lip

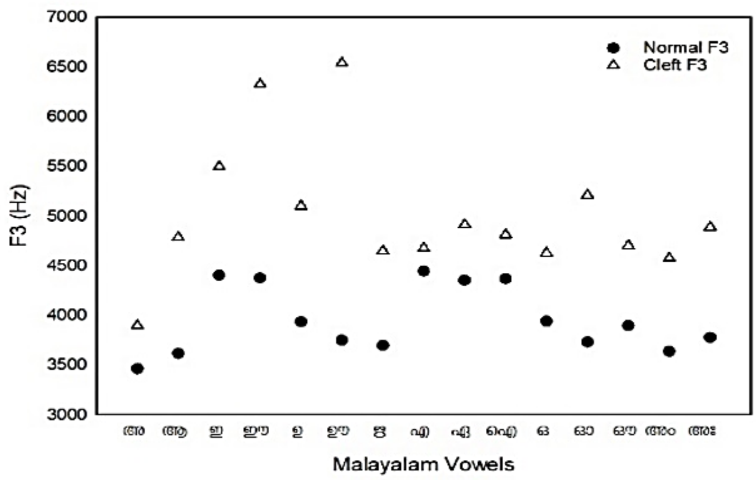

Fig. 5: F3 values for normal and cleft lip

F1 values for normal speech and cleft lip are represented graphically in Fig.4. 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 5, The F2 frequency of cleft lip patients is higher than normal speech because of the more front position of the tongue, as seen in the figure, however, the difference is rather slight in some letters. F3 is compared in Fig.6. where the third formant seems to be higher for the cleft lip subject and in each case there is a clear difference. In some utterances, the F3 deference is greater, and the lip has a greater influence on how that letter is articulated. F3 difference is higher for long vowels than short vowels.

Finally, the findings of this study show a strong relationship between formant frequency and Malayalam vowel phoneme sound of a patient with cleft lips, whose cleft lip surgery was performed earlier. Some key discoveries emerge from the measurement and analysis. Because of their hypernasality, in cleft-lip subjects, the first formant (F1) will be lower than normal voice, while F2 and F3 would be higher. F2 and F3 formant frequencies play a key role in estimating the severity of cleft lip. The pronunciation difficulty will vary according to F3 variations, i.e., the articulation difficulty will be higher for higher F3. For words or utterances that rely more on the case of the 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 mouthparts are oriented at the time of pronunciation. Because of articulation difficulties, several consonant articulations were substituted with another letter.

4. Spectral Analysis of Cleft Lip Speech with Normal Speech

Voice samples from the healthy public were collected and compared to defective speech candidates. 20 healthy persons and 15 people with speech defect who speak Malayalam as their native language participated in the study. The speech samples were recorded with a high-quality microphone in a quiet environment. For phonemic character investigations, test words were created to place nasal vowels and consonants in various phonologic conditions. The test letters and words include Malayalam Vowels (Swaraksharam) and a few short words. Lips and tongue are more influenced and evaluated when pronouncing words. The samples are taken from healthy and cleft-lip participants at a sampling frequency of 16 kHz.

For each speech sample, plots of time domain spectrum were obtained and converted them to frequency domain spectrum using Short Time Fourier Transform (STFT). In this study, the comparison between normal speech and defective speech had done based on time domain and frequency domain spectral analysis.

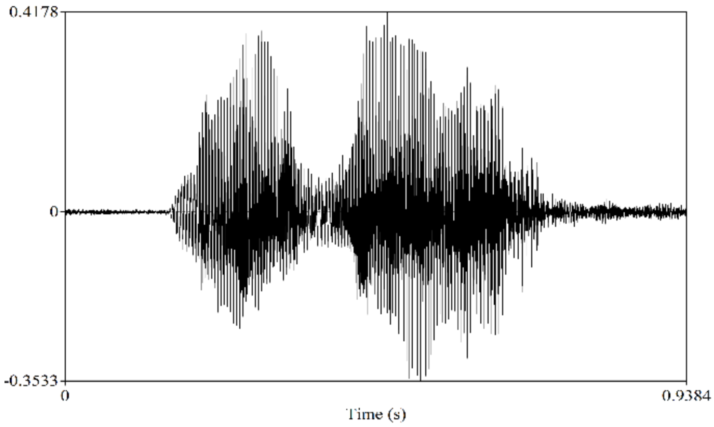

Fig. 6: Normal speech signal

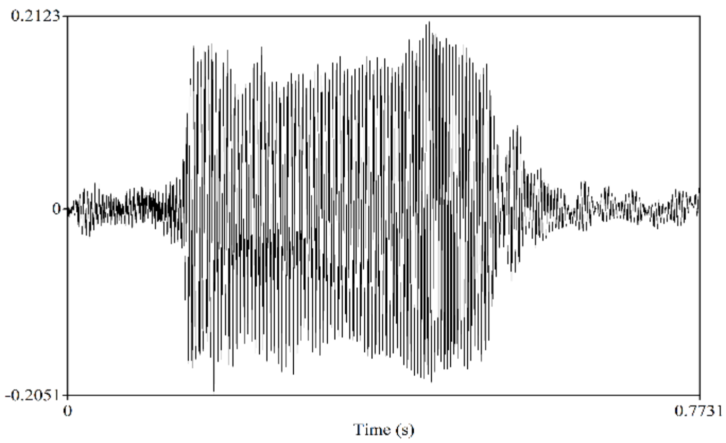

Fig. 7: Cleft lip speech signal

Figures 6 and 7 show the normal and cleft lip speech signal for Malayalam word മഴ (maɻa) respectively. There is a spreading of the continuous speech signal in the cleft lip speech spectrum.

