Decision Support Tool to Recognize Speech Disorder of Children in Sri Lanka

Indunil Jayalath

169313N

Supervised By:

Mr. Saminda Premaratne

Faculty of Information Technology

University of Moratuwa

March 2019

Decision Support Tool to Recognize Speech Disorder of Children in Sri Lanka

Indunil Jayalath

169313N

Supervised By:

Mr. Saminda Premaratne

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Master of Science in Information Technology.

Faculty of Information Technology

University of Moratuwa

March 2019

Abstract

This research focuses on the development and management of an intelligent system for voice disorders and has addressed the important issues such as diagnosing and treatment procedures of voice disorders. The diagnostic methods are currently limited to manual procedures; and therefore, some form of IT-assisted tool with intelligent decision making would be a valuable addition to the treatment process. Furthermore, the ability of the tool to monitor the progress of the treatment plans would further enhance the value addition.

The physical properties of a voice signal can be analyzed using the digital signal processing analysis such as pitch frequencies, formant frequencies, and Mel frequencies. These physical properties are varying with age, cultural background, geographical location, and language. Therefore, these parameters have to be fixed when performing an experiment. In order to fix these parameters, the students' samples for the research can be selected from a single population. The members of the sample will comprise students whose age is less than 7 and greater than 5 years. In addition to it, pre-determined sample words called phonetic balance can be chosen for the voice recordings.

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student (s)	Signature of Student (s)
H.D.I.M.Jayalath	…………………………..
	…………………………..
	Date:

Supervised by	………………………………
Name of Supervisor(s)	Signature of Supervisor(s)
Mr.Saminda Premaratne	
Senior Lecturer	
Faculty of Information Technology	
The University of Moratuwa	……………………………..
	Date:

Acknowledgments

I would like to thank the many individuals who made this research project possible and gave their supportive hands all along the research study and made the journey very stimulating and enjoyable. Firstly, I am very grateful to the Department of Information Technology of the University of Moratuwa, for giving me the opportunity to carry out this research study. I would like to express my sincere gratitude to my supervisor Mr. Saminda Premarathne for his valuable guidance and immense support.
I owe a sincere debt to Mr. W. V. Sinhanatha, Educational Audiologist and Speech Therapist and Chief Project Officer at National Institute of Education and Ms. S.A. Malani Chandralatha, Linguistics Expert, for their academic guidance in making a success of this research project.

My special thanks go to Mr. Sunil Piyatissa, Principal Mahaweli Maha Vidyalaya, Polgolla, Mr. S. Palamakumbura, Principal Kengalle Primary College, Kandy, Principal Madawala Primary College Alawwa, Hanthihawa Kanishta Vidyalaya Halmillawewa, and Ms. Nalini Pieris, Teacher at Pathadumbara Maha Vidyalaya, for their assistance to me to perform the speech recordings.

Further my thanks and appreciations are due to my family for their understanding, motivation and patience during this exercise, as well as to my colleagues and friends who helped me in the preparation of this dissertation.

Table of Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Introduction

Speech disorders are the problem of pronouncing words accurately and difficulties of expressing ideas. It can be treated at the early stage, i.e. at the age of 4 to 6 years. It is an important task to identify the problem of the speech disorder prior to preparing the treatment plan. The speech pathologist is responsible for assessing, diagnosing, treating and providing a treatment plan for such treatments.

Speech disorders mainly involve the difficulties of producing the speech such as swallowing, stuttering, cluttering, articulations disorders, and dyspraxia. Most of these problems occur because of timing between pronouncing two words, nerve and muscle control of the vocal system. It is recommended, to consult a speech-language pathologist to create a treatment plan. Speech-language pathologists analyze the type of the problem such as lack of fluency, articulation, or motor skills by recording an audio tape or by listing during conversations. A computer-based decision support system can be used to identify the symptoms of voice disorders. The treatment plan can be designed with the aid of a decision support system, which involves recognizing vowels, consonant sounds, short words and phrases of the speech. The pathologist is responsible for choosing appropriate words and sentences to identify articulator complexity, voice patterns of each patient. The treatments for speech disorders can only be provided at the early stage of a child i.e. age between 6 to 8 years.

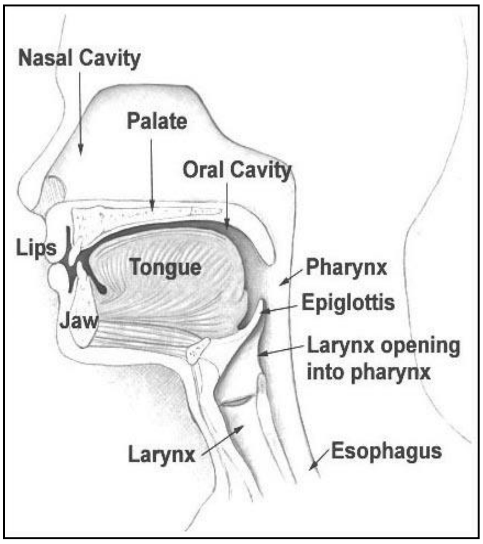


Figure 1-1: The cartilages of larynx, seen from the right-hand side

The physical properties of a voice signal can be analyzed using the digital signal processing such as pitch frequencies, formant frequencies, and Mel frequencies. These physical properties are varying with the age, cultural background, geographical location, and language. These parameters have to be considered when designing an experiment related to voice disorders. According to Australian Dyspraxia Association (2010), the disease has been defined as "a speech disorder that affects the programming, sequencing and initiating of movements required to make speech sounds". Identification of the voice disorder is difficult, because this has been ignored by the parents and the teachers when they found pronunciation problem of a child, and they think that that the problem will outgrow with the development of child's vocabulary. However, the problem becomes worse, because the treatments recommended for the voice disorders have not been provided at the early stage. Otherwise no significant results is achieved.

The conventional method of recognizing the symptoms of voice disorders is by listening to voice conversations. It is difficult to follow the symptoms at the childhood because some patients pronounce some words or sequence of words clearly in a certain situation, but the same words or sequence of words would not be pronounced at a different situation. In such a situation a computer-based system can be recommended to recognize symptoms of voice disorders. It can be done by comparing default attributes of pronouncing words or sequence of words of ordinary children with a disordered voice child. Therefore, a sample of pronouncing words have to be chosen first and its attributes have to be stored in the computer-based system, where it can be used for the above-mentioned comparison.

1.2 Aims and Objectives

The diagnostic and progress monitoring tools to monitor the problems relevant to voice disorder are currently limited to manual methods. Information Technology assisted solution would be extremely useful and effective for the professionals involved in this discipline such as speech pathologists. The aim of this research is to identify a sample of ordinary children's' default voice attributes for certain words and store on a computerized database and, perform analysis such as Pitch Analysis, Formant Frequency Analysis, and Mel Frequency Analysis for the children who are having certain voice disorders.

In order to satisfy the above requirements following research, objectives have been formulated.

- Choose a suitable phonetic balance (i.e., appropriate words including vowels and consonants) for the voice recording of normal children with the age group 6 to 8 years.
- Trace the properties of a voice such as frequency and pitch, in order to use it to identify voice disorder.
- Use the above mentioned properties to design an intelligence system to recognize voice disorder and monitor the progress of voice disorders.

The frequent voice disorders in children have been identified due to the problems in the larynx. Therefore, most of the ENT specialists conduct investigations based on the larynx. The Fluharty-2 is a preliminary Speech and Language Screening test, which tests the physical attributes of the voice and provide a simple conclusion like "sound normal; recheck may be necessary". Due to the lack of intelligence based investigation tools, reduces the investigation procedures of voice disorders. These conventional one-line summaries failed to provide detailed investigations about the voice disorders. These primitive investigations do not provide a detailed study about the respiration, phonation, and resonance. The notable symptom of voice disorder is the inability or difficulty to put sounds and syllables together in order to produce a proper sentence. People with voice disorder can speak with inconsistent mistakes. For example, frequent repetition of words and unusual varying rhythms are common features of their speech. Due to these problems there some other problems can also be noticed in their speech such as poor vocabulary, incorrect grammar, difficulty in clearly organizing spoken information, problems of reading, writing, spellings, problems of language coordination and poor mathematical skills. Identification of symptoms is very important before providing treatments for voice disorder. It is recommended to consult an ENT specialist first and he will conduct initial investigations. He/she then directs the patient to a Speech-Language Pathologist (Speech-Language Therapist) who then plays the key role of diagnosing the disease. Due to the lack of intelligence devices at present, most therapists observe the child's speech over a period of time. Parents are also advised what steps have to be taken to improve the speech of the child. Children with development voice disorder will not outgrow their problem on their own, unless suitable treatments are provided. Speech Language Therapists use different techniques for these patients. No single approach has been recommended. There are different approaches recommended to treat these patients, but all take considerable time. The Nuffield Dyspraxia Program is one of the popular approaches used by therapists. It is a systematic approach with an assessment method, and it is generally recommended for children whose age between 3 to 7 years. The treatment plan will be designed by the speech therapist. It consists of two parts. For the first part, speech characteristics will be identified. It includes an examination of child's usage of vowel and consonants, articulations, vowel distortions, a breakdown in sequencing words, omissions and substitutions, voice difficulties such as volume, length, pitch, and quality, glottal stop insertions, resonance difficulties in terms of rate, rhythm, stress and unintelligible speech. Speech Language Therapists use manual methods to identify voice impairments. However, it is difficult because a person's language differs from age, gender, cultural background, and geographical location.

The second stage of the Nuffield Dyspraxia Program is providing basic oral-motor exercises for the person to improve and develop all areas of speech related 'apparatus' of the person. It includes a list of exercises to improve lip shapes, and movements. Different types of apparatus are also used when performing these exercises. Other exercises will be given for the tongue, soft palate, and larynx and to improve breath control. The progress of these exercises has to be monitored. Otherwise the therapist will not easily trace the progress and sometimes the parents of the child may have negative attitudes of the training program. It is an important issue to trace the progress of voice disorder patients' when they are undergoing training program. However, the diagnostic and progress monitoring tools in current use are limited to manual methods.

1.3 Background

The frequent voice disorders in children have been identified due to the problems in the larynx. Therefore most of the ENT specialists conduct investigations based on the larynx. The Fluharty-2 is a preliminary Speech and Language Screening test, which tests the physical attributes of the voice and provide a simple conclusion like "sound normal; recheck may be necessary". Due to the lack of intelligence based investigation tools, it reduces the investigation procedures of voice disorders. These conventional one-line summaries failed to provide detailed investigations about the voice disorders. These primitive investigations do not provide a detailed study about the respiration, phonation, and resonance.

Voice dyspraxia is a common voice disorder in children, and it is also known as Apraxia. Dyspraxia is not due to the weakness or paralysis of the speech muscles either in the face or lips or tongue. It is a problem of saying some words or sentences by the person. Two types of dyspraxia have been identified. Acquired and Developmental Dyspraxia are the two types. Acquired dyspraxia is common in adults, but it can affect a person at any age. It occurs due to a stroke, head injury, tumor or other illness affecting the brain. This illness sometimes results in damage in a certain part of the brain. These patients have certain impairments in speaking and sometimes completely lose the ability to pronounce words. These damages result in reducing the strength of muscles and relevant nerves which are important to produce sound. The treatment plan for the Acquired Dyspraxia must be managed by an ENT surgeon together with a Voice and Language Therapist.

The Development Apraxia of Speech (DAS) is a disease in pre-school children and is present from birth. It is more common in boys than in girls. The main cause for the DAS is not yet known, but treatment is recommended. Two arguments have been placed for the root cause for the DAS. The first argument is that it occurs due to the problem of language development of the child. The second argument is it occurs due to the neurological disorder of the child. But these two arguments have not yet been confirmed. Most of the countries nonprofit making organizations have been established to the development programs of voice Dyspraxia patients.

The notable symptom of DAS is the inability or difficulty to put sounds and syllables together in order to produce a proper sentence. People with DAS can speak with inconsistent mistakes. For example, frequent repetition of words and unusual varying rhythms are common features of their speech. Due to these problems some other problems can also be noticed in their speech such as poor vocabulary, incorrect grammar, difficulty in clearly organizing spoken information, problems of reading, writing, spelling, problems of language coordination and poor mathematical skills.

Identification of symptoms is very important before providing treatments for patients. It is recommended to consult an ENT specialist first, and he will conduct initial investigations. He/she then directs the patient to a Speech-Language Pathologist (Speech-Language Therapist) who then plays the key role of diagnosing the disease. Due to the lack of intelligence devices in the present, most therapists observe the child's speech over a period of time. Parents are also advised what steps have to be taken to improve the speech of the child. Informal testing for voice disorders, speech-language therapist may ask the patient to repeat certain word or list of words. In addition to this, brain imaging tests like Magnetic Resonance Imaging (MRI) will also be used for the acquired voice dyspraxia patients' for the identification procedure. Development dyspraxia patients are not usually requested to do MRI Scanning procedures.

Children with development voice dyspraxia will not outgrow their problem on their own unless suitable treatments are provided. Speech Language Therapists use different techniques for these patients. No single approach has been recommended. There are different approaches recommended to treat these patients, but all take considerable times. The Nuffield Dyspraxia Program is one of the popular approaches used by therapists. It is a systematic approach with an assessment method, and it is generally recommended for children whose age between 3 to 7 years. The treatment plan will be designed by the speech therapist. It consists of two parts. For the first part, speech characteristics will be identified. It includes an examination of child's usage of vowel and consonants, articulations, vowel distortions, a breakdown in sequencing words, omissions and substitutions, voice difficulties such as volume, length, pitch, and quality, glottal stop insertions, resonance difficulties in terms of rate, rhythm, stress and unintelligible speech. Speech Language Therapists use manual methods to identify voice impairments. However, it is difficult because a person's language differs due to age, gender, cultural background, and geographical location.

The second stage of the Nuffield Dyspraxia Program is providing basic oro-motor exercises for the person to improve and develop all areas of speech related 'apparatus' of the person. It includes a list of exercises to improve lip shapes, and movements. Different types of apparatus are also used when performing these exercises. Other exercises will be given for the tongue, soft palate, and larynx and to improve breath control. The progress of these exercises has to be monitored. Otherwise the therapist will not easily trace the progress, and sometimes the parents of the child may have negative attitudes of the training program. It is an important issue to trace the progress of voice disorder patients' when they are undergoing a training program. However, the diagnostic and progress monitoring tools in current use are limited to manual methods.

