DEVELOPMENT OF AN AUTOMATIC SPEECH PROCESSING BASED PROCEDURE FOR ASSESSMENT OF STUTTERING IN MALAYALAM SPEAKING ADULTS THROUGH VIRTUAL MODE

Ms. Darsitha S alias Sneha S

Reg. No. : 20SLP014

Dissertation Submitted in Part Fulfillment of Degree of Master of Science

(Speech-Language Pathology)

University of Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSURU - 570006

AUGUST 2022

CERTIFICATE

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Malayalam speaking adults through virtual mode**" is a bonafide work submitted in part fulfillment for degree of Master of Science (Speech Language Pathology) of the student Registration Number: **20SLP014**. This has been carried out under the guidance of the faculty of this institute and has not been submitted earlier to any other University for the award of any other diploma or degree.

Prof. M. Pushpavathi

Mysuru

Director

August 2022

All India Institute of Speech and Hearing

Manasagangothri, Mysuru - 570006

CERTIFICATE

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Malayalam speaking adults through virtual mode**" has been prepared under our supervision and guidance. It is also being certified that this dissertation has not been submitted earlier to any other University for the award of any other diploma or degree.

Co-Guide

Dr. Sangeetha Mahesh

Associate Professor & HOD

Department of Clinical Services

Guide

Dr. Ajish K. Abraham

Professor of Electronics and Acoustics

Department of Electronics

All India Institute of Speech and Hearing,

Manasagangothri, Mysuru -570006.

Mysuru

August 2022

DECLARATION

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Malayalam speaking adults through virtual mode**" is the result of my own study under the guidance of Dr. Ajish K. Abraham, Professor of Electronics and Acoustics**,** Department of Electronics and co-guidance of Dr. Sangeetha Mahesh, Associate Professor & Head, Department of Clinical Services, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier to any other University for the award of any other diploma or degree.

Mysuru **Register No. : 20SLP014**

August, 2022

Dedicated to

Achan, Amma &

Vava

For the unconditional love and support

ACKNOWLEDGEMENT

It is impossible to extend enough thanks to my family, my Amma, Achan and Vava who gave me the encouragement I needed throughout this process. Thank you Acha and Amma for always standing by my side and supporting me, I am proud to be your daughter.

I extend my sincere gratitude to my guide Professor Ajish K. Abraham for his unwavering support and guidance. This study would not have been possible without your help, sir. I am blessed to have you as my mentor. I will always appreciate everything you have taught me. Thank you so much sir for being kind and inspiring. I also want to thank my co-guide, Dr. Sangeetha Mahesh, who never lost hope in me and provided me the extra support I needed. Thank you ma'am for having faith in me.

Nothing is possible without god's grace…I am blessed and I thank god, for every day for everything that happens for me.

I acknowledge Dr. M Pushpavathi, Director of All India Institute of Speech and Hearing, for providing a platform to conduct dissertation study.

I would like to extend my sincere thanks to National Institute of Speech and Hearing, especially Dr. Vinitha Mary George HOD of Department of audiology and Speech Language Pathology, Nirmal sir, and Anne ma'am for helping me in data collection.

Special thanks to Dr Reuben Thomas Varghese, Ms. Hina, Ms. Neeraja, and Ms. Revathi, for their inputs with respect to validation and verification of the quality rating and Dr. Vasanthalakshmi for her assistance and patient guidance with the statistical part.

I would like to thank Hina ma'am and Veda di, for their constant support and guidance throughout the study. I am overwhelmed with your kind gestures towards me. Thanks for all the help and motivation you have given me.

Kichu, thank you for listening to me, understanding me and supporting me so well. You have never failed to convince me that I am strong enough to overcome the struggles and find my own way. Thanks for being the shoulder I can always depend on.

My crazy girls Shinsi, Saamie, Abna, Adithya, Malu, Dyuthi, Lekshmi , and Aiswarya you guys made my bachelor life so memorable. Thanks for making my college life extraordinary.

Shinsi and Dyuthi…. thanks for being the best companions at AIISH. I will cherish all the moments we had together. Thank you joel for the surprises and all the help.

Thanks to Monisha and Audrey who took time out of their busy schedule for helping me in the process.

A special thanks to you Audrey, for helping me despite compromising on your sleeps. Jayasree….. thank you for sharing your craziness with me, it was a ray of happiness in the middle of chaos.

My dear friends Sujisha, Audrey, Swathi, Nayana, Erica and Trupti who were there with me throughout my MSc life, without you my MSc life wouldn't have been this fun.

Aiswarya, Alfiya, and Devika, it was a wonderful experience to get to know more about you, thank you for all the lovely memories throughout this journey.