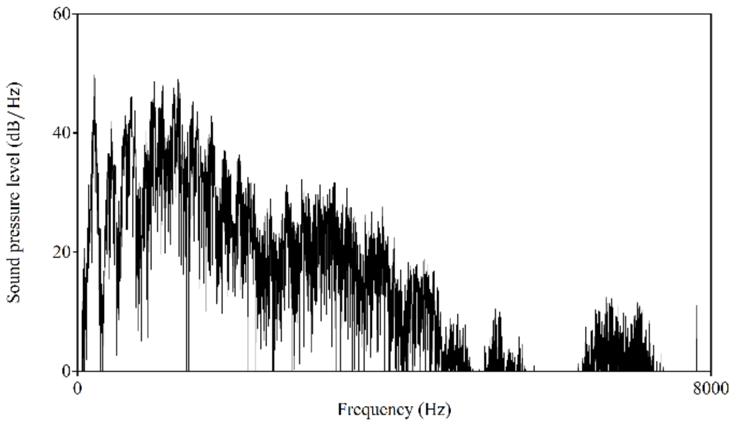

Fig. 8: Spectrum of normal speech signal

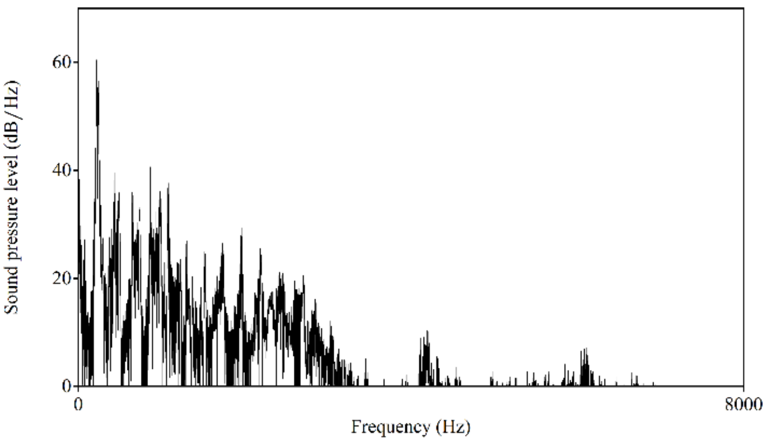

Fig. 9. Spectrum of cleft lip speech signal

Figures 8 and 9 show the spectrum of normal and cleft lip speech samples respectively. It is the relation between sound pressure level and frequency. Sound pressure levels, measured in decibels (dB), a way of validating the intensity of a sound occurrence in decibels.

For normal speech, the sound pressure level is in the range of 30-60dB. From the figures 8 and 9, the sound pressure level of normal speech signal is between 30-60dB, but in the case of cleft lip speech signal, the sound pressure level comparatively less than that of normal speech.

The purpose of this paper is to compare the cleft lip speech with normal speech signal. From the results, it is concluded that in time domain analysis the cleft lip speech signal is broadened than the normal speech signal. In frequency domain analysis, it is clearly seen that the sound pressure level of cleft lip speech signal is lower than the normal speech signal. This is due to the speaker's difficulty to pronounce actual sounds of words and letters due to the influence of extra sounds, elongating words, frequent pauses, distorting sounds, hoarseness, omissions, and substitutions of sounds.

5. Defective Speech Classification & Recognition

Defective Speech Recognition is a huge challenge for distinguishing and interpreting the speech conditions. Fortunately, deep learning is proved to have great ability to deal with acoustic features. For instance, Long Short Term Memory (LSTM) has an advantage of solving time series acoustic features and Convolutional Neural Network (CNN) can discover the local structure among different features.

The model has created on a dataset from the Audio Database of Defective Speech in Malayalam Language. Based on recent studies, Mel-Spectrogram helps to extract important features from audio data and those features were used in the CNN-LSTM model.

This shows how to create a 2-D CNN-LSTM network for speech classification tasks by combining a 2-D convolutional neural network (CNN) with a long short-term memory (LSTM) layer. A CNN processes sequence data by applying sliding convolutional filters to the input. A CNN can learn features from both spatial and time dimensions. An LSTM network processes sequence data by looping over time steps and learning long-term dependencies between time steps. A CNN-LSTM network use convolutional and LSTM layers to learn from the training data.

To train a CNN-LSTM network with audio data, extract auditory-based spectrograms from the raw audio data and then train the network using the spectrograms. This diagram illustrates the network application.

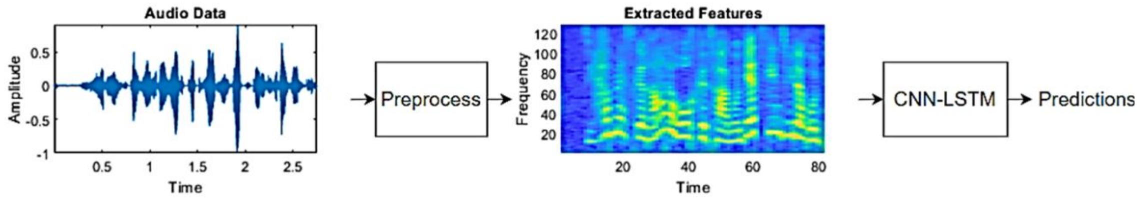

Fig. 10. CNN-LSTM network application

It trains a 2-D CNN-LSTM network to recognize the speech condition (Normal / defect) of spoken text by using the Speech database. The speech conditions are text-independent, which means that the data contains no textual clues that indicate the whether it is normal or defect. The dataset contains 380 utterances labelled with one of these speech conditions: cleft lip & palate before surgery, cleft lip & palate after surgery and undergoing speech therapy, or normal.

```
Speech_condition: 380×1 categorical

    Values:

        CLP after sugery & ST        80
        CLP before Surgery          120
        Normal_speech               180
```