Chapter 2

2 Literature Review

There are different types of voice disorders. Most of the voice disorders have to be treated at an early stage. The treatment plan for voice disorders has to be created by the involvement of a professionally qualified speech-language therapist. Diagnosing and monitoring of the treatments provided for the patient have to be monitored, and the required adjustments will be performed by the therapist. It would be very useful to get the assistance of an intelligent tool when providing treatments for these patients. Due to the inadequacy of the knowledge gained by the current researches, it is difficult to design a fully-fledged computer-based decision tool.

The literature survey is the initial step in this regard, and it has been organized under four main headings.

1. Literature survey to recognize the relevant factors of digital signal processing
2. Literature survey to recognize the treatment methods for voice disorders
3. Literature review to identify the important relevant factors in linguistics studies which have a direct impact to start this kind of research.
4. Literature survey to identify the relevant areas in computer science to write computer programs for the research.

2.1 Literature survey to recognize the relevant factors of digital signal processing

Digital signal processing (DSP) is the mathematical handling of an information signal to change it in some way. These signals characterized by the illustration of discrete time, frequency or discrete domain signals by a sequence of numbers or characters and the processing of these information signals.

The objective of digital signal processing is generally to measure, filter or compress real-time continuous analog signals. The first step is to convert the analog signal to digital signal form, sampling and then digitizing it with an analog-to-digital converter, which converts the analog signal into a flow of numbers. The required output signal is another analog signal, which requires a digital to analog converter. Digital signal processing is more complex than the analog signal processing, and digital signal processing has a range of discrete values, but the application of computational power to DSP offers many advantages over analog processing, such as the detection and correction of errors in the transmission and data compression.

Digital Signal process is one amongst the foremost powerful technologies that may form science and engineering within the ordinal century. Changes have already been created during a broad vary of field: communications, medical imaging, measuring instrument, hi-fi music reproduction, and oil prospecting. Those areas have developed using deep DSP technology, with its own algorithms, arithmetic, and specialized techniques [26].

2.1.1 Digital Signal Processing and Voice Disorders

Pathological voice recognition has been received good attention from researchers within the last decade. Speech process has tried to be a wonderful tool for voice disorder detection. There are three systems contributors to the assembly of the speech: the respiratory system, the laryngeal system and also the supra-laryngeal system (the articulators). The system nervous additionally controls the prosody. This one schematically covers the variations of height (intonation, melody), the variations of intensity (accentuation) and also the temporal progress (pauses, debit, and rhythm) [17].

The analysis of the voice disorder stays basically clinical. The instrumental measures square measure spilled very little in follow clinic. the foremost used square measure the acoustic and mechanics measures. The speech analysis is complicated and has been forgotten for an extended time. an issue results to analyze within the literature the various treatment result (medical or surgical). Indeed, several studies do not come back a specific analysis of the speech [1].

Features assessment of a voice disorder is that the disorder carries on a patient's capability to speak are an important step to conceive a program of its management. A method of the prosperous assessment permits the diagnostician of the speech to diagnose the voice disorder, confirm the relative potency of many treatment approaches and formulate a prognosis. Physicians usually use invasive techniques like an examination to diagnose the symptoms of vocal fold disorders. However, it's attainable to identify disorders using features of speech signals.

Digital signal process typically approaches the matter of voice disorders in 2 steps: feature extraction followed by feature matching every word within the incoming audio signal analyzed to spot its options, and these parameters are then compared with the keep knowledge.

2.1.2 Feature extraction and classification

Feature extraction is a method of reworking input information into reduced illustration set of informative features that involve the speech signal process technique. For the aim of speech recognition, speech samples are divided into frames, and features are extracted from every frame. Totally different classic techniques are used to extract the vocal parameters therefore on create the classification of the pathological voices like pitch and formant detection [17]. During feature extraction, speech features area unit extracted into a sequence of feature vectors so as to be classified in the classification stage. Before the feature extraction method takes place, three basic steps of pre-processing, i.e. pre-emphasis, framing and windowing are executed. The primary feature extraction utilized in this analysis is Mel Frequency Cepstral constant (MFCC). This procedure produces twelve mel cepstrum coefficients, one log energy co-efficient, their delta and delta-delta coefficients [1].

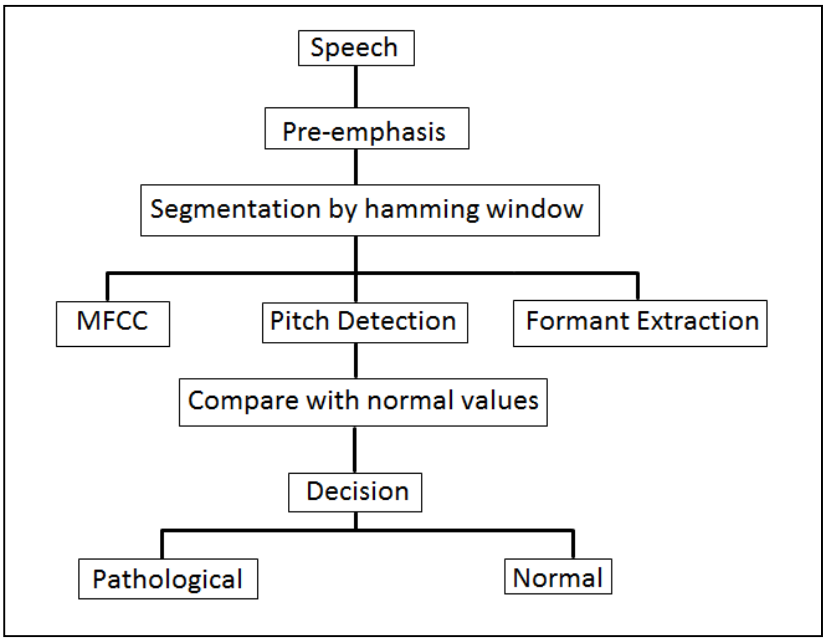


Figure 2-1: Speech processing algorithm

2.1.3 Mel Frequency Cepstral Coefficient (MFCC)

In sound process, the mel-frequency cepstrum (MFC) could be illustration of the short power spectrum of a sound, supported a linear cosine function of a log power spectrum on a nonlinear mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that jointly conjure associate MFC. They're derived from a sort of cepstral illustration of the sound clip. The distinction between the cepstrum and therefore the mel-frequency cepstrum is that within the MFC, the frequency bands are equally spaced on the mel scale, that approximates the human sense modality system's response additional closely than the linearly-spaced frequency bands utilized in the traditional cepstrum. This frequency distortion will provide a higher illustration of sound, as an example, in audio compression.

MFCCs are usually derived as follows [14]:

1. Get the Fourier transform of a signal.
2. Map the powers of the spectrum obtained higher than onto the mel scale, exploitation triangular overlapping windows.
3. Take the logs of the powers at every of the Mel frequencies.
4. Get the discrete cosine transform of the list of Mel log powers, as if it were an indication.
5. The MFCCs are the amplitudes of the ensuing spectrum.

MFCCs are usually used as features in speech recognition systems, like the systems which may automatically acknowledge numbers spoken into a telephone. They're conjointly common in speaker recognition, that is that the task of recognizing individuals from their voices.

MFCCs also are progressively finding uses in music data retrieval applications like genre classification, audio similarity measures.

2.1.4 Pitch Analysis

A pitch detection algorithmic is associate to estimate the pitch or fundamental of a quasiperiodic or virtually periodic signal, typically a recording of speech or musical note. This may be tired of the time domain or the frequency domain or each the 2 domains.

In the time domain, a pitch detection algorithmic usually estimates the amount of a quasiperiodic signal, then inverts that worth to relinquish the frequency.

One straightforward approach would be to measure the gap between the zero crossing points of the signal. However, this doesn't work well with that area unit composed of multiple sine functions with differing periods. Nonetheless, there are cases within which zero-crossing are often a helpful measure, e.g., in some speech applications wherever one supply is assumed. The algorithm's simplicity makes it "cheap" to implement.

More refined approaches compare segments of the signal with alternative segments offset by a shot amount to seek out a match. AMDF (average magnitude distinction function), ASMDF (Average square Mean distinction Function), and alternative similar autocorrelation algorithms work this manner. These algorithms will offer quite correct results for extremely periodic signals. However, they need false detection issues, and might typically cope badly with clamorous signals (depending on the implementation), and in their basic implementations don't deal well with polyphonic sounds (which involve multiple musical notes of various pitches).

Current time-domain pitch detector algorithms tend to create upon the fundamental strategies mentioned on top of, with further refinements to bring the performance additional in line with a person's assessment of pitch. for instance, the algorithmic principle rule and therefore the MPM algorithmic rule are each primarily based upon autocorrelation.

2.1.5 Formant Frequency

Formants are the distinctive or meaty frequency parts of human speech and of singing. By definition, the data that humans need to differentiate between vowels will be diagrammatical strictly quantitatively by the frequency content of the vowel sounds. In the speech, these are the characteristic partials that determine vowels to the attender. Most of those formants are made by tube and chamber resonance. The formant with the bottom frequency is named f1, the second f2, and therefore the third f3. most frequently the 2 initial formants, f1 and f2, are enough to clear up the vowel. These 2 formants verify the standard of vowels in terms of the open/close and front/back dimensions, therefore, the primary formant f1 features a higher frequency for an open vowel and a lower frequency for a detailed vowel and therefore the second formant f2 features a higher frequency for a front vowel, and a lower frequency for a back-vowel Vowels can nearly always have four or additional distinguishable formants; generally, there are quite six. However, the primary 2 formants are most vital in determinant vowel quality, and usually, this can be often displayed in terms of a plot of the primary formant against the second formant, although this can be not decent to capture some aspects of vowel quality, like rounding.

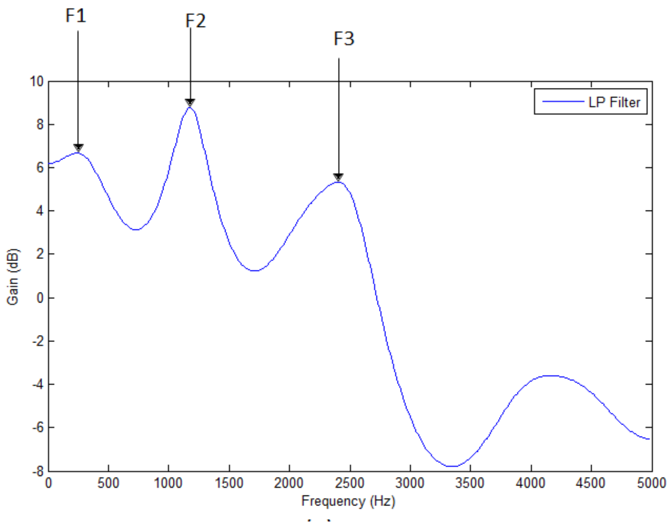


Figure 2-2: First Three Formant Frequencies

2.2 Literature survey to recognize the treatment methods for voice disorders

Voice disorders occur not because of the disorder of the nerves or muscles. Medical experts believe it is a problem of coordination and movements of larynx, lips, tongue, and soft palate. The main feature of this drawback is difficulty in making individual speech sounds and in sequencing sounds along in words. The speech characteristics of patients are;

- Limited range of vowels and consonant
- Overuse of one sound
- Vowel distortions
- The breakdown in sequencing in words, particularly as length increase
- Errors of omission and substitution
- Glottal stop insertions and substitutions
- Voice difficulties affecting volume, length, pitch and quality
- Resonance difficulties are affecting the overall tone of the speech [25].

Speech-language therapists use totally different medical care ways to treat children with the disorder. One common method is the Nuffield Dyspraxia Program is widely used by the therapists in overseas. It offers a systematic approach with numerous exercises, and the progress of the treatments are also monitored. These exercise programs include exercises for lips, tongue, soft palate, larynx, and breathe control [25].

Speech-language therapists play an impotent role in diagnosing voice disorders. There is no single aspect or test that can be used to identify voice disorders. Investigations proceed with the examinations of one or more symptoms and usually ask the patient to pronounce some words. Same words are repeatedly tested and study the improvements. However, these treatments have to be proceeding for a long period. Therefore, initial investigations are important to diagnose voice disorders. Brain imaging test, for example, Magnetic Resonance Imaging (MRI) can be used to help distinguish the voice disorders. [28]

Speech therapists use different methods to treat voice disorders. No single method has been established to be the most effective. However, most of the treatment plans include different types of exercises for the larynx and vocal tract. It is difficult to monitor progress of treatment plans. [28]

The two researchers have addressed same issues. The treatment plans, diagnosing difficulties and progress monitoring are the most important issues. The development of an intelligent tool would be very useful in identification and progress monitoring process.

The majority of children's voice disorders are identified by the parents or their teachers. Sometimes these difficulties are not considered for further investigations and this would be the main reason that the problem to become worse. It is recommended to consult a medical expert in this area prior to consult a speech-language therapist. The medical expert performs initial investigations and he directs the patient for voice therapy actually the patient need that. The Fluharty-2 Preschool Speech-Language Screening Test is one of the common screen tests are performed speech-language therapists. It is a preliminary level investigation and a generalized comment like "sound normal; recheck may be necessary" will be the output of this test. This conventional one-line comment fails to address the required parameters of a voice. A comprehensive comment based on the voice-related subsystems of respiration, phonation, and resonance may be more useful for further investigations. [14]

These researches addressed the same issue and attempted to explain the importance of developing an intelligent tool.