Yasha and Bhavani thank you for being so kind and sweet.

I also thank my UG (Renovators) and PG (Master artifacts) batch mates for all the good moments, and all the juniors and seniors who helped me directly or indirectly during the study.

Last but not the least, I want to thank all the participants for their cooperation, patience, and enthusiasm during the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The current research aimed to develop and evaluate a virtual mode approach using automatic speech processing for assessing fluency parameters in Malayalam-speaking adults who stutter. Objectives of the study were; a. to assess the fluency parameters such as frequency and duration of filled pauses, blocks, repetitions and prolongations, using the stuttering severity instrument – 4th edition (SSI – 4) in virtual mode, b. to assess the fluency parameters using automatic speech processing from the standard Malayalam passage spoken by 15 adults with stuttering (age range of 18-35 years), recorded through Zoom app, c. to compare the values of fluency parameters obtained from automatic speech processing with those derived via perceptual assessment, and d. to analyze the online assessment's efficacy based on feedback from participants and three SLPs. Each participant was asked to read the standard Malayalam passage presented using the Zoom app's presentation mode. The session was recorded and stored on a computer for further analysis. The results of perceptual evaluation and automatic speech processing were compared using Wilcoxon signed rank test. The results revealed no significant difference between the detection of stuttering events through perceptual evaluation and automatic speech processing. The overall accuracy in detection of fluency parameters ranged from 36% to 44%. Also technical and clinical quality of the online assessment was found to be satisfactory by the participants as well as SLPs. The study established the viability of doing online assessments for fluency disorder. Hence the study's findings will promote telepractice for stuttering evaluation, particularly in epidemic situations like COVID-19 where it is difficult to perform a conventional face-to-face examination.

CHAPTER I

INTRODUCTION

Stuttering is characterized by involuntary, audible or silent repetitions or prolongations of sounds or words, resulting in a breakdown in verbal expressive fluency. These are difficult to control and may be accompanied by additional movements as well as negative feelings like fear, anxiety, or frustration (Wingate, 1964). Fluency disorder is characterized by primary (core features) and secondary behaviors. Repetition of sounds, syllables, or the entire word, as well as prolongations of single sounds or blocks of airflow or voicing during speech, are all primary behaviors. Interjections, circumlocutions or word avoidances, and abnormal breathing patterns are also observed along with the primary behaviors.

Dysfluencies can be found in both normally fluent speakers and people who stutter. To distinguish between typical and atypical dysfluencies, the words "stutter-like dysfluency" (SLD) and "other like dysfluency" (OLD) have been coined. SLDs include prolongations, syllable repetitions, part word repetitions, and blocks, whereas OLDs include phrase repetitions, revisions, and interjections. SLDs are not commonly seen in normally fluent speakers. OLDs are frequently seen in the speech of both persons with stuttering and normally fluent speakers.

Yairi and Ambrose (1999), based on the review of 11 studies, found that the average age of onset was 42 months. Approximately 2% of children aged 3 to 17 years have stuttering (Zablotsky et al., 2019). Stuttering is observed in 0.78 percent of persons between the ages of 21 and 50, and in 0.37 percent of those who are aged 51 and above (Craig et al., 2002).

1.1 Conventional method for assessment of stuttering

The fourth edition of the Stuttering Severity Instrument (SSI-4) is a reliable

and extensively used tool for assessing stuttering severity. The SSI-4 evaluates four aspects of a person's speech: frequency, duration, physical concomitants, and naturalness of speech (Riley, 2009). The person being evaluated using SSI-4 will be asked to read a standardized passage in his native language. The following are assessed using perceptual evaluation from the spoken passage:

- The frequency of stuttering - measured based on the number of syllables and words stuttered.
- Dysfluencies' duration (Typical dysfluencies include hesitations, filler words, revisions, and phrase and word repetitions; atypical dysfluencies include blocks, prolongations, and sound syllable and word repetitions.)
- Disturbance in the forward flow of speech
- Reduced or no eye contact
- Struggle behaviors.
- Behavioral patterns of avoidance

1.2 Limitations of the conventional method

The following limitations have been observed:

- When the person is reading a standard text or speaking spontaneously, it is difficult for the clinician to manually count the stuttering syllables.
- When a client exhibits dysfluency, determining the syllable boundary can be challenging.
- Only a skilled clinician will be able to determine the stuttered events.
- There are no recognized factors that influence the naturalness of speech.
- It only considers overt expression characteristics.

- The conventional assessment takes a long time to complete and is thus time consuming for the client.
- Inter-judge variability.