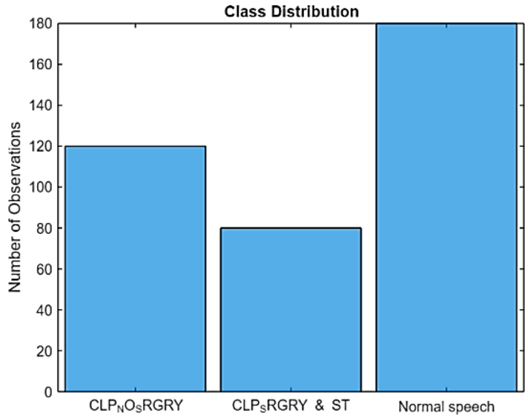

Fig. 11. Class distribution of dataset

2-D CNN LSTM Architecture

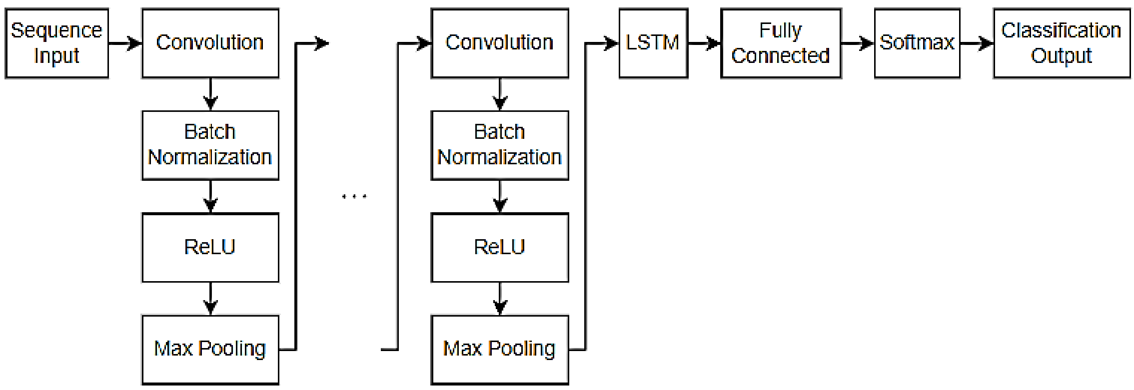

Fig. 12. CNN-LSTM architecture

For sequence input, specify a sequence input layer with an input size matching the input data. For ensure that the network supports the training data, set the MinLength option to the length of the shortest sequence in the training data and learn spatial relations in the 1-D image sequences, use a 2-D CNN architecture with four repeating blocks of convolutional, batch normalization, ReLU, and max pooling layers. Specify an increasing number of filters for the third and fourth convolutional layers. To learn long-term dependencies in the 1-D image sequences, include an LSTM layer with 256 hidden units and to map the sequences to a single value for prediction, output only the last time step by setting the OutputMode option to "last". For classification, included a fully connected layer with a size equal to the number of classes. To convert the output to vectors of probabilities, included a softmax layer; also, included a classification layer.

For specifying the training options using the trainingOptions function: first Train a network using the Adam solver with a mini-batch size of 32 for three epochs and Train with an initial learning rate of 0.005 and reduce the learning rate in a piecewise manner after two epochs. To avoid overfitting the training data, specify an L2 regularization term with a value of 0.0005 and To prevent padding values affecting the last time steps of the sequences that the LSTM layer outputs, left-pad the training sequences. Then Shuffled the data every epoch and validate the training progress using the validation data once per epoch. Displayed the training progress in a plot and suppress verbose output.

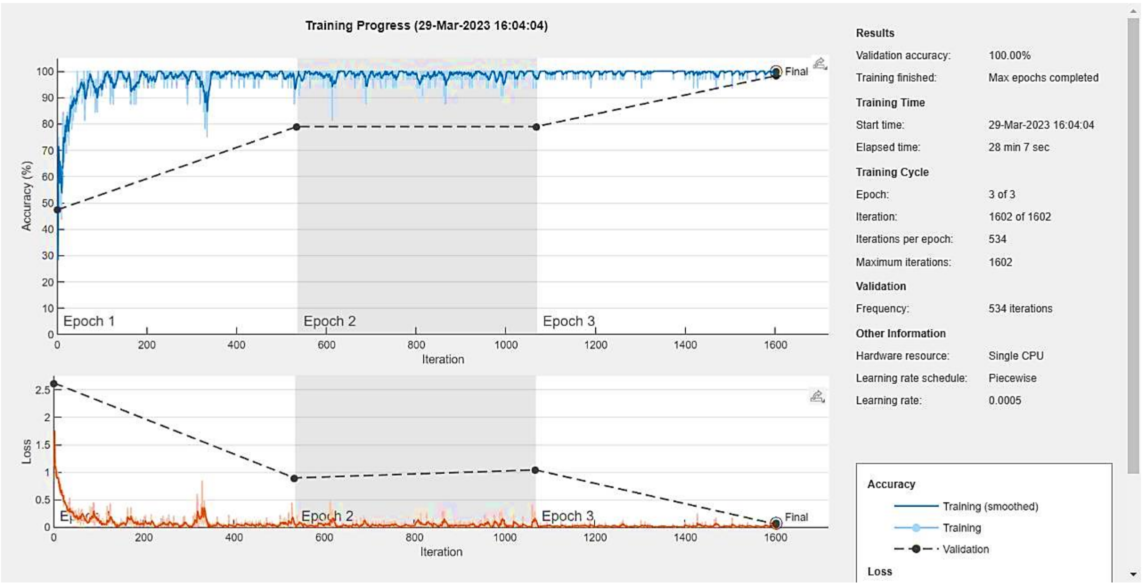

Fig. 13. Training progress

The whole data is divided as follows, 60% of the data is used for training, 15% of the data is used for validation and the remaining 25% of data is used for testing.

The figure shows the training progress during data validation. As seen in the plot, 100 % accuracy is achieved by 15 % of the data validation.

Tested the classification accuracy of the model by comparing the predictions on the held-out test set with the true labels for each test observation and classified the test data using the trained network.

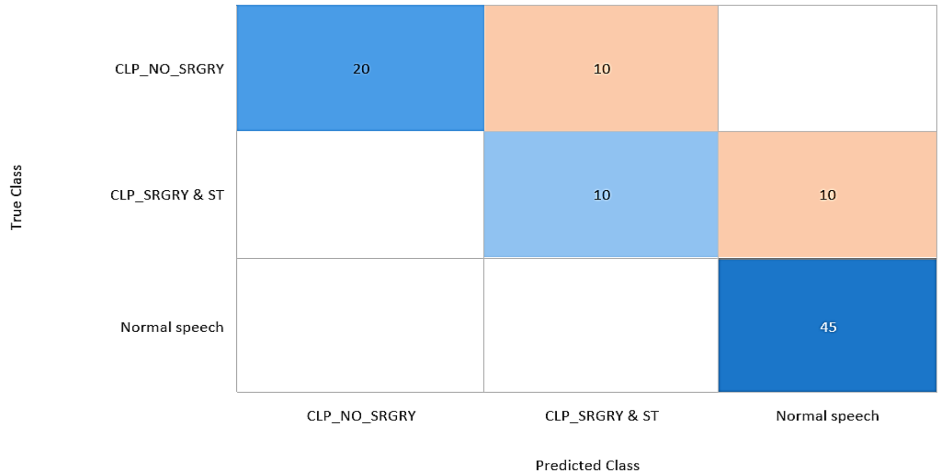

Fig. 14. Confusion matrix for the examined dataset

This confusion matrix shown in the figure summarises speech defect recognition and classification. 25% of the data is used as test data. It plots all the predicted and actual values of the speech classifier. The dataset has 3 classes. Cleft lip and palate before surgery, cleft lip and palate after surgery with speech therapy, and third one is normal speech. In figure, class 1 in which most of the data is predicted as actual data. But some speech samples of cleft lip and palate before surgery were predicted as class 2. Similarly, some of the data of class 2 is predicted as actual and some of them are predicted as normal speech. But it can be seen that there is a clear improvement of the cleft lip and palate after surgery. And, in normal speech class, all the data were predicted correctly. The classification accuracy is defined as the ratio of the total number of predicted class to the actual class, here it is 79%.