2.3 Literature review to identify the important relevant factors in linguistics studies which have a direct impact to start this kind of research

It is significant to recognize the basic distinction between spoken and written language. Both spoken and written languages have syllables, vowels and consonants. Speech and writing are separate channels through which language messages can be passed. Persons normally speak when breathing out. Air passing outside to the lungs must go through the larynx, where its flow is controlled by the vocal folds. Voicing makes a distinction between consonants such as voiced and voice less and it is utilized in many languages of the world. Articulation is another aspect of voice production. Articulation is sound produced by the vocal tract; these are generally voiceless. Voice sounds have complex periodic waveforms, and voicing is the only source of periodic energy in speech. Persons having voice disorders have certain problems of these wave forms. [19]

2.4 Literature survey to identify the relevant areas in computer science to write computer programs for the research.

Early identification of the system requirement is an important issue when writing computer programs. It is important to identify the functional non-functional requirements. Functional needs for a system describe the practicality or services that the system is anticipated to supply. Non-functional needs are those needs that don't seem to be directly involved with the precise operate delivered by the system. They may be relevant to reliability, response time, and about the storage. For the documentation purposes, different types of methods are used by the systems designers. Most of these methods are followed by the styles used in programming languages. For example, PDL (Program Description Language) is one method used to describe the details about a computer system. The PDL is a language derived from a programming language like Java or Ada. This would not be an easy task due to the difficulties of the syntaxes of PDL. The Use Case Diagram is the method described in UML (Unified Modeling Language), is used by the most of the designers to identify and describe user requirements. The UML (Unified Modeling Language) is a universally accepted method for designing object-oriented programming applications. The Use Case Diagrams can also be used as a technique for the requirement identification process. [7]

Verification and validation is another important aspect of computer programming. Testing is the main process for the identification of verifications. There are 2 types of testing techniques; black box and white box testing. Black box testing involves the correctness of the output produced by the program. White-box involves the testing of internal program structures. [7]

System implementation is bringing the system into real-world application. When it is going to be done, the designer has to examine, how his perspective models will directly be implemented or any prior adjustments have to be applied for perspective models. Programming is generally recognized as the major aspect of system construction. Majority of the users assume programming is the only task of the system development and once it is completed the product can be directly supplied to the market. There may be some bugs in the systems that may have a certain impact on the final product, and finally, it would produce some errors in the output or result in malfunctioning of the product. Before supplying the product, its quality has to be verified, and a guarantee has to be provided. This guarantee would be a quality assurance about the operations, output and security of the product. [21]

Testing is an important task in software development. It is not differed until all programs are finished. It has to be performed when individual modules are developed. There are three different levels of testing that have to be performed when developing software; stub testing, program or unit testing and systems testing. Stub testing is performed for individual modules of the software. All program stubs are tested as integrated modules in program or unit testing. System testing is performed on an entire system. The proper functionality of an individual program does not mean the entire system works properly. The integrated test shows the correctness of the whole system in terms of inputs and outputs. It would be helpful for the end users to recognize the system's inputs and recommended output formats suggested by the system designers. All end users too worried always about the final output. Therefore the system testing has to be performed. The correctness of the program stubs recommends the modes of parameter passing methods and the correctness of internal control structures of program modules. The developed test database will be used for this purpose. [21]

The two authors have explained the importance of the process of verification and validation. Two different methods have been explained by the two authors. The concept of stub testing, program and unit testing directly relate with the white-box testing.

3 Technologies Used for Tool

3.1 Digital Signal Processing

Speech recognition systems mostly used Digital Signal Processing techniques. DSP is the mathematical operation of an information signal in terms of identification of the illustration of discrete time, discrete frequency or other discrete domains signals by a sequence of numbers or symbols and processing of information signals. The first step of digital signal processing is converting an analog signal into digital format. In which a digital-to-analog converter is required. Voice recognition systems are based on the analysis performed on the phonetically related evaluations.

The algorithms discussed in Digital Signal Processing (DSP) are often used to examine the parameters of a digital signal. Since the concepts of digital signal processing are going to be applied for an analog signal, it has to be transformed into a digital signal first. Speech is a complex form of an analog signal. Study pertinent to speech recognition is also transformed the analog signal into digital and further applied algorithms described in digital signal processing. In this regard, the researcher has used the pitch frequency analysis, the formant frequency analysis, and the MFCC analysis described in digital signal processing.

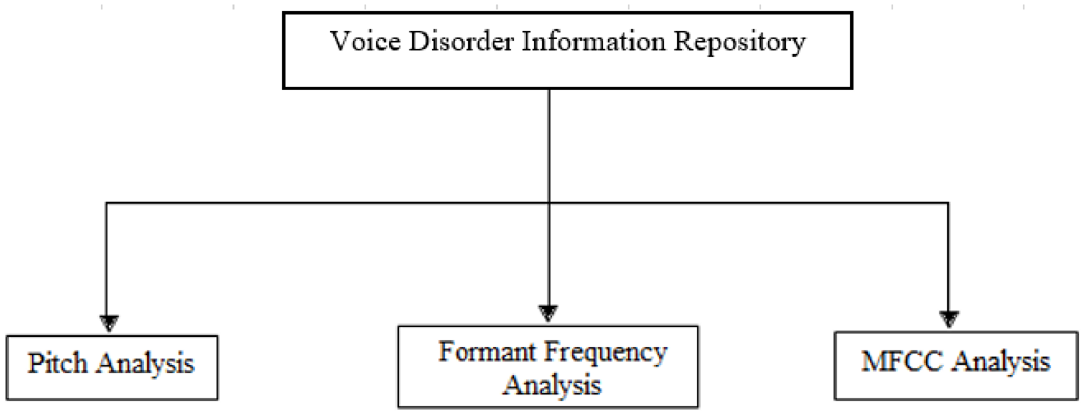


Figure 3-1:Domain-Specific Architectural Model for the proposed automatic tool

3.2 MATLAB

MATLAB and signal process merchandise facilitate to investigate signals from a variety of information sources. MATLAB will use to acquire, measure, transform, filter, and visualize signals

- Preprocess and filter signals before analysis.

- Explore and extract options for information analytics and machine learning applications.
- Analyze trends and find out patterns in signals.
- Visualize and measure time and frequency characteristics of signals.

MathWorks provides style apps, DSP algorithm libraries, and I/O interfaces for real time processing of streaming signals in MATLAB and Simulink. MATLAB will quickly style and simulate streaming algorithms for audio, video, instrumentation, sensible sensors, wearable devices, and alternative electronic systems.

Chapter 4

4 Analysis and Design

4.1 Trace the types of voice disorders in children in Sri Lanka whose age is less than seven years.

Voice is produced by the vocal folds in the larynx. Vocal folds are thin muscle bands in the lower part of the mouth also known as larynx. When air is passing in and out through these muscle bands voice is produced. A voice disorder is a dissimilarity of a voice comparative to a voice produced by an ordinary person in the same age and same sex. Children can have different types of voice disorders, but most of them are harmless and are caused by excessive use of harsh voice, but some of them have to be treated. (Royal Children Medical Hospital Melbourne, 2010)

Voice disorders are generally divided into three main categories: organic, functional and combination of the two. It has to be treated at the early stage i.e. before 7 years of the patient to achieve the best results. Medical experts believe no medicine to treat voice disorders, but properly managed exercises can be used to treat these patients. Usually, doctors advice these patients to get treatments from Speech and Language Therapists (SLT).

It is an important issue to get treatments at early childhood because this affects the child's communication and learning abilities. Early diagnostic tools in current use are limited to diagnose this disease.

4.2 Identify the special attributes and treatment methods for common voice disorders

The vibration of vocal folds in the larynx is the main reason to generate human voice. The three main characteristics: pitch, intensity, and quality can be used as parameters to evaluate the quality of a sound. Pitch is generally known as the tone of the voice. Inappropriate pitch patterns like low or high are the root cause of a monotonous sound. In addition to it, repeated pitch patterns are also a reason for a monotonous sound.

Loudness is the volume or intensity of a voice. Due to hearing loss or a structural defect of vocal cords a person may speak unnecessary louder than normal. These may be the root causes to have an unnecessary louder voice. Due to these reasons some patients are reluctant to talk in public.

The quality of the sound directly relates to the functionality of the vocal cords. Breathing problems may have a certain problem to vibrate of vocal cords. Sometimes it produces a nasal sound or voice resonates inside the mouth.

However, most voice disorders based due to a problem of one parameter; voice dyspraxia is a combination of all. According to a paper published recently by Williams Pams(2011) voice dyspraxia is a combination of problems in pitch, quality, length, the volume of the sounds. The other characteristics of voice dyspraxia patients have limited range of consonant and vowel, misuse of one sound, vowel distortion, failure in sequencing in words, errors of omission of certain sounds, resonance difficulties of overall tone of speech and unintelligible speech.

It is recommended to consult an ENT specialist or a general practitioner to discover whether the patient has a voice disorder. He then directs the patient to a Speech and Language Therapist (SLT). Most of the government hospitals in Sri Lanka have a separate treatment unit to provide treatments for voice disorder patients. It is an important requirement to provide the treatment as early as possible to obtain successful results. The speech-language therapist initially accesses to recognize the present difficulties of the patient and describes which type of treatment procedure is suitable for successful results.

After the initial assessment, a treatment plan will be designed by the SLT for the patient. An internationally accepted treatment procedure called Nuffield Dyspraxia is used by most of the SLTs to treat patients. It includes different types of exercises for the lips, tongue, soft palate, larynx of the patient. The treatments have to be continued for a certain period, and the progress of the patient has to be monitored from time to time. It is important to measure the progress of the patient, and certain adjustments will be applied for the treatment plan by the SLT. Nevertheless, the monitoring procedure has to be performed based on the evaluation of parameters of the patient's voice. It is performed by listing to the patient's voice.

4.3 Importance of making a phonetic balance

Phonetic science is concerned with the objective, description, and analysis of all aspects of speech. The speech is a sound generated by a human with the intention of expressing a certain idea. The sounds in speech are two kinds: vowels and consonants. Speech sounds grouped into syllables, which are the smallest units that can be pronounced in a natural way. Recording a sound using a microphone can be used to further analyze its physical properties by taking various measurements. Speech and writing use the same syllables, but the brain uses two channel to produce it. These two channels have different uses and different characteristics. Writing has a direct relationship with our culture but avoids mistakes by following grammatical rules which enhance the readability.

Humans produce speech when passing out air through the larynx, the vocal folds in the larynx vibrates when the air is going out. Sounds produced by the vocal folds are called voiced when the vocal folds are opened voiceless sounds are generated. To generate a voiceless sound, the vocal tract is used. Articulation is another term used to express the sounds generated by vocal tract. Most of the languages vowels are voiced sounds; consonants are voiceless. Voiced sounds have complex periodic waveforms, and voicing is the only source of periodic energy in speech.

The core areas of pronunciation for the consonants are the lips, upper front teeth, soft palate, alveolar ridge, hard palate, and uvula. The involvement of the tongue is different, and it moves to five different places; the tip, blade, front, back and root when producing consonants. Some instrumental techniques like X-rays and MRI can be used to identify any problems of producing articulations. The problems of articulation are more complex than to the sounds produced by the vocal folds. Expert comments are very important to identify the real problem of sound generation, and he will direct the patient for speech therapy.

Speech therapy includes nonsurgical techniques to improve the human voice. Speech and Language Therapists (SLT) are formally trained persons to provide speech therapy for patients. Speech therapy involves improvements for articulations, improving fluency disorders, improving voice disorders (e.g. unnecessary loudness of speech) and, improving language disorder (missing words when speaking). Voice therapy includes different types of exercises for the larynx, lips, tongue, soft palate and breath control. SLTs use list of words and usually ask the patient to pronounce the list and record the sound. At different time intervals same list is used to pronounce and study the improvements of the speech. Choosing the words for the list is important issue because the language efficacy is depend on the age, the environment and the gender. This list of words is known as ‘phonetic balance’ and it consists of proper mixture of vowels and consonants of words. For this regard, linguistics expert advises have been obtained to choose a proper phonetic balance. Refer appendix A for further information. The phonetic balance consists of 10 words including a variation of two to six character words. Refer Table 1 for the phonetic balance.

Table 4-1:Phonetic Balance for ten words

Word	How to pronounce in English	No of vowels	No of consents
ගස්	Gas	1	2
ලේන්	Len	1	2
පොත්	Poth	1	2
අළුත්	Aluth	2	2
දෙන්නේ	Denne	2	3
දෙකටක්	Dekatak	3	4
කුඩුවක්	Kuduvak	3	4
පන්තිය	Panthiya	3	4
හිටවන්න	Hitavanna	4	5
මුකුණුවැන්න	mukunuvanna	5	6

Majority of Sri Lankans’ native language is Sinhala. Sinhala is an Indo-Aryan language spoken primarily in Sri Lanka. According to statistics from 2018, approximately 75 percent of the population of Sri Lanka is native Sinhala speakers. Therefore, Sinhala is only concerned as a language for the phonetic balance.

4.4 Conceptual view of the architectural design of proposed automated tool

The concept of the architectural design is used to identify the major components of the system and the communication of each other. It can also be used as a preliminary tool to commence the analysis of components. Further, it is also be used to describe the stakeholder communications. The architecture is a high level arrangement of the system, may be used as a emphasis for discussion by a range of different stake holders. An industry expert has been interviewed to recognize the significant features, requirements, and stake holder communication of the automated tool. Refer appendix A for further information.

Speech recognition systems mostly used Digital Signal Processing (DSP) techniques. The first step of digital signal processing is converting an analogue signal into digital format. In which a digital-to-analog converter is required. Voice recognition systems are based on the analysis performed on the phonetic related evaluations.

The algorithms discussed in Digital Signal Processing (DSP) are often used to examine the parameters of a digital signal. Since the concepts of digital signal processing are going to be applied for an analog signal, it has to be transformed into a digital signal first. Speech is a complex form of an analog signal. Study pertinent to speech recognition is also to transform the analog signal into digital and further apply algorithms described in digital signal processing. In this regard, the researcher has used the pitch frequency analysis, the formant frequency analysis, and the MFCC analysis described in digital signal processing.

The finalized domain specific architectural model for the proposed automatic tool is specified below (figure 5). It consists of three unique models, which perform different calculations and analysis based on digital signal processing methods.

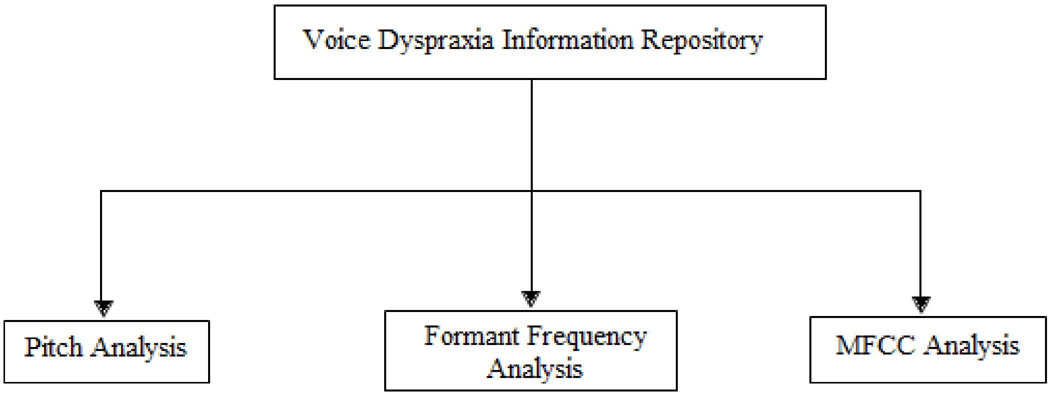


Figure 4-1: Domain Specific Architectural Model for the proposed automatic tool

A number of researchers have proposed the use of architectural design to describe internal steps of the system. They used the term Architectural Description Language (ADL) for such type of internal descriptions. Unified Modeling Language (UML) is frequently used for architectural descriptions. (pp. 217, Sommerille, 2001). The UML notations and required narrations are included in subsection 4.6.

4.5 Identify the functional and non-functional requirements of proposed automated tool.

Software system requirements are often classified as functional and non-functional requirements.

4.5.1 Functional Requirements

The functionality or the services expected to be provided are classified under functional requirements. Use Case diagrams are widely used to recognize software functions of a system. In this project, Use Case diagrams have been used to identify the functional requirements of the software.

Use case diagram is an analysis tool for fact finding and identifying of events and responses of the system. Person or an external system that interact with the system is called 'actors' in Use Case diagrams.

Three use case diagrams have been developed for each component in the domain specific architectural model, i.e. Use Case Diagram for Pitch Analysis, Use Case Diagram for Formant Frequency Analysis, and Use Case Diagram for Mel Frequency Analysis.

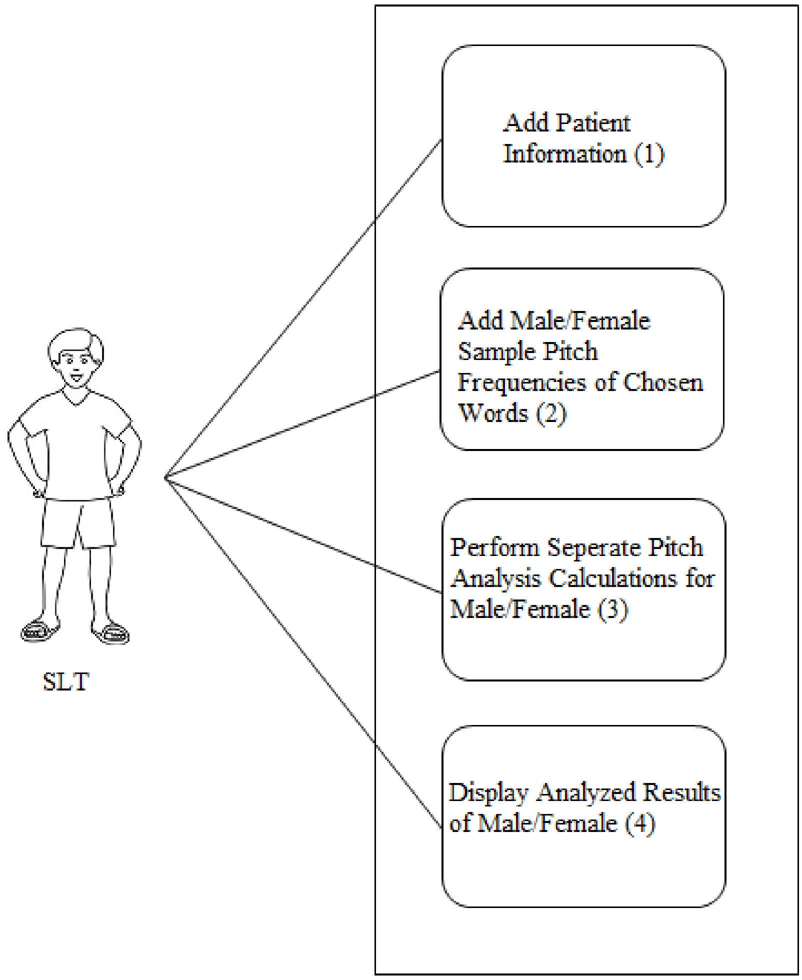


Figure 4-2: Use Case Diagram for Pitch Analysis

The requirements list of Pitch Analysis module is mentioned in Table 2 with a column to demonstrate which use cases provide the functionality of requirement.

Table 4-2: Requirements List for Male & Female Pitch Analysis

No.	**Requirement**	**Use Case (Figure6)**
1	To record name, age, gender and pitch frequencies of 10 words of the patient	(1)
2	To record sample pitch frequencies of ordinary children. Frequencies of teb words are included to the sample.	(2)
3	To perform pitch analysis of male/female patients with sample data. (Refer algorithm 1 of detailed design)	(3)
4	To obtain a report on performed pitch analysis	(4)

The requirements list of Formant Frequency Analysis module is mentioned in Table 3 with a column to illustrate which use case provide the functionality of requirement.

Table 4-3: Requirements List for Male & Female Formant Analysis

No	**Requirement**	**Use Case (Figure7)**
5	To record name, age, gender and formant frequency analysis of 10 words of the patient	(5)
6	To record sample formant frequency analysis of ordinary children.	(6)
7	To perform formant frequency analysis of male/female patients with sample data. (Refer algorithm 2 of detailed design)	(7)
8	To obtain a report on performed Formant analysis	(8)

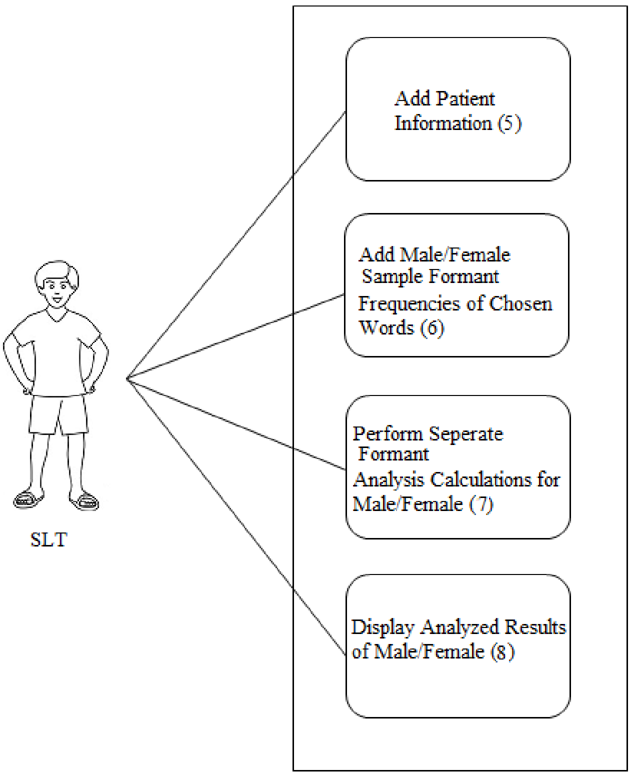


Figure 4-3:Use Case Diagram for Formant Frequency Analysis

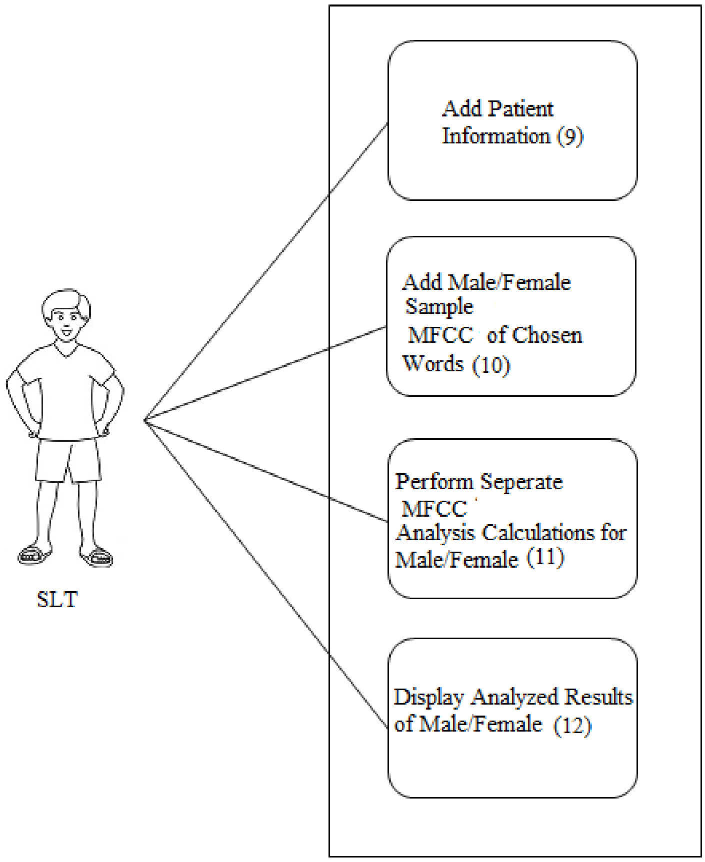


Figure 4-4: Use Case Diagram for Mel Frequency Analysis

The requirements list of Mel Frequency Analysis module is mentioned in Table 4 with a column to illustrate which use case provide the functionality of requirement.

Table 4-4: Requirements List for Male & Female Mel Frequency Analysis

No.	Requirement	Use Case (Figure8)
9	To record name, age, gender and Mel frequency analysis of 10 words of the patient	(9)
10	To record sample Mel frequency analysis of ordinary children.	(10)
11	To perform Mel frequency analysis of male/female patients with sample data. (Refer algorithm 3 of detailed design)	(11)
12	To obtain a report on performed Mel frequency analysis	(12)

4.5.2 Non- Functional Requirements

Non-functional requirements are those used to support implementation of functional requirements. In this regard performance, security and management of data volumes are considered. Non-functional requirements have been identified and mentioned in Table 5.

Table 4-5: Non-Functional Requirements

No.	Category	Requirement
1	Performance	System must support to implement on desktop computers which run on Windows 7, Windows 8 & Windows 10 operating System. The memory size is minimum 2GB. There is no need to connect the system with a network. System must support for individuals and appropriate GUI facilities must be included to operate.
2	Security	Data validation routines must be included to all data entry screens. Password protection is required for the database. Patients' history files must be maintained. Backup facilities must be included.
3	Data management	Required analysis (i.e. pitch analysis, formant frequency analysis, and Mel frequency analysis) of patients' inputs must be performed using MATLAB before transferring to Voice disorder Information Repository.

4.5.3 Summary of Requirements

4.5.3.1 Functional Requirements

1. To record name, age, gender and pitch frequencies of 10 words of the patient.
2. To record sample pitch frequencies of ordinary children. Frequencies of ten words are included to the sample.
3. To perform pitch analysis of male/female patients with sample data. (Refer algorithm 1 of detailed design)
4. To obtain a report on performed pitch analysis
5. To record name, age, gender and formant frequency analysis of 10 words of the patient
6. To record sample formant frequency analysis of ordinary children.

7. To perform formant frequency analysis of male/female patients with sample data. (Refer algorithm 2 of detailed design)
8. To obtain a report on performed Formant Frequency analysis
9. To record name, age, gender and Mel frequency analysis of 10 words of the patient.
10. To record sample Mel frequency analysis of ordinary children.
11. To perform Mel frequency analysis of male/female patients with sample data. (Refer algorithm 3 of detailed design)
12. To obtain a report on performed Mel Frequency

4.5.3.2 Non Functional Requirements

1. System must support to implement on desktop computers which run on Windows 7,8,10 operating System. The memory size is minimum 2GB. There is no need to connect the system with a network. System must support for individuals and appropriate GUI facilities must be included to operate.
2. Data validation routines must be included to all data entry screens. Password protection is required for the database. Patients' history files must be maintained. Backup facilities must be included.
3. Required analysis (i.e. pitch analysis, formant frequency analysis, and Mel frequency analysis) of patients' inputs must be performed using MATLAB before transferring to Voice disorder Information Repository.

4.6 Detailed Design Specification of Proposed Automated Tool

At this stage various decisions have to be made regarding the design specifications. Possible program components of the proposed automated tool are given in figure 9. The requirements recognized for the proposed system has been further refined to identify the program components of the proposed system. This architecture has two layers. The bottom layer has been designed to store relevant data of ordinary children and patients. The two upper layers i.e. Voice Input and Patient's Data Analysis modules have been designed to support for voice data inputs and required data analysis of patients. Choosing programming languages for the individual component development are another import task. MATLAB has been chosen for Voice Input module. It is easy to use MATLAB for DSP based analysis such as pitch analysis, formant analysis and Mel frequency analysis.

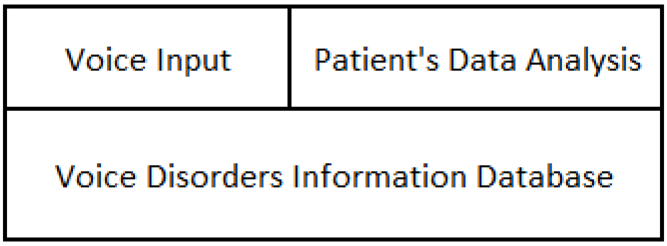


Figure 4-5: Program Components of Proposed Automated Tool

It is difficult to understand a complex program module when viewed as whole. Therefore, dividing it into separate sub modules is important I systems analysis. The detailed design specification of each program component is given below.

4.6.1 Detailed Design Specification of Voice Input Component

The DSP analysis is included to this module. It consists of Pitch Analysis, Formant Analysis and Mel Frequency Analysis and its programs. The algorithms of each analysis are given below. All program codes have been developed using MATLAB. Figure 10 depicts the detail modules of Voice input components.

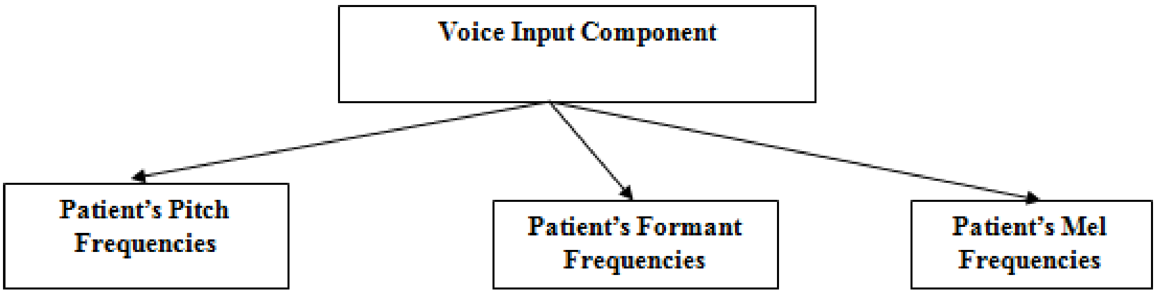


Figure 4-6: Program modules of Voice Input Component

Algorithm of each software module is given below.

4.6.1.1 Algorithm of Pitch Analysis

Pitch is obtained by getting the highest of autocorrelation sometimes the initial speech file is divided into frames and inflection is derived by plot of peaks from frames.