References

1. Jaya Bharti, Sangita Srivasatava & Manini Srivasatava (2022). “Language and Speech Disorders in Children”, Department of Psychology, A.N.D.N.N.M.M. Kanpur, IJARIIE-ISSN(O)-2395-4396, Vol-8 Issue-1.
2. Digital-Signal-Processing- 4lh edition-John.G. Proakis- and -Diinitis G.Manolakis.
3. SPEECH AND AUDIO SIGNAL PROCESSING Processing and Perception of Speech and Music Second Edition BEN GOLD Massachusetts Institute of Technology Lincoln Laboratory),NELSON MORGAN (International Computer Science Institute and University of California at Berkeley), DAN ELLIS(Columbia University and International Computer Science Institute).
4. Speech and Audio Signal Processing- A.R. Jayan(Professor Department of Electronics and Communication, Government Engineering College Engineering Sreckrishnapuram Palakkad, Kerala).
5. A Comparative study of Recognition Technique used for development of automatic Stuttered speech dysflucncy recognition system (Swapnil D. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, Vishal B. Waghmare, Ganesh B. Janvale, Babasaheb Sonawane).
6. Recognition and Classification Of Speech And Its Related Monica Mundada, Fluency Disorders Bharti Gawali „Sangramsing Kayte Department of CS &. IT, Ambedkar Dr.Babasaheb Marathwada University Aurangabad .MS (India).
7. OVERVIEW OF ANALYSIS AND CLASSIFICATION OF STUTTERED SPEECH - G. MANJUL (Assoc Prof, Dcpt. of ECE,) M. Shiva Kumar (Prof & Head, Dept. of EIE, GSSSIETW, Mysuru.).
8. The Scientist and Engineer's Guide to Digital Signal Processing.
9. Communication Disorders Pal Caldwell, SLP Speech and Language Pathologist.
10. Clustering dysfluency in 2.1 to 3 year old Kannada speaking children- Anjana B. Ram(Speech and Hearing), Savithri S.R(Speech and hearing).
11. 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.
12. 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.
13. 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.
14. 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.

15. 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.
16. 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.
17. 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.
18. Semer N, "Practical plastic surgery for non-surgeons," Philadelphia: Hanley&Belfus, Inc; 2001. pp. 235–43.
19. 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
20. 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.
21. 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.
22. A.R. Jayan, "Speech and Audio Signal Processing," Department of Electronics and Communication Engineering, Government Engineering College Sreekrishnapuram, Palakkad, Kerala, 2017.
23. J. D. Markel and A. H. Gray, "Linear Prediction of Speech," New York: Springer-Verlag, (Issss976), 1977. 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
24. 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.
25. 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.
26. 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.

27. 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.
28. 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.
29. O’Shaughnessy, “Linear predictive coding”. IEEE Potentials, 7(1), 29–32. doi:10.1109/45.1890.
30. Ethnicity Group. “Cepstrum Method”. 1998 http://www.owlnet.rice.edu/~elec532/PROJECTS98/speech/cepstrum/ cepstrum.html
31. 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.

NO. OF PUBLICATION OUT OF THE PROJECT

a) Papers Presented at National-International Conferences/ Seminars

Sl. N	Title of the Paperpresented	Name of the Conference / Seminar	Organised b	Date
1	Classification of Speech Defects in Speech Impaired Children	UGC Sponsored International Conference on Electronics and Advanced Signal Processing (ICESP2020) [ISBN 978-81-945111-2-0]	Organized by Department of Electronics, Baselios Poulose II Catholics College, Piravam, Kerala State, India in Technical association with IETE Kochi Centre, at BPC College, Piravam	13^{th} -14^{th} Feb 2020
2	Gender Based Formant Analysis of Malayalam Vowel Phonemes	International Conference on Statistical Machine Learning and its Applications to Covid-19	Department of Statistics, WMO Arts and Science College, Muttil, Wayanad, Kerala State	6^{th} Oct2021

b) Publications

Sl. No.	Title of the Paper presented	Title of Journal	Volume Pages, Year	ISSN No.
1	Formant Analysis of Malayalam Phonemes for Cleft Lip	International Journal of Scientific Development and Research	Volume 8 Issue 1 PP 731-739	2455-2631

c) **Communicated Journal Papers**

Sl. No.	Title of the Paperpresented	Title of Journal
1	Cleft Lip and Palate Speech Defect Recognition Using 2D CNN-LSTM Network	IEEE/ACM Transactions on Audio, Speech, and Language Processing.
2	Formant Analysis of Malayalam Vowel Phoneme Using LPC Model	Voice and Speech Review
3	Spectral Analysis of Defective and Normal Speech	Voice and Speech Review

Conducted Programs

➢ One Week FDP International Conference

A Faculty Development Programme (FDP) as well as International Conference was held to enable the faculty members for updating their research and pedagogical skills. Higher education institutions are starting to adapt and respond to professionals already in employment, mature learners and the demand from faculty enrolled for skills for employability programmes. To equip faculty members with the required skills and knowledge, WMO Arts and Science College Muttil, Wayanad will pursue Faculty Development Programme being its primary objective.

The one week FDP was based on Research Methodology and Technical Writing Skills whereas the International Conference was on Recent Trends in Speech Processing, Signal Analysis and Acoustics. The aim is to promote innovative culture for trans-disciplinary research and support the creation of new knowledge, improving the quality of doctoral research and inculcating innovative and cognitive thinking. Teachers, researchers and students of various universities attended the conference. Vice chancellor Dr. M.K Jayaraj, University of Calicut inaugurated the seven-day faculty development programme and international conference organised by STRIDE lab, WMO Arts and Science College. Muttil

Various invited talks, relevant to the theme are:

- “Introduction to Research Methods” - Dr. T.P. Muhammad Fareed, Principal, WMO Arts and Science College
- “Data collection and Analysis” - Raju N, Retired Professor, Department of Statistics
- “Application of Machine Learning and Remote Sensing in Agriculture” – Mr. Chandrasekharan C. P., Application Engineer
- “Spatial Data Analysis” – Dr. Sebastian George, Associate Professor, Department of Statistics, Campus Director, Kannur University
- “Technical Writing” – Dr. Umashankar K, Associate Professor, Department of Behavioural Science, Mahe, Bangalore
- “Technical Writing using Latex”- Dr. Hamsa K.K, Retd. Associate Professor, Farook College, Kozhikode
- “Qualitative Research” – Dr. T.C. Mini, Head of Centre of Excellence of Statistical Research & Dept. of Statistics
- “Speech Signal Processing & Automatic Speech Recognition” – Dr. Rajeev Rajan, Associate Professor, GEC,Barton Hill, IEEE Fellow & Ms. Kavya Manohar, Digital University, Kerala
- “Digital Signal Processing Demystified a MATLAB Based Approach & Advanced Topics and Recent Research Trends in Signal Processing” – Dr. M.V. Rajesh, Associate Professor, Dept. of Electronics Engineering, CEC, Alapuzha, IETE Fellow & Vice Chairman, Kochi centre.