The autocorrelation technique may be a methodology for estimating the dominating frequency during a advanced signal, in addition as its variance. Specifically, it calculates the primary 2 moments of the ability spectrum, specifically the mean and variance.

Several works have enforced pitch extraction algorithms supported computing the short-time autocorrelation perform of the speech signal. First, the speech is generally low-passed filtered at a frequency of regarding 1kHz, that is well higher than the utmost anticipated frequency vary for pitch. Filtering helps to scale back the results of the upper formats and any extraneous high-frequency noise. The signal is windowed using an suitable soft window (such as Hamming) of duration 20 to 30 ms and a typical autocorrelation function is given by,

$$R(k) = \sum_{n=-\infty}^{\infty} x[n].\, x[n+k]$$

The autocorrelation function provides a measure of the correlation of a signal with a delayed copy of itself. In the case of voiced speech, the main peak in short-time autocorrelation function normally occurs at a lag equal to the pitch-period. This peak is therefore detected and its time position gives the pitch period of the input speech.

To calculate the pitch,

- First Low pass filtered speech signal
- Use a 30msec segment, choose a segment every 20msec that means the overlap between segments is 10msec.
- Calculate the pitch from every segmented frame
- Pitch calculate using the autocorrelation technique

<u>LPF at 900Hz</u>

```
[xf0, yf0] = butter(4, 900/(fs/2));
Aseg = filter(xf0, yf0, Aseg);
```

<u>Find the clipping level, C</u>

```
l13 = len/3;
max1 = max(abs(Aseg(1:l13)));
```

```
l23 = 2 * len/3;
max2 = max(abs(Aseg(l23:len)));
if max1>max2
C=max2;
else
C= max1;
```

Center clip waveform, and compute the autocorrelation

```
clp = zero(len,1);
ind1 =find(Aseg>=C);
clp(ind1) = Aseg(ind1) -C;
ind2 = find(Aseg <= -C);
clp(ind2) = Aseg(ind2)+C;
en = norm(clp,2)^2;
RR = xcorr(clp);
n = len;
```

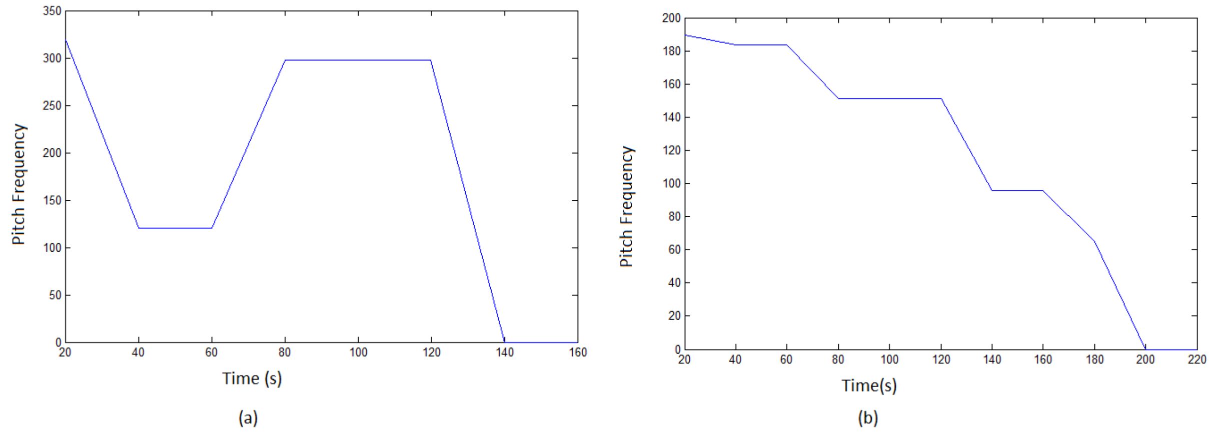


Figure 4-7: Pitch values of (a) normal word “Gas”(ගෑස්) and (b)Pathological word “Gas”(ගෑස්)

Figure 11 shows the pitch variation by application of the autocorrelation technique analysis of ordinary and pathological male sounds. The high distortion and the difference of the pitch around the expected value reveal a state of the glottis signal anomaly, resulting of a laryngeal pathology.

4.6.1.2 Algorithm of Formant Frequency Analysis

A formant frequency is identifying or significant frequency element of human speech. it's the characteristic harmonic that identifies vowels to the auditor. This follows from the definition that the data humans need to differentiate between vowels are often delineate strictly quantitatively by the frequency content of the vowel sounds. Therefore, formant frequencies are significant features and formant extraction is a very important characteristic of speech process.

Since ordinary children and pathological children have totally different formant positions for vowels, so formant positions is accustomed confirm the voice disorder of children. Therefore, excellence between ordinary and pathological can be described by the situation within the frequency domain of the primary three formants for vowels.

When using formant analysis for identification of voice disorder, the matter is largely counteracted to two components. The first part is formant extraction that uses a method that performs a detection of energy concentration rather than the classic peak choosing technique. The second part is that the normal or pathological detection supported the placement of the primary second and third formants.

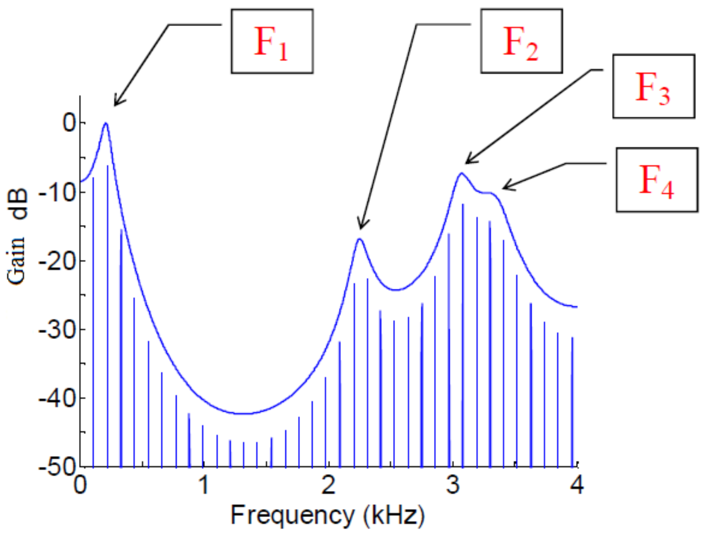
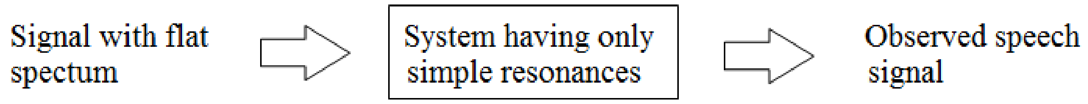


Figure 4-8: Formant frequency positions

Estimation of formant frequencies is usually harder than estimation of fundamental. the matter is that formant frequencies area unit properties of the vocal tract system and wish to be inferred from the speech signal instead of simply measured. The spectral form of the vocal tract excitation powerfully influences the determined spectral envelope, specified we have a tendency to cannot guarantee that each one vocal tract resonances can cause peaks within the determined spectral envelope, or that each one peaks within the spectral envelope area unit caused by vocal tract resonances.

The dominant technique of formant frequency calculation is based on modeling the speech signal as if it were generated by a particular kind of source and filter:



This type of analysis is termed source-filter separation, and within the case of formant frequency estimation we have a tendency to have an interest solely within the modeled system and therefore the frequencies of its resonances. To search out the most effective matching system, have a tendency to use a technique of analysis known as Linear Prediction. Linear prediction models the signal as if it were generated by a signal of minimum energy being went through the purely-recursive IIR filter.

Demonstration of idea by using LPC to find the best IIR filter from a section of speech signal and then plotting the filter's frequency response.

[x,fs]=wavread('.wav');

fs=10000;

Calculate Linear prediction filter

nocoef=2+fs/1000;

x=lpc(x,nocoef);

Plot frequency response

[h1,f1]=freqz(1,x,512,fs);

plot(f,20*log10(abs(h1)+eps));

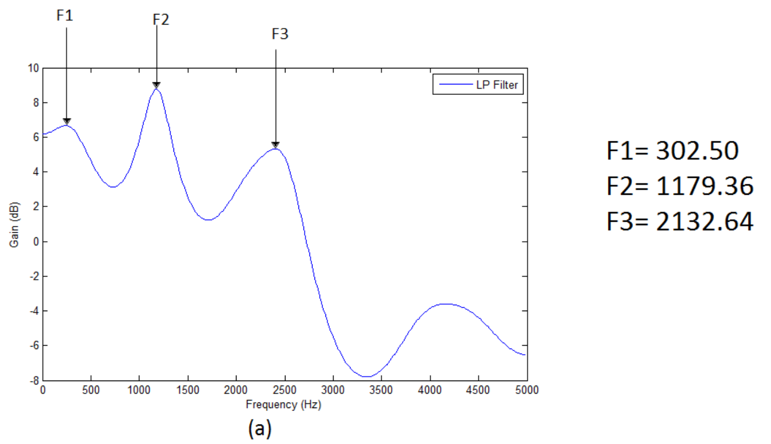
To calculate the formant frequencies from the filter, first need to calculate the locations of the resonances which make up the filter. This involves treating the filter coefficients as a polynomial and solving for the roots of the polynomial. Following code shows how to calculate estimated formant frequencies from the LP filter,

<u>Calculate frequencies by root-solving</u>

```
R1=roots(x);
R1=R1(imag(R1)>0.01);          % only look for roots >0Hz up to fs/2
freq=sort(atan2(imag(R1),real(R1))*fs/(2*pi));
```

<u>Convert to Hz and sort</u>

```
for j=1:length(freq)
fprintf('Formant %d Frequency %.1f\n',j,freq(j));
end
```



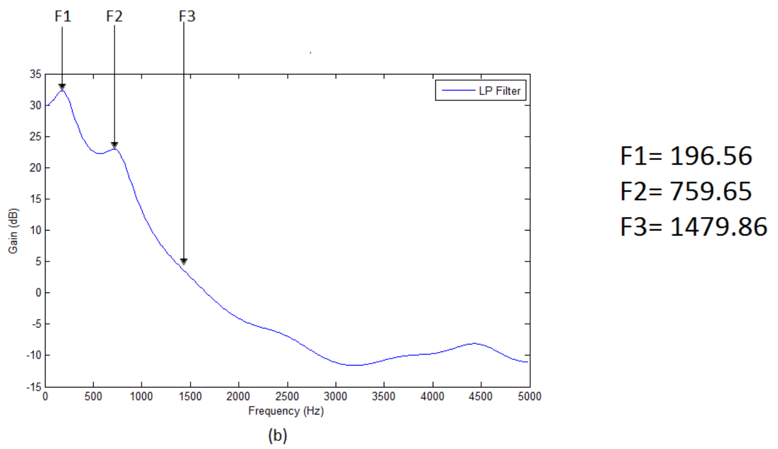


Figure 4-9: (a)First three Formant frequencies for normal word “Gas”(ගෑස්) and (b) Pathological word “Gas”(ගෑස්)

Table 4-6: First three formant frequency rangers for selected ten words

Word	F1 (Hz)	F2 (Hz)	F3 (Hz)
Gas (ගස්)	130- 330	270-1150	850-2600
Len (ලේන්)	75-150	500-2500	650-2500
Poth (පොත්)	130-250	250-1400	1200-2800
Aluth (අලුත්)	90-920	250-1200	950-2900
Denne (දෙන්නේ)	90-150	600-900	700-1400
Dekatak (දෙකටක්)	130-900	250-1400	800-2900
Kuduvak (කුඩුවක්)	30-350	150-1500	950-2800
Panthiya (පන්තිය)	130-350	200-900	500-1200
Hitavanna (හිටවන්න)	150-350	250-1000	200-1200
Mukunuvanna (මුකුණුවැන්න)	150-250	500-750	800-1500

4.6.1.3 Algorithm of Mel Frequency Cepstral Coefficients Analysis

The mel-frequency cepstrum (MFC) could be an illustration of the short term power spectrum of a sound, supported a cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) area unit coefficients that jointly constitution of MFC. MFCC is predicated on human hearing perceptions that cannot understand frequencies over 1Khz. In different words, in MFCC is predicated on identified variation of the human ear's essential information measure with frequency [8-10]. MFCC has two sorts of filter that spaced linearly at low frequency below a thousand Hertz and index spacing on top of 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture vital characteristic of phonetic in speech.

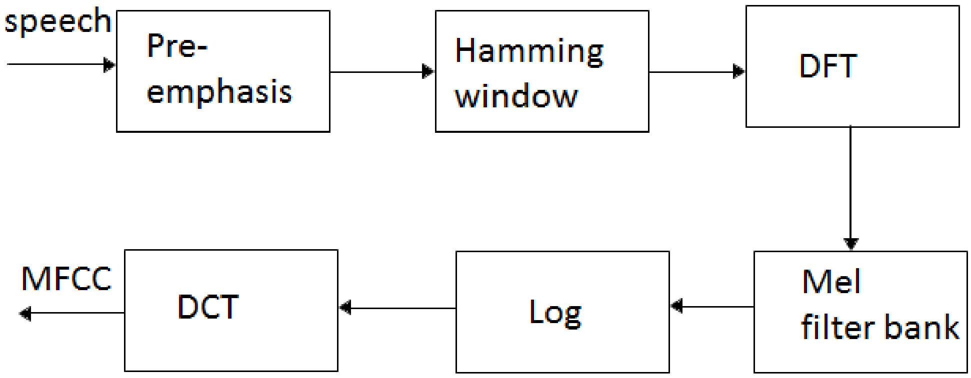


Figure 4-10:MFCC Block Diagram

As shown in Figure 14, MFCC consists of six computational steps. Each step has its function and mathematical approaches as discussed briefly in the following:

4.6.1.3.1 Pre – Emphasis

This stage, processes the passing of sound signal through a filter which highlights higher frequencies. Pre emphasis will increase the energy of sound signal at higher frequency.

The sound signal s(n) is sent to the a high-pass filter:

$$S_o(n) = s(n) - a \times s(n-1)$$

Where $S_o(n)$ is the output signal and the value of 'a' is generally between 0.9 and 1.0.

The Z transform of the filter is

$$H(Z) = 1 - a \times Z^{-1}$$

The aim of pre-emphasis is to compensate the high frequency half that was suppressed throughout the sound production mechanism of humans. Mover, it also can amplify the importance formants.

4.6.1.3.2 Hamming windowing

Hamming window is used as window shape by considering the next step in feature extraction processing chain and integrates all the neighbouring frequency lines. The Hamming window equation is as follow:

If the window is defined as HW(n), $0 \leq n \leq N-1$ where

$N = number\ of\ samples\ in\ each\ frame$

$Y[n] = output\ signal$

$S[n] = input\ signal$

$HW(n) = Hamming\ window$

Then the result of output signal is as follow:

$$Y(n) = S(n) \times W(n)$$

$$W(n, \alpha) = (1 - \alpha) - \alpha cos\left[\frac{2\pi n}{N - 1}\right]\ 0 \leq n \leq N - 1$$

In practice, the value of α is set to 0.46. MATLAB also provides the command *hamming* for generating the curve of a Hamming window.

4.6.1.3.3 Fast Fourier Transform (FFT)

Fast Fourier transform is used to convert every frame of N samples from time domain to frequency domain. FFT is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response

H[n] in the time domain. Following equation is support to above statement,

$$Y(w) = FFT[h(t) \times X(t)]$$

$$Y(w) = H(w) * X(w)$$

If $X(w), H(w), Y(w)$ are the Fourier Transform of $X(t), H(t)$ and $Y(t)$ respectively.

4.6.1.4 Mel Filter Bank Processing

Mel-frequency scale may be a perceptually actuated scale that is linear below 1 kilohertz, and logarithm above, with equal numbers of samples below above 1 kilohertz. It represents the pitch of a tone as a function of its acoustics frequency. One mel is described as 1000th of the pitch of a 1 kHz tone. Mel-scale frequency can be estimated using below equation.

$$B(f) = 2595 log_{10}\left(1 + \frac{f}{700}\right)$$

This non-linear transformation can be demonstrated in Figure 15. It described that equally spaced values on mel frequency scale correspond to non equally spaced frequencies. This is the inverse function of the above equation, which is given below equation,

$$f = 700\left(10^{\frac{f_{mel}}{2595}} - 1\right)$$

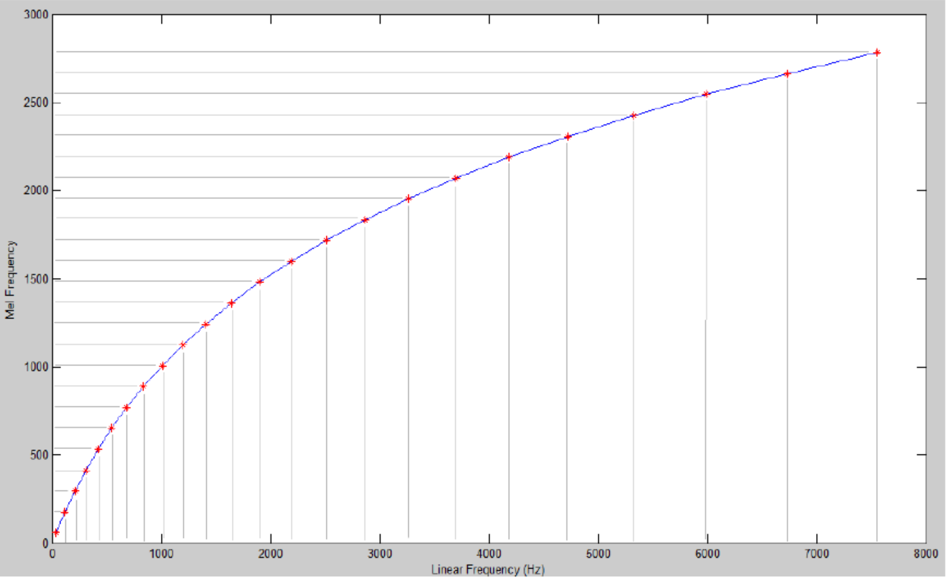


Figure 4-11: Mel-to-Linear Frequency scale transformation

So, it is hoped that mel scale more closely models the sensitivity of the human ear than a purely linear scale, and provides for greater discriminatory capability between speech segments.

4.6.1.4.1 Discrete Cosine Transform

Discrete Cosine Transform is used to convert the log Mel spectrum into time domain. The outcome of the process is called Mel Frequency Cepstrum Coefficient. The set of output coefficients are called acoustic vectors. Hence, every input utterance is transformed into a sequence of acoustic vector.

4.6.1.4.2 Delta Energy and Delta Spectrum

The speech signal and the frames changes, such as the slope of a formant at its transitions. Hence, it is required to add features related to the change in cepstral features over time. Thirteen delta or velocity features (twelve cepstral features plus energy), and thirty-nine features a double delta or acceleration feature were added. The energy in a frame for a signal 'S' in a window from time sample 'T1' to time sample 'T2', is represented at the equation below,

$$Energy = \Sigma S^2[T]$$

Every thirteen delta features signify the change between frames in the above equation corresponding cepstral or energy feature, while each of the thirty-nine double delta features represents the change between frames in the corresponding delta features

$$d(t) = \frac{c(t+1) - c(t+1)}{2}$$

There are two types of speech signal patterns have been found which are normal and pathological along with their spectrograms. The spectrograms are used to show the spectral density of a signal varies with the time which showed by the darkness of the plot of the frequency analysis. During the period of voiced sound, frequency energy of spectral is seen in the spectrogram while in silence period, the spectral cannot be detected. The speech signal pattern and its spectrogram of normal and disorder words for word *Gas(ගස්)* and word "Mukunuvanna"(මුකුණුවැන්න) are depicted in Figure 16 and Figure 17 respectively. The dissimilarity between normal and pathological for word *Gas(ගස්)* is the characteristic of unusual pausing pattern for pathological word as shown in Figure 17. For word "Mukunuvanna"(මුකුණුවැන්න), there is an obvious difference between the normal and disorder utterances in the syllables of word "Mukunuvanna"(මුකුණුවැන්න) as shown in the following Figure 17.

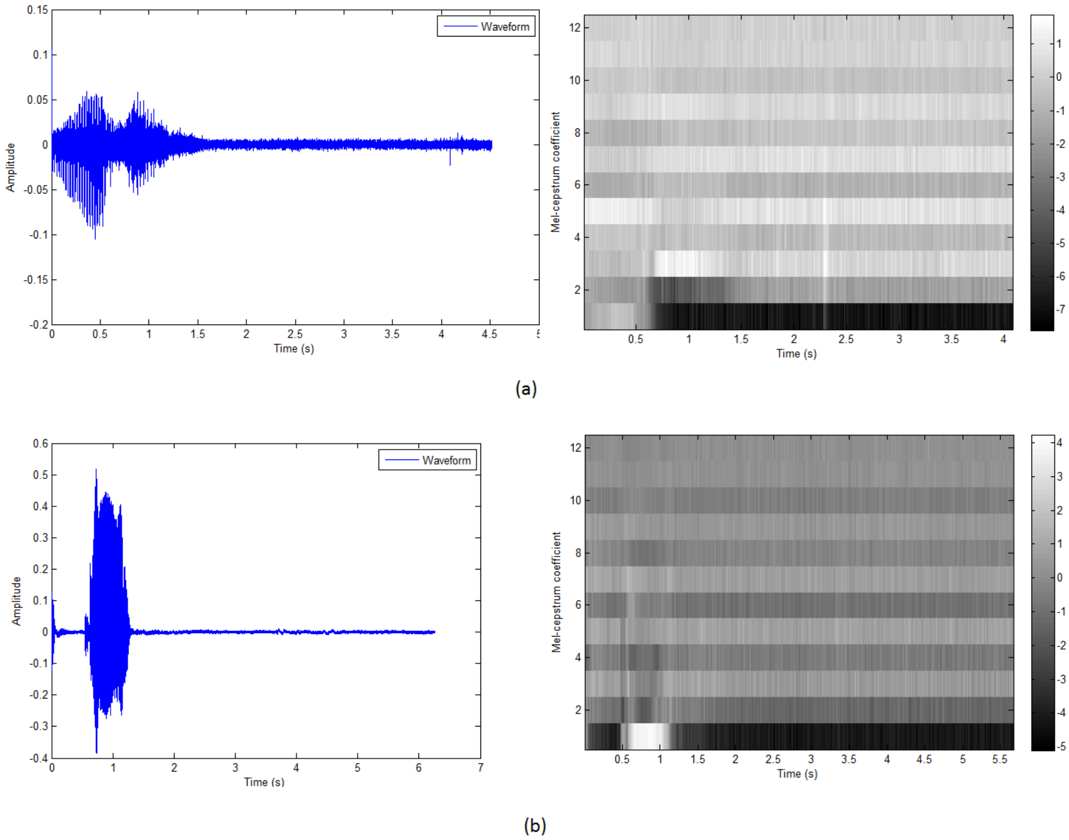


Figure 4-12: Speech signal pattern and spectrogram of (a) normal word "Gas"(ගස්) and (b) Disorder word "Gas"(ගස්)

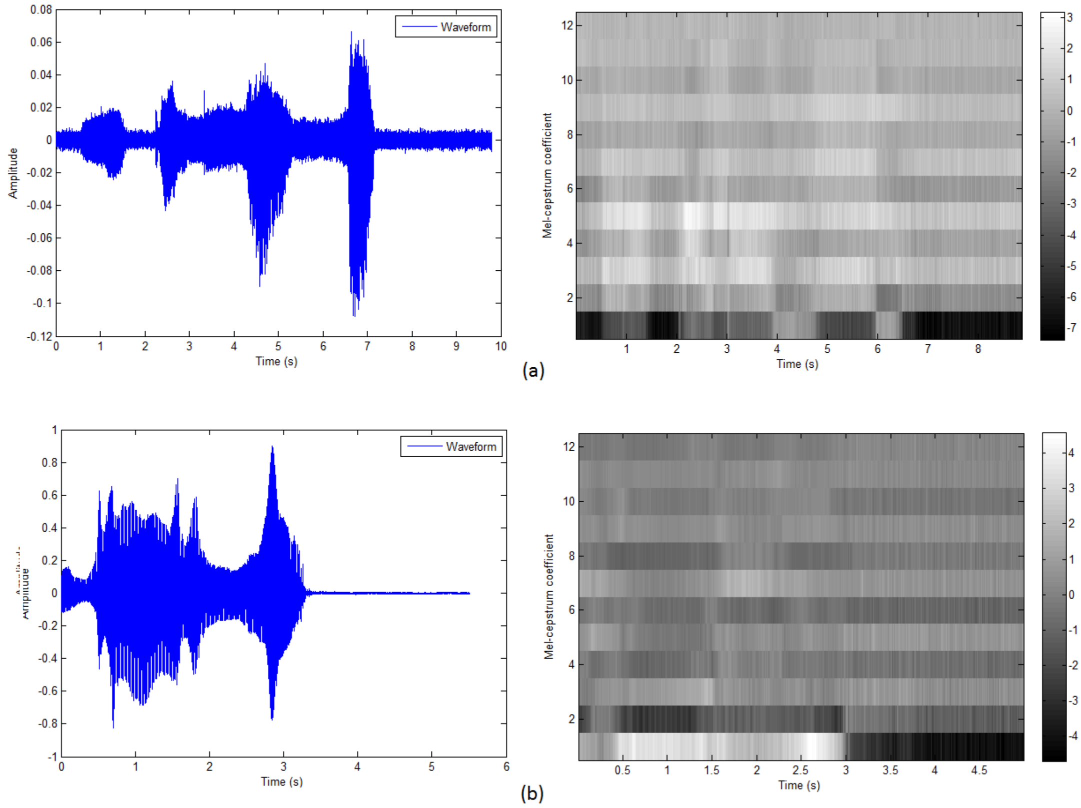


Figure 4-13: Speech signal pattern and spectrogram of (a) normal word "Mukunuvanna"(මුකුණුවැන්න) and (b) Disorder word "Mukunuvanna"(මුකුණුවැන්න)

4.6.2 Detailed design specification of Patient's Data Analysis

This module has been developed for the application programs developed to perform analysis of the patients' data. It has been developed using Matlab programming language to. The main module consists of three different sub modules; Patient's Pitch Analysis, Patient's Formant Frequency Analysis, and Patient's Mel Frequency Analysis. Detail specifications of each module are given below. Refer figure 18.

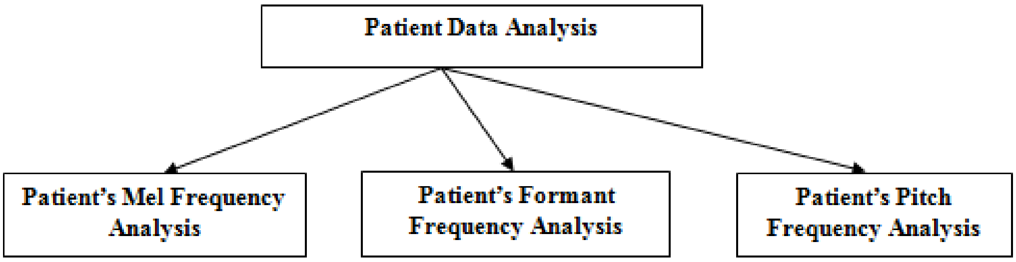


Figure 4-14: Program modules Patient Data analysis

Compare the deviation of patient's each extracted feature within the standard deviations of ordinary children's data. Mathematically it is denoted as;

The average of extracted feature of ordinary children = f

Standard deviation of above range = s

Patient's extracted feature = pf

Therefore, when each instance of extracted feature satisfies following inequalities it is included under one of these categories.

$(f - s) <= pf <= (f + s) \rightarrow$ perfect

$(f - 2s) <= pf <= (f + 2s) \rightarrow$ average

$(f - 3s) <= pf <= (f + 3s) \rightarrow$ poor

This analysis will be performed for all 10 words and find the total number of perfect, average and poor values.

Chapter 5

5 Implementation

5.1 Data Collection

Initially got the records of 100 students including girls and boys in Kandy, Kurunegala and Colombo districts and 10 patients. Each data set consisted of selected 10 words. These values were recorded as .wav format. The selected patients included both males and females between 5 to 8 years of age.