Also, teachers and students from various department presented papers on different topics.

Classification of Speech Defects in Speech Impaired Children

Neena K V
Department of Electronics
WMO Arts and Science College
University of Calicut
Wayanad, Kerala, India

Basila mol M B
Department of Physics
WMO Arts and Science College
University of Calicut
Wayanad, Kerala, India

Dr. Nobert Thomas Pallath
Department of Electronics
WMO Arts and Science College
University of Calicut
Wayanad, Kerala, India
nobert.pallath@gmail.com

Abstract— Speech is an integral part of communication. Speech defect is a disorder in which there is a problem in producing voice or how a person reproduces sound. The power of speech has advanced our species in incalculable ways as our earlier homosapien ancestors began communicating close to 3.5 million years ago. Speech is a tool by which one communicate about needs and enable one to share their desire Abstract - can with the people around them. It is a means by which one can connect and build intimate relationships with others also it is an unavoidable part of one's daily life. Speech disorders are frequently frustrating for the speaker when they know what to speak but are impaired in doing so. Having a speech impediment can increase stress and anxiety levels in social situations for the speakers. The main focus of this study is to identify speech defects in Malayalam language. In this paper, the comparison of normal and abnormal voice is done by speech spectrum analysis.

I. Introduction

Speech disorder is a situation in which a person is aware of what should be spoken but is unable to articulate thoughts. Speech impairment is common among both children and adults Some of the speech disorders can be corrected by earlier treatment. Some of the speech disorders are found to be hereditary which are developed over a period of time. Vocal cord damage, brain damage, stroke, respiratory weakness etc. are considered to be some of the causes of speech impairments. Normal speech is usually found to be effortless, but it is actually a complex process that needs precise timing, nerve and muscle control. Language impairments are different from speech impairments. Language impairments is an impairment in understanding and sharing thoughts. Sometimes, an individual may have both speech and language impairments. Repeating sounds which is most often seen in people who stutter, adding extra sounds, elongating words, making jerky movements while talking usually involving the head, blinking several times while talking, visible frustration when trying to communicate, taking frequent pauses when talking, distorting sounds when talking, hoarseness raspy or gravely sounding voice etc., are considered to be some of the common symptoms of speech impairments. The proposed study concentrates on speech impairments among children and modeling of speech defects. Some physiological disorders like orofacial, myofunctional disorders, neurological disorders like aphasia dysarthria and speech sound disorder and functional disorders like preschool language disorder, language-based learning disabilities and selective mutism.

Articulation problems evolve when the patient produces sounds, syllables, or words incorrectly so that listener does not understand what is being said. The attention of the listener is paid more to the way, the words sound than to what they mean. There are many types of sound errors but most of the mistakes fall into one of the three categories i.e., omissions, substitutions or distortions.

Articulation problems may be as a result of a variety of physical handicaps such as cleft palate, cerebral palsy, hearing loss or they may be related to other problems appearing in the mouth such as dental problems. Stuttering (alalia syllabaries) also known as stammering is a speech defect in which the flow of speech is infringed by involuntary repetitions and extensions of sounds, syllables, words or phrases. Also, a silent pause in which the person suffering from stuttering is unable to produce sounds. Aphasia is a language disorder resulting from the brain damage. It affects the parts of brain responsible for language. The main aim of this project is to develop a system that identifies defective speech and to interpret the speech correctly using comparison methods. Analysis of the speech signal is done using features obtained from the extraction of speech signal and detection can be carried out using decoding of the result obtained from the features.

II. Methods And Materials

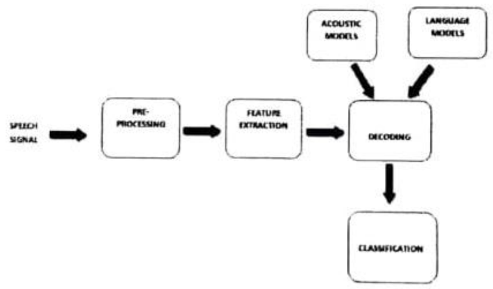

Fig. 1 Block diagram of the speech recognition system

A. *Signal Pre-processing*

This is nothing but the pre-processing step used in the speech signal before analysis. It is mainly used for the noise removal in the speech signal used for analysis. The signal processing step works on the speech signal to reduce the effects of the environment as well as the effects of the channel.

The most common problem in speech processing is the effect of interference noise in the signals. This noise masks the speech signal and reduces its intelligibility. It may be produced by acoustical sources such as ventilation equipment, traffic, crowds and commonly, reverberation and echoes. It can also arise electronically from thermal noise, tape hiss or distortion products. If the sound system has unusually large peaks in its frequency response, the speech signal can even end up masking itself. It is important to cancel the noise which may combine the signal in order to obtain a good quality signal. It can be achieved by active noise cancellation or by adaptive noise cancellation. Different filter techniques can be used to Remove noise from the speech signal.

B. Feature Extraction

The goal of feature analysis is to extract a set of salient features that characterize the spectral properties of the various speech sounds and that can be efficiently measured. The feature extraction stage parameterizes speech waveform, so that the relevant information is enhanced and the non-relevant information is mitigated. There are methods that attempt to extract parameters of a speech production model or to simulate the effect that the speech signal has on the speech perception system or just to use a signal-based method to describe the signal in terms of its fundamental components.

Regardless the method employed to extract features from the speech signal; the features are usually extracted from short segments of the speech signal. This approach comes from the fact that most signal processing techniques assume stationary of the vocal tract, but speech is non- stationary due to constant movement of the articulators during speech production. However, due to the physical limitations on the movement rate, a segment of speech sufficiently short can be considered equivalent to a stationary process. It is like if the segment is extracted from the speech sound during its production.

C. Decoding

After analyzing the salient features from the feature extraction step, the decoding of defective speech signal can be carried out in the decoding stage. Decoding of the original speech signal is done from an existing pool of acoustic and language models.

D. Classification

Classification of the speech signal can be done using two methods; namely rule based classification and Artificial Neural Network classification. Artificial neural network based analysis can be used for classifying defective and non-defective speech signal. This step is performed after retrieving original speech signal. Classification of the speech signal is based on the features extracted from the input speech signal.

III. Experimental Setup And Results

A. Analysis of Speech Samples

Speech is a quasi-periodic signal which lies in the band range between 300 Hz and 3500 Hz. In earlier studies, dysfluency analysis were done by identifying single instanees and also the dysfluencies to join or cluster together on same or adjacent words. Some studies have thrown light on the clustering phenomenon in preschoolers' speech. A study has been conducted on dysfluencies in the speech of 10-15 year old children.

B. Comparison of Defective Speech with Normal Speech

Four children (3 girls and 1 boy) in the age group 10-15 years participated in the study, the children are students of special school, WMO Orphanage, Muttil and they are partially or completely deaf and dumb and also they have some speech problems Speech samples were elicited using the sample phrases developed during the study. Simple questions were asked to involve the child in conversation and for word utterances, pictures, flash cards and cartoons were used. Different speech samples of same words from selected students were collected. Five words from each child was considered for the study. The speech samples from the children were audio recorded using a digital wave recording. For each speech sample, plots of time domain spectrum were obtained and converted them to frequency domain spectrum using Short Time Fourier Transform (STFT).

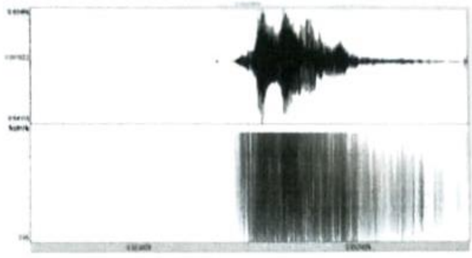

Fig 2 Spectrogram of normal speech "hello"

As shown in the figure 2 is a spectrogram of normal speech "hello" using STFT has been obtained. In the plot above amplitude versus time is shown, which indicates the signal strength or the amplitude of voice with respect to time. The plot below shows the STFT which has been obtained using hamming window, which divides the signal into sections of fixed length. Frequency analysis of normal "hello" has been obtained in figure 3, which shows the amplitude gain versus frequency plot.

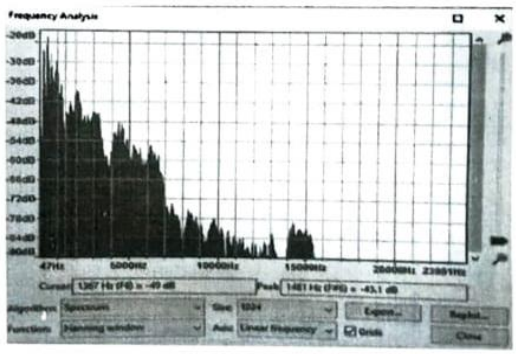

Fig 3 Frequency analysis of normal "hello"

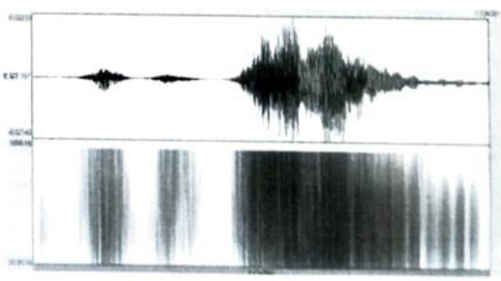

Fig 4 Spectrogram of defective speech "hello"

The plot above is the amplitude versus time plot and the plot below shows the STFT which has been obtained using hamming window.

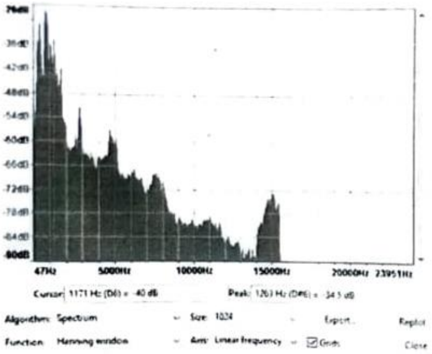

Fig 5. Frequency analysis of defective speech "hello"

In figure 5 frequency analysis of defective "hello" has been obtained. Comparison between normal speech and defective speech has been done based on time domain and frequency domain spectral analysis. As can be seen spreading of the continuous speech signal in the defective speech spectrum. Also, from the sample obtained after applying STFT reveals that the samples are spread. For a specified interval between 12x 104 and zero. Here it can be inferred that the defective speech having the characteristics of tongue tideness has a spread spectrum when compared with normal speech.

IV. CONCLUSION

In this work a study based on different defective speech sample were studied. It has been found that each kind of defect produces a continuous speech spectrum in time domain as well as in frequency domain. From the shape of the envelope and the density of samples, the particular speech defect can be identified. An attempt has been made to classify the various speech defects which help in the further study defective speech. From the analysis carried out, it is evident that, there are different speech production models in the production of defective speech which are modification of the speech production model in normal speech. The work can be extended for the determination of specific speech defects using the defective production model. The expert system designed will be able to identify the factors responsible for various speech disorders. The work is beneficial to know the area of improvement in. of childhood abnormal speech at the early stages of childhood.

REFERENCES

[1] Digital-Signal-Processing- 4th edition-John.G. Proakis- and -Dimitis G.Manolakis.

[2] SPEECH AND AUDIO SIGNAL PROCESSING Processing and Perception of Speech and Music Second Edition BEN GOLD Massachusetts Institute of Technology Lincoln Laboratory),NELSON MORGAN (International Computer Science Institute and University of California at Berkeley), DAN ELLIS(Columbia University and International Computer Science Institute).

[3] Speech and Audio Signal Processing- A.R. Jayan(Professor Department of Electronics and Communication, Government Engineering College Engineering Sreekrishnapuram Palakkad, Kerala).

[4] A Comparative study of Recognition Technique used for development of automatic Stuttered speech dysfluency recognition system (Swapnil D. Waghmare, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, Vishal B. Waghmare, Ganesh B. Janvale, Babasaheb Sonawane).