5.2 Graphical User Interface (GUI)

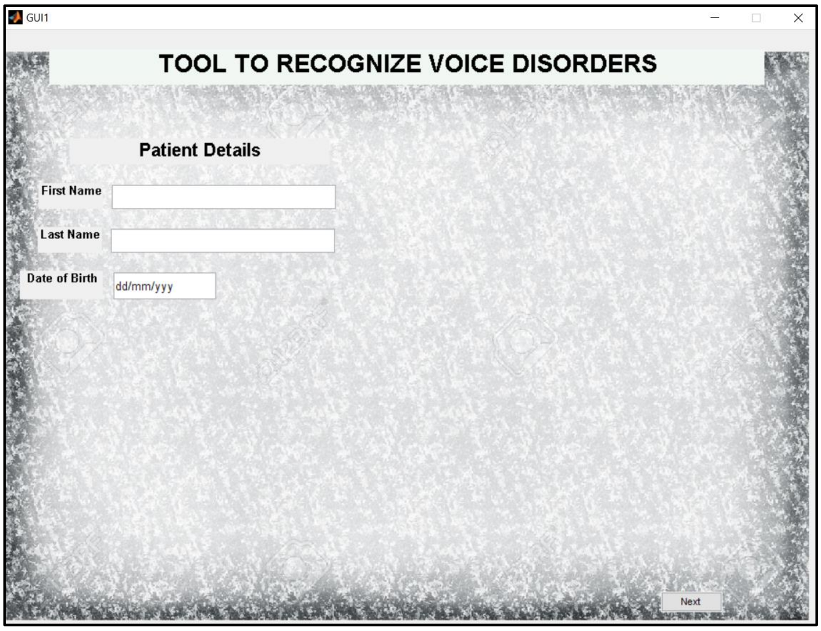


Figure 5-1: Patient Data Collection Screen

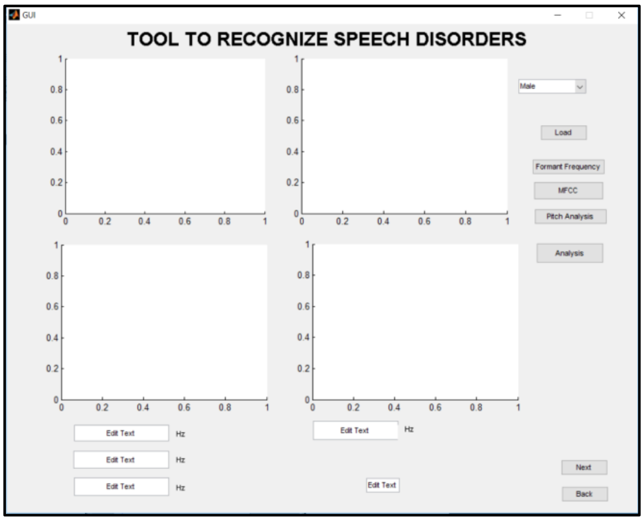


Figure 5-2: Load the Input Signal and Extract the Selected Three Features

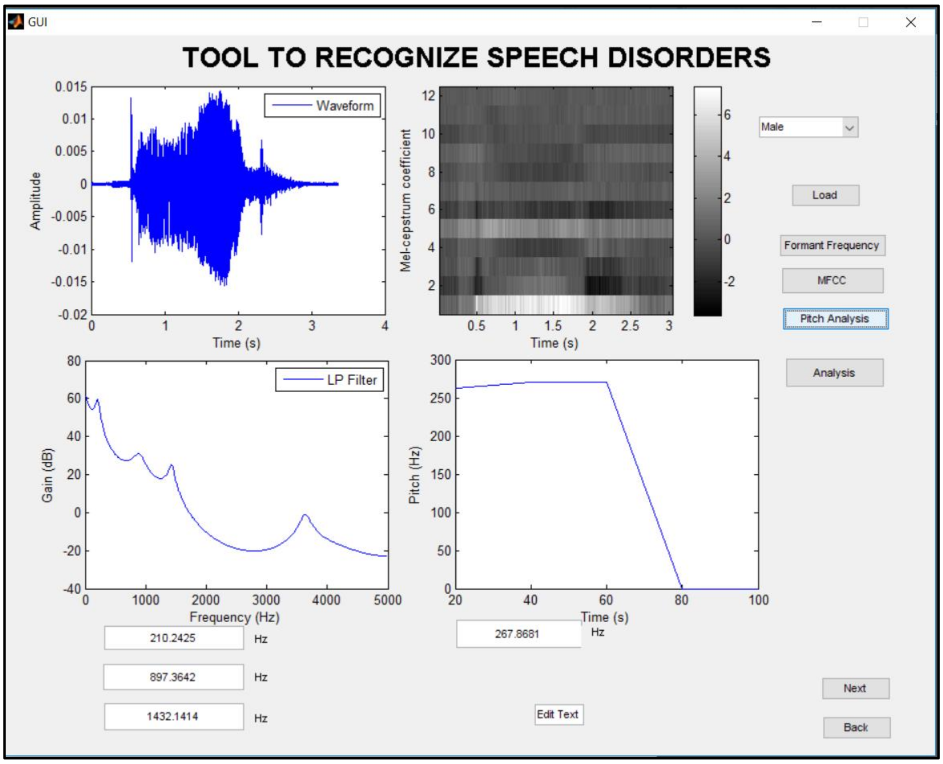


Figure 5-3: Feature Extraction for One Word

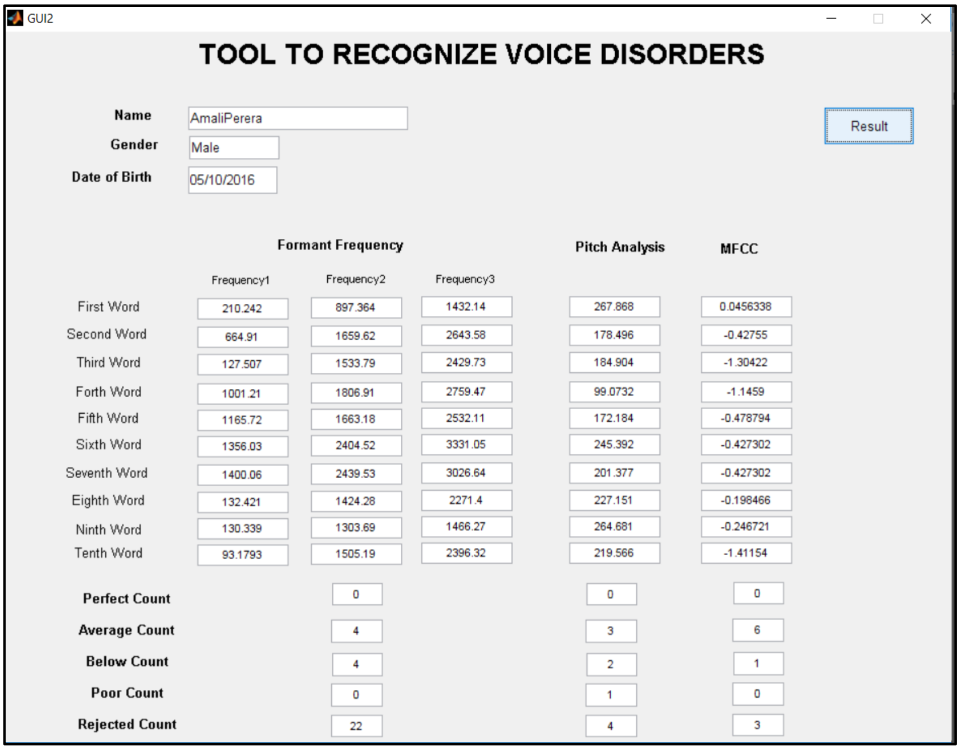


Figure 5-4: Result Screen for One Patient

6 Evaluation

6.1 Formant Frequency

6.1.1 Database update of Formant Frequencies of Ordinary Children

Sample consists of 70 children (members), their speech recordings of 10 words have considered for this database. The average frequency of each word with the relevant standard deviation are included in the Table 7.

Table 6-1: Average formant frequency values and Standard deviations of ordinary children for the ten words in the phonetic balance

	F1		F2		F3	
Words	**A1**	**S1**	**A2**	**S2**	**A3**	**S3**
ගස්	252.426	56.123	759.86	223.014	1407.72	266.724
ලේන්	111.831	16.793	734.745	58.223	1098.18	263.487
අළුත්	206.052	61.803	738.751	209.647	1364.18	306.82
දෙකටක්	278.041	97.457	632.311	186.514	1123.83	276.671
දෙන්නේ	129.078	14.167	774.693	66.027	1092.6	124.587
හිටවන්න	269.772	60.96	731.034	196.887	1206.73	315.096
කුඩුවක්	274.86	67.007	697.131	209.497	1254.03	354.17
පන්තිය	234.593	66.995	561.57	234.507	1026.4	177.897
පොත්	170.095	30.294	897.409	283.939	1854.61	670.491
මුකුණුවැන්න	185.492	23.558	673.121	88.478	1176.15	238.433

Key

F1 – Formant Frequency 1

F2 – Formant Frequency 2

F3 – Formant Frequency 3

S1, S2, S3 – Standard Deviations

A1, A2, A3 – Averages

6.1.2 Comparison between disorder and ordinary children

The following analysis performed to find whether the frequencies are within the ranges of relevant standard deviations.

$$F == A \rightarrow Perfect$$

$$A - S \leq F \leq A + S \rightarrow Average$$

$$(A - 2 \times S) \leq F < (A - S) OR\ (A + S) < F \leq (A + 2 \times S) \rightarrow Below\ Count$$

$$(A - 3 \times S) \leq F < (A - 2 \times S) OR\ (A + 2 \times S) < F \leq (A + 3 \times S) \rightarrow Poor\ Count$$

$$F < (A - 3 \times S)\ OR\ F > (A + 3 \times S) \rightarrow Reject\ Count$$

Table 6-2:Formant Frequency values of ordinary child 1 for the ten words in phonetic balance

Words	**F1**	**F2**	**F3**
ගස්	210.242	897.364	1432.14
ලේන්	88.7155	836.111	2409.69
අළුත්	104.815	391.577	1160.16
දෙකටක්	894.079	1370.5	2834.86
දෙන්නේ	135.006	787.28	1060.68
හිටවන්න	263.295	731.108	1374.85
කුඩුවක්	282.336	832.824	1436.42
පන්තිය	72.9927	961.113	1353.57
පොත්	164.591	1107.39	1390.03
මුකුණුවැන්න	346.359	875.469	1413.82

Table 6-3: Comparison of Table 7 and Table 8

Perfect Count	0
Average Count	17
Below Count	6
Poor Count	2
Rejected Count	5

Table 6-4: Formant Frequency values of disorder child 1 for the ten words in phonetic balance

Words	F1	F2	F3
ගස්	187.172	1307.85	2092.76
ලේන්	664.91	1659.62	2643.58
අළුත්	127.507	1533.79	2429.73
දෙකටක්	1001.21	1806.91	2759.47
දෙන්නේ	1165.72	1663.18	2532.11
හිටවන්න	1356.03	2404.52	3331.05
කුඩුවක්	1400.06	2439.53	3026.64
පන්තිය	132.421	1424.28	2271.4
පොත්	130.339	1303.69	1466.27
මුකුණුවැන්න	93.1793	1505.19	2396.32

Table Count 6-5: Comparison of Table 7 and Table 10

Perfect	0
Average Count	1
Below Count	5
Poor Count	2
Rejected Count	22

Based on 30 ordinary children and 10 disorder children following results can be obtained

	Normal Children	Disorder Children
Average Count	(min) - 50% (15 words)	(max) – 16% (5 words)
Rejected Count	(max) – 16% (5 words)	(min) – 66% (20 words)

6.2 Pitch Analysis

6.2.1 Database update of Pitch Frequencies of ordinary children

Sample consists of 70 children (members), their speech recordings of 10 words have considered for this database, the average frequency of each word with the relevant standard deviation are included in the Table 12.

Table 6-6: Average pitch values and Standard deviations of ordinary children for the ten words in phonetic balance

Words	Average Pitch Values (A)	Standard Deviation (S)
ගස්	245.419	31.772
ලේන්	262.738	22.037
අළුත්	276.657	22.665
දෙකටක්	244.051	28.413
දෙන්නේ	255.519	17.958
හිටවන්න	256.401	17.854
කුඩුවක්	242.786	37.331
පන්තිය	257.075	19.837
පොත්	236.512	28.876
මුකුණුවැන්න	268.188	16.37

6.2.2 Comparision between normal and disorder children

The following analysis performed to find whether the frequencies are within the ranges of relevant standard deviations.

$$F == A \rightarrow Perfect$$

$$A - S \leq F \leq A + S \rightarrow Average$$

$$(A - 2 \times S) \leq F < (A - S) OR\ (A + S) < F \leq (A + 2 \times S) \rightarrow Below\ Count$$

$$(A - 3 \times S) \leq F < (A - 2 \times S) OR\ (A + 2 \times S) < F \leq (A + 3 \times S) \rightarrow Poor\ Count$$

$$F < (A - 3 \times S)\ OR\ F > (A + 3 \times S) \rightarrow Reject\ Count$$

Table 6-7: Pitch values of ordinary child 1 for the ten words in phonetic balance

Words	Pitch values (F)
ගස්	267.868
ලේන්	253.497
අළුත්	294.907
දෙකටක්	115.474
දෙන්නේ	249.684
හිටවන්න	266.62
කුඩුවක්	278.205
පන්තිය	260.177
පොත්	302.182
මුකුණුවැන්න	273.857

Table 6-8: Comparison of Table 12 and Table 13

Perfect Count	0
Average Count	8
Below Count	0
Poor Count	1
Rejected Count	1

Table 6-9: Pitch values of disorder child 1 for the ten words in phonetic balance

Words	Pitch values (F)
ගස්	197.068
ලේන්	178.496
අළුත්	184.904
දෙකටක්	99.0732
දෙන්නේ	172.184
හිටවන්න	245.392
කුඩුවක්	201.377
පන්තිය	227.151
පොත්	264.681
මුකුණුවැන්න	219.566

Table 6-10: Comparison of Table 12 and Table 15

Perfect Count	0
Average Count	2
Below Count	3
Poor Count	1
Rejected Count	4

Based on 30 ordinary children and 10 disorder children following results can be obtained.

	Normal Children	Disorder Children
Average Count	(min) - 50% (5 words)	(max) – 30% (3 words)
Rejected Count	(max) – 30% (3 words)	(min) – 30% (3 words)

6.3 MFCC Analysis

Sample consists of 70 children (members), their speech recordings of 10 words have been considered for this database. The average frequency of each word with the relevant standard deviation are included in the Table 17.