[5] Recognition and Classification Of Speech And Its Related Monica Mundada, Fluency Disorders Bharti Gawali „Sangramsing Kayte Department of CS & IT, Ambedkar Dr.Babasaheb Marathwada University Aurangabad .MS (India).

[6] OVERVIEW OF ANALYSIS AND CLASSIFICATION OF STUTTERED SPEECH - G. MANJUL(Assoc Prof, Dcpt. of ECE,) M. Shiva Kumar (Prof & Head, Dept. of EIE, GSSSIETW, Mysuru.)

[7] The Scientist and Engineer's Guide to Digital Signal Processing.

[8] Communication Disorders Pal Caldwell, SLP Speech and Language Pathologist.

[9] Clustering dysfluency in 2.1 to 3 year old Kannada speaking children-Anjana B. Ram(Speech and Hearing), Savithri S.R(Speech and hearing).

Gender Based Formant Analysis of Malayalam Vowel Phoneme

Dr. Nobert Thomas Pallath *Department of Electronics WMO Arts and Science CollegeUniversity of Calicut* Wayanad, Kerala, India nobert.pallath@gmail.com
Amala Thomas*Department of Electronics WMO Arts and Science CollegeUniversity of Calicut)* Wayanad, Kerala, India amalathomas178@gmail.com
Ehijas KV *Department of Electronics WMO Arts and Science CollegeUniversity of Calicut)* Wayanad, Kerala, India kvehijas@gmail.com

***Abstract*— Malayalam belongs to the Southern branch of the Dravidian language family. Although closely related to Tamil, Sanskrit influenced Malayalam than Tamil. Malayalam is one of the 22 languages and 14 regional languages of India. Malayalam has 10 vowel phonemes. A vowel is a sound produced by the voiced excitement of the open vocal tract. In speech and speaker recognition, the Formant frequency plays a significant role. In this paper, the first four formants F1, F2, F3, and F4 are analysed using the LPC model in Malayalam vowel phonemes. 40 normal adults (20 males and 20 females) in the age group of 18-45 years participated in this study.**

Keywords*— *Vowel Phoneme, Formants, LPC, Spectrum, Filter, Acoustic vowel space

I. INTRODUCTION

Malayalam (IPA: mælə'jɑ:ləm) is one of the classical languages in India, it is the official language of the state of Kerala and Lakshadweep Islands. Malayalam is one of the toughest languages in India and is known to be a tongue-twisting language since it is difficult to pronounce certain words. When compared with South Indian languages like Tamil, Telugu, and Kannada which are comparably easy to learn, Malayalam is found to be difficult. Malayalam consists of 36 consonants and 15 vowel symbols. Though there are 15 vowel symbols in the Malayalam script, the number of vowel phonemes are only 10 [1].

A phoneme is the smallest unit of sound which makes a difference in meaning when used a language. A phoneme can be classified into two types, vowels and consonants [2].

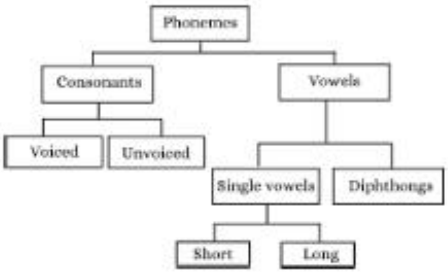

Figure 1. Phoneme Division Flowchart

Sounds is classified as voiced or unvoiced. When produced a sound causes the vocal cords in the larynx to vibrate to produce a voiced sound and, if it is unvoiced sound, there is no vibration. In contrast, it can be found that although consonants may be voiced or unvoiced, all vowel sounds are voiced. It may be single (Monophthongs) or a combination (Diphthongs) [3]. Table 1 shows the Malayalam vowel letters and the corresponding vowel signs of Malayalam script, in ISO 15919 transcriptions inthe International Phonetic Alphabet (IPA). Single vowel sounds may be short or long and the symbol /:/ is used to denote thelong vowel.

Vowels can be produced by controlling the airflow through mouth. It can be adjusted by the position of tongue, rounding of lips,and degree of opened mouth. The vowels can be classified based on the position of tongue in terms of how far it is forward andhow high it is.

Malayalam vowel consists of 5 short vowels (അ a /a/, ഇ i /i/, ഉ **u** /u/, എ **e** /e/, ഒ **o** /o/) and 5 long vowels (ആ ā /a:/, ഈ ī /i:/, ഊ ū /u:/, ഏ ē /e:/, ഓ ō /o:/). ഋ (/ri/, ṛ) is a vocalic consonant, which is not officially a vowel. The vowel

അം (**aṁ** /am/) is an Anusvaram and അഃ (**aḥ** /ah/) is a Visargam. Malayalam has 2 diphthongs (ഐ ai /aj/, ഔ au /au̯ /). A diphthong, also known as a gliding vowel, forms by the movement of tongue, lips, and jaw from one pure vowel sound to another. It is a combination of two adjacent vowel sounds within the same syllable. The written symbols of Malayalam vowel phonemes and their positions are shown in Table 2.

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.

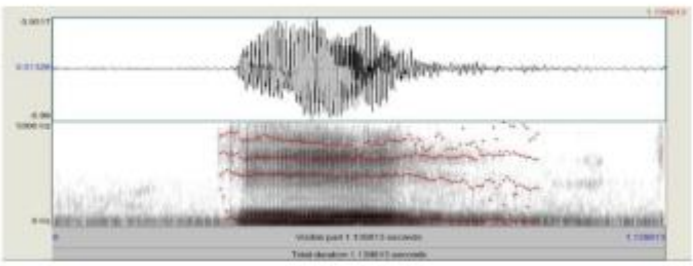

Figure 2. An example of Frequency spectrum and formant frequency plot of a Malayalam vowel phoneme

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), $0 \leq n \leq N-1$, then the result of windowing is the signal:

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

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

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

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

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:

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

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

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

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

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

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

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

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.

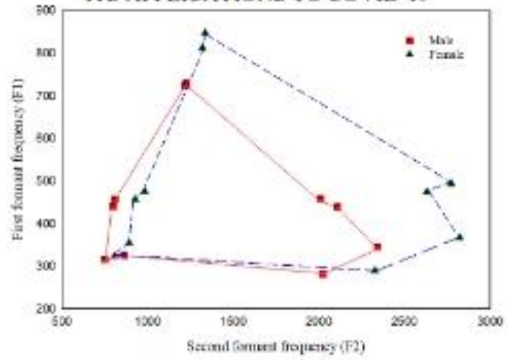

Figure 3. Male and Female acoustic vowel space in Malayalam

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.

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].

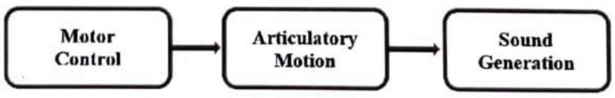

Fig. 1. Block diagram of human speech production mechanism

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.

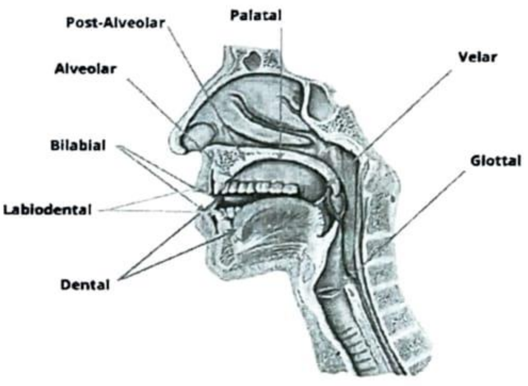

Fig. 2. Anatomy of Speech

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.

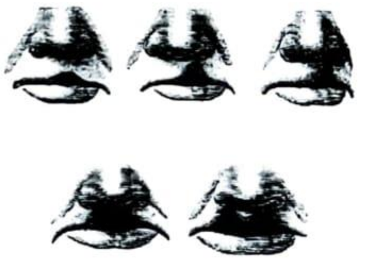

Fig. 3. Cleft Lip

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]

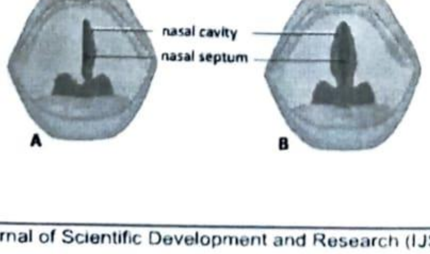

Fig. 4. (A) Unilateral cleft palate. (B) Bilateral cleft palate

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 /ˈfouniːm/ 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/, ന /n̪a/, ണ /ɳa/, ഞ /ɲa/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]:

$$F_n = a_n(F_0 + b_n) \tag{1}$$

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).

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

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

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

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

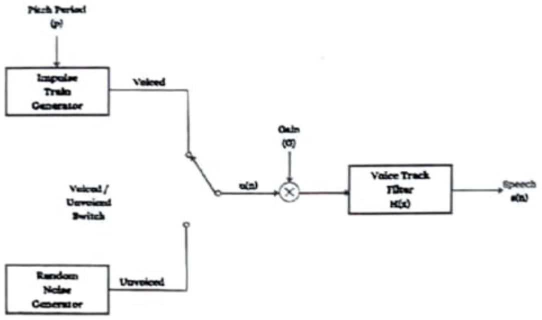

Fig. 5. Model of speech production for LP analysis

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):

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

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

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

   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	aː
ി, ഇ	i	ീ, ഈ	iː
ു, ഉ	u	ൂ, ഊ	uː
െ, എ	e	േ, ഏ	eː
ൊ, ഒ	o	ോ, ഓ	oː
Vocalic Consonant	***IPA***	***Anusvaram***	***IPA***
ൃ, ഋ	r̥	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	ʋa
Retroflex	ള	ḷa
	ഴ	ɻa
	റ	ra

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 *r̥*	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 *ṭa*	753	242	1603	2234	2928	4777
ല *la*	818	276	1488	2109	2954	4859
പ *pa*	837	211	1711	2179	4083	4961
വ *ʋa*	787	180	1538	1163	3324	2515

മ *ma*	574	235	1370	1364	2895	5727
റ *ra*	769	266	1700	2096	3235	4746
ള *ḷa*	826	256	1361	1848	3090	2398
ഴ *ḻa*	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 ഡ 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.

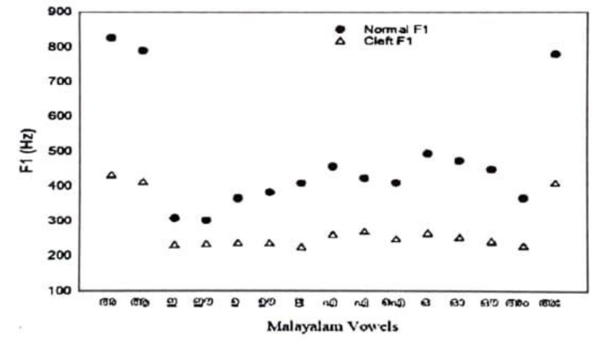

Fig. 6. F1 values for normal and cleft lip

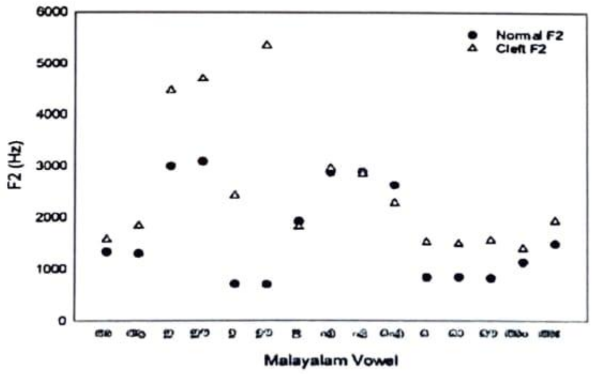

Fig. 7. F2 values for normal and cleft lip

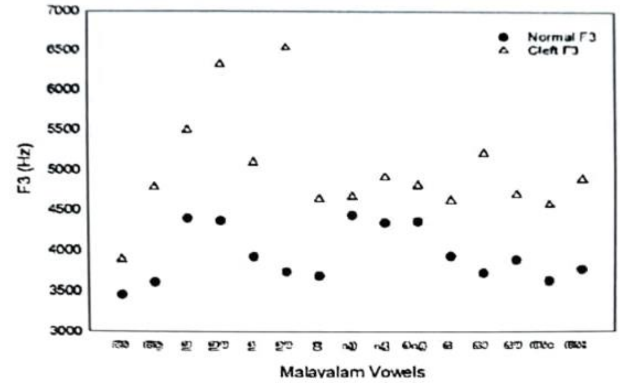

Fig. 8. F3 values for normal and cleft lip

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], "Proposai for a Malayalam Script Root Zone Label Generation Ruleset (LGR)," version 4.0, 2020.