6.3.1 Database update of MFCC of normal children

Table 6-11: Average MFCC values and Standard deviations of ordinary children for the ten words in phonetic balance

Words	Average MFCC Values (A)	Standard Deviation (S)
ගස්	-0.3268	0.3269;
ලේන්	-0.0446;	0.5398;
අළුත්	-0.312;	0.224;
දෙකටක්	0.044;	0.252;
දෙන්නේ	-0.0641;	0.4454;
හිටවන්න	0.1351;	0.6539;
කුඩුවක්	-0.3069;	0.250;
පන්තිය	-0.6109;	0.5129;
පොත්	-0.4829;	0.2625;
මුකුණුවැන්න	0.0218;	0.3792;

6.3.2 Comparision between ordinary and disorder children

The following analysis performed to find whether the frequencies are within the ranges of relevant standard deviations.

$$F == A \rightarrow Perfect$$

$$A - S \leq F \leq A + S \rightarrow Average$$

$$(A - 2 \times S) \leq F < (A - S) OR\ (A + S) < F \leq (A + 2 \times S) \rightarrow Below\ Count$$

$$(A - 3 \times S) \leq F < (A - 2 \times S) OR\ (A + 2 \times S) < F \leq (A + 3 \times S) \rightarrow Poor\ Count$$

$$F < (A - 3 \times S)\ OR\ F > (A + 3 \times S) \rightarrow Reject\ Count$$

Table 6-12: MFCC values of ordinary child 1 for the ten words in phonetic balance

Words	MFCC values (F)
ගස්	0.045634
ලේන්	-0.43671
අළුත්	-0.44825
දෙකටක්	0.027928
දෙන්නේ	-0.02784
හිටවන්න	0.720056
කුඩුවක්	-0.90505
පන්තිය	-0.28914
පොත්	-0.84278
මුකුණුවැන්න	-0.30736

Table 6-13: Comparison of Table 17 and Table 18

Perfect Count	0
Average Count	4
Below Count	5
Poor Count	1
Rejected Count	0

Table 6-14: MFCC values of disorder child 1 for the ten words in phonetic balance

Words	MFCC values (F)
ගස්	-0.90574
ලේන්	-1.41154
අළුත්	-0.42755
දෙකටක්	-1.30422
දෙන්නේ	-1.1459
හිටවන්න	-0.47879
කුඩුවක්	-0.4273
පන්තිය	-1.3764
පොත්	-0.19847
මුකුණුවැන්න	-0.24672

Table 6-15: Comparison of Table 17 and Table 20

Perfect Count	0
Average Count	6
Below Count	1
Poor Count	0
Rejected Count	3

6.4 Critique

The main objective of this chapter is to provide an evaluation of the development process of the intelligent tool. The discussion is organized based on the following topics.

- Verification and validation of analysis process of digital signal processing
- How much user requirements are successfully achieved

Successfulness of dependability attributes like availability, reliability safety and security

6.4.1 Verification and validation of analysis process of digital signal processing

The voice disorders have a strong relationship to the linguistics studies. Many researchers have conducted various researches on linguistics studies. But these researches did not address the issues relevant to digital signal processing. The researcher has found researches pertinent to digital signal processing with a certain background to linguistics studies. In which the researchers have applied three concepts i.e. pitch frequency analysis, formant frequency analysis and Mel frequency analysis for the identification process of voice disorders. (Refer literature review for more information). The developer has used these technologies as formal methods to develop the intelligent tool.

The MATLAB programs used for the Mel frequency analysis had been developed by the "mathwork" software group. The programs for the other two analyses such as pitch analysis, and formant frequency analysis were developed by the researcher. For these programs, no input validation routines were included. These validations are difficult to include, because the inputs for these programs are from sound files. It is difficult to develop validation routines for sound files.

6.4.2 How much user requirements are successfully achieved

Two Speech Language Therapists (SPTs) were initially interviewed by the researcher to identify the requirements for the intelligent tools. They were unable to provide a clear image or the features to be included to the expected tool. Later the researcher interviewed a professional speech language therapist/audiologist and he has provided a list of requirements and operational features are to be included to the automated tool. (Refer appendix A for the detailed discussion). In addition to it, the researcher has investigated other intelligent tools available in the market which can be used for sound analysis. The icSpeech Analyzer, sound analysis software trial version available to download, was examined to identify the features

6.4.3 Successfulness of dependability attributes like availability, reliability, safety, and reliability

This topic focuses to introduce the steps used by the researcher to improve the software quality. The quality management process can be evaluated in terms of the validation routines applied to the programs by the developer. The validations are measured by the dependability attributes like availability, reliability, safety and security.

Availability is a critical measurement. This suggests that the architecture should be designed to include redundant components so that it is possible to replace and update components without stopping the system.

The reliability of the system depends on the validation routines applied to the software and the number applied test routines to find bugs in the system. The developer has divided the software modules into two parts; voice input component and patient data analysis module. The researcher has applied data validation routines for data analysis modules. To identify runtime errors of text file readings were managed by exception handlers; ‘File Not Found Exceptions’. The researcher has used validation routes for data entry screens also. In addition to these validation routines, error messages were included for the users to provide appropriate directions.

Safety is a critical issue for most of the bio medical systems. However, safety will not be a critical issue of this system because the information provided by the system would not be used for oral medication or any vaccination process.

Security will not be a critical issue for this system because the system has not been designed to run on network systems or internet. System has been designed to run on standalone computer systems.

Software standards are important to quality assurance as they represent an identification of ‘best practice’. Software standards compel the user to follow compulsory routines. Therefore it automatically enhances the quality or provides a guarantee of the operability of the system. There is no universally applicable standard for software development. Most standards cover the relationships between internal and external software attributes like portability, usability, reliability, and maintainability. It is recommended to follow a quality management process in relation to analysis and software development, if the product is going to be implemented in the real world.

Chapter 7

7 Conclusion and Future Works

The focus of this study is to design an intelligent tool to identify a common voice disorder. Early medication is recommended to obtain successful results of these patients. The tool has been developed to assist for speech language therapists when diagnosing process is carried out in voice disorders. The treatments provided for these patients usually take a long period and difficult to monitor the progress. Therefore, this intelligent tool can also assist to recognize progress of the treatment plans.

This research can be applied the principles of digital signal processing such as pitch frequency analysis, formant frequency analysis and Mel frequency analysis as formal methods. The required software for the application can be developed on the MATLAB software. The main significance of this study is that the native language of Sri Lanka i.e. Sinhala has not been tried out before for such a study.

The tool has been used to diagnose a sample recordings of disorder children, 90 percent of the pitch frequencies have a significant deviation comparison to members of ordinary sample. Other significant features of formant frequency and MFCC frequency have same features with a significant deviation to the calculated results of normal children.

The limitation of this study is that the pronunciation styles of any language differ from the person’s age, gender, cultural background and the geographical location. Therefore, further research is needed when implementation of such a product in the real world.

References

[1] A. Hafiza, M. Saod and D. A. Ramil, "Preliminary Study on Classification of Apraxia Speech using Support Vector Machine," *Australian Journal of Basic and Applied Science*, 2011.

[2] Chiu Ying Lay and Ng Hian James, " Gender Classification from Speech"

[3] Dorra Garouri, Med Ali Kammoun and Ahmed Ben Hamida, "A Comparative Study of Formant Frequencies Estimation Techniques", University of Sfax,Tunisia. Rep. 27.May.2006.

[4] Dr. Peter Assmann, "HCS 7367", Speech Perception lab. 2011

[5] Esfandiar Zavarehe. *Formant Extraction and Speech Recognition* [Online]. Available: http://dea.brunel.ac.uk/cmsp/home_esfandiar/formants.htm

[6] Hosom JP, Shriberg L and Green JR. *Diagnostic Assessment of Childhood Apraxia of Speech Using Automatic Speech Recognition (ASR) Methods.* [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17066124

[7] I. Sommerville, Software Engineering 6th Ed, Addison-Wesley Publishers, 2001.

[8] J. Harrington and S. Cassidy, "The acoustic source in speech production," *Techniques in Speech Acoustics*, pp. 34-40, 1999.

[9] J. Hosom, J. Green and L. Shriberg, "Diagnostic Assessment of Childhood Apraxia of Speech Using Automatic Speech Recognition (ASR) Methods.," [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17066124. [Accessed 10 11 2017].

[10] J. Nayak and P. S. Bhat, "Identification of Voice Disorders using Speech Samples," Manipal, 2003.

[11] Jonathan Harrington and Steve Cassidy, "The acoustic source in speech production" in *Techniques in Speech Acoustics,* Kluwer Academic Publishers, 1999, pp 34-40

[12] J. Price, "Design of an Automatic Speech Recognition System Using MATLAB," Maryland, 2005.

[13] Kishore Parahallad, “Speech Technology: A Practical Introduction”, Carnegia Mellon University & International Institute of Information Technology Hyderabad.

[14] L. Lee, J. C. Stemple, L. Glaze and L. N. Kelchner, "Quick Screen for Voice and Supplementary Documents for Identifying Pediatric Voice Disorders," *Language, Speech and Hearing Services in Schools,* vol. 35, pp. 308-31, 2004.

[15] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, “Voice Recognition Alogorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques”, Journal of Computing, vol 2,

[16] Lotfi Salhi, Talbi Mourad, and Adnene Cherif, “Voice Disorder Identification Using Multilayer Neural Network”, Signal Processing Laboratory Sciences Faculty of Tunis, University Tunis ELManar, Tunisia. The International Arab Journal of Informational Technology, Vol 7. 2, April 2010

[17] L. Salh, M. Talbi and A. Cherif, "Voice Disorder Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks," *World Academy of Science, Engineering and Science,* pp. 330-338, 2008.

[18] L. Salhi, T. Mourad and C. Adnene, "Voice Disorders Identification Using Multilayer Neural Network," *The International Arab Journal of Information Technology,* vol. 7, no. 2, pp. 177-185, 2010.

[19] M. Ashby and J. Maiment, Introducing Phonetic Science, Cambridge, 2005.

[20] M. A. Wahed, "Computer Aided Recognition of Pathological Voice," in *31st National Radio Science Conference,* Egypt, 2014.

[21] M. K. Arjmandi, M. Pooyan, H. Mohammadnejad and M. Vali, "Voice Disorders Identification Based on Different eature Reduction Methodoloies and Support Vector Machine," in *ICEE,* Tehran, 2010.

[22] Nick Bardici and Bjorn Skarin, “Speech Recognition using Hidden Markov Model”, school of Engineering, Department of Telecommunications and Signal Processing. Blekinge Institute of Technology. March 2006.

[23] Noelia Alcaraz Meseguer, “Speech Analysis for Automatic Speech Recognition”, Norwegian University of Science and Technology. Rep. July 2009

[24] N. Souissi and A. Cherif, "Speech Recognition System Based on Short -term

Cepstral Parameters, Feature Reduction Method and Artificial Neural Networks," in *2nd International Conference Advanced Technologies*, Tunisia, 2016.

[25] P. Williams, "Developmental Verbal Dyspraxia," London.

[26] S. W. Smith, "The Breadth and Depth of DSP," in *The Scientist & Engineer's Guide to Digital Signal Processing* , California, California Technical Pub, 1997, p. 629.

[27] UCL Department of Phonetic and Linguistics, *Introduction to Programming with MATLAB.*[Online].Available: http://www.phon.ucl.ac.uk/courses/spsci/matlab/lect1.html

[28] U. D. o. H. a. Human Services, "National Institute on Deafness and Other Communication Disorders," National Institute of Health, 31 October 2017. [Online]. Available: https://www.nidcd.nih.gov/health/apraxia-speech. [Accessed 17 November 2017].

[29] Whitten, Bentley and Dittman, System Analysis Design Methods 6th Ed, Tata McGraw-Hill, 2004.

Appendices

Appendix A

Mr W. V. Singhanatha (B.Ed., M.Ed., MA) is an Educational Audiologist & Speech Therapist and Chief Project Officer, National Institute of Education, Maharagama. The discussion was held on 16/07/2018 at Mattegoda Housing Scheme, Mattegoda

He has addressed following issues;

1. Voice disorder is a difficulty of in making and coordinating the precise movements to make a clear speech. There are no evidences to prove that the problem will arise due to the problems of nerves or muscles or problems of hearing. The treatments for the disease mainly depend on the exercises provided to improve the movements of larynx, lips, tongue and soft palate. The treatment plans must be designed by a professionally qualified speech language therapist.
2. It would be very useful to develop a computerized intelligent system and which can be directly used in the diagnosing process. Further, the tool can also be used to monitor the progress of the treatment plans. One common method of identifying childhood communication disorders is through mass screening. These screening methods provide some basic information about the quality of the sound. But this will not be adequate to make a strong decision to identify the attributes of the qualities of the child's speech. The physical properties like respiration, phonation, and resonance are more suitable parameters to analyze a voice disorder. Therefore, if a tool can be developed based on these parameters and it would be more useful for the clinicians and speech language therapists to recognize the speech disorder.
3. Prior to design the tool, it would more helpful to identify the features of the computerized packages developed in this area. For example, IcSpeech Analyzer is a computerized system provides analysis based on the digital signal processing principles such as Fast Fourier Transform (FFT). Pratt and MATLAB are other useful packages to perform analysis based on the principles of digital signal processing.
4. A person's pronouncing style depends on the age, gender, cultural background, and geographic location. When a research is going to design, the

impact of these parameters has to be fixed. Members of both normal and pathological have to be chosen for the experiment. However to mitigate the impact of above mentioned parameters, two groups (samples) have to be chosen from single population. It is difficult to record speech for a large number of words. Therefore a list of words has to be selected for the speech recording. The list of words is known as phonetic balance. It is suitable to get expert advices to create the phonetic balance.

Ms. S. A. Malani Chadralatha (BA, MA) is a Teacher, Ministry of Education and Linguistics Expert. The discussion was held on 6/8/2018 at Balagolla Kandy.

She has addressed the following issues with regard to make the phonetic balance.

1. Sounds in speech are of two kinds; vowels and consonants. Different language has different ways of placing these vowels and consonants together to write words. In Sinhala vowels are known as ‘panakuru’ and consonants are known as ‘gathakuru’. Placing these vowels and consonants produce words. In Sinhala certain words are produced by placing only consonants. But these consonants are mixed with vowel sound; therefore these words can also be produced.
2. The phonetic balance for this experiment has been created by using the recommended reading book of the year 2 students. The words included in the reading book are concerned as the expected language proficiency of year 2 students.
3. Following words have been chosen for the phonetic balance.

Two character words - ගස්, ලේන්, පොත්
Three character words – අළුත්, අන්න, දෙන්නේ
Four character words – දෙකටක්, කුඩුවක්, පන්තිය
Five character words - හිටවන්න
Six character words- මුකුණුවැන්න