1.3 Assessment of stuttering using automatic speech processing

There have been studies that have sought to use automatic speech processing to test fluency issues (Bayerl et al., 2020; Chee et al., 2009; Surya & Varghese, 2016). These researchers used the following methods: Mel Frequency Cepstral Coefficient (MFCC) based feature extraction and identifying fluency disorders using LDA based classifier and k-nearest neighbors (k-NN) (Chee et al., 2009; Surya & Varghese, 2016); Linear Prediction Cepstral Coefficients (LPCC) based feature extraction and identifying fluency disorders using LDA & k-NN (Chee et al., 2009).

Because speech output (in stuttering) is non-linear due to involuntary silent pauses, repetition, and lengthening of words, non-linear characteristics such as wavelet packet transform with sample entropy (Hariharan et al., 2013) have been employed to detect dysfluencies. Deep learning architectures have recently been applied to both text and signal level characteristics (Alharbi & Farrahi, 2018; Kourkounakis et al., 2020). Kourkounakis et al. (2020) classified a 4-second stutter file into one of six dysfluencies using spectrogram characteristics. Automatic speech processing can be used to accurately detect stuttering, according to the studies mentioned above. However there hasn't been any attempt towards automatic speech processing based stuttering detection among Malayalam-speaking adults.

1.4 Need for the study

Telepractice is proving to be a viable service delivery approach, with many recipients expressing satisfaction. No research on online stuttering assessment in

Malayalam-speaking adults using automatic speech recognition techniques has been reported. Online evaluation methods need to be examined because traditional face-to-face examinations are difficult during pandemics like COVID- 19. Therefore it is necessary to investigate the feasibility of performing assessments online involving automatic techniques.

Hence the current study intends to create and evaluate a virtual mode approach utilizing automatic speech processing for assessing fluency metrics in Malayalam-speaking adults.

1.5 Aim of the study

To create and test a virtual mode approach based on automatic speech processing for assessing stuttering events in Malayalam-speaking adults.

1.6 Objectives of the study

- To assess the fluency parameters such as frequency and duration of blocks, repetitions and prolongations, using the stuttering severity instrument – 4th edition (SSI – 4) through perceptual evaluation in virtual mode.
- To assess the fluency parameters such as frequency and duration of blocks, repetitions and prolongations, using automatic speech processing from the standard Malayalam passage spoken by 15 adults with stuttering, recorded through Zoom app.
- To compare the values of fluency parameters obtained from automatic speech processing with those derived via perceptual assessment.
- To analyze the online assessment's efficacy based on feedback from participants and three Speech Language Pathologists.

CHAPTER II

REVIEW OF LITERATURE

Interruptions of speech such as repetition, hesitation, or prolongation of sound that can occur in both typically developing people and those who stutter, is called as disfluency (Guitar, 2013). According to Van Riper (1982), any “disruption of the simultaneous and successive programming of muscular movements required to produce a speech sound or its link to the next sound in a word” is referred to as stuttering.

2.1 Types of Disfluencies

Stuttering comprises primary (core) and secondary behaviors. The primary behaviors include repetition (A repeated sound, syllable, or single-syllabic word, a speaker appears to be "stuck" on that sound and keeps repeating it until the next one can be made), prolongations (a stuttering event in which sound or air flow continues while articulator movement is stopped) or blocks (A stuttering event in which an inappropriate stoppage of air flow or voice and often articulator movement as well) while speaking (Guitar, 2013). Secondary behaviors include reactions of a speaker to his or her repetitions, prolongations, and blocks in an effort to finish them promptly or avoid them entirely. Such reactions can start off as a random effort, but they quickly develop into well-honed routines. Escape and avoidance behaviors are two types of secondary behaviors (Guitar, 2013). When the speaker is already stuttering, the speaker attempts to end a stutter and finish the word, this is referred to as escape behavior, whereas when a speaker anticipates stuttering on a word or in a scenario, he or she tries to avoid stuttering. Interjections of extra sounds, such as "uh," before the word on which stuttering is expected, are typical examples of word-based avoidances (Guitar, 2013).

2.2 Causes of disfluencies

Scientists are still trying to figure out what causes stuttering, although they have a lot of leads. Developmental stuttering develops before puberty, usually between the age of two and five, without any obvious brain injury or other known reason (Büchel & Sommer, 2004). After a defined brain injury, such as stroke, intracerebral hemorrhage, or head trauma, neurogenic or acquired stuttering develops. It is a rare occurrence that has been documented after lesions in a range of brain areas (Ciabarra, 2000; Grant et al., 1999). There is considerable evidence that stuttering has a genetic basis—that is, something is inherited that increases the likelihood of a child stuttering. Another indication about the nature of stuttering is that the majority of stuttering starts between the age of two and five. As a result, the development of stuttering coincides with the occurrence of a number of typical early childhood stresses. During a period of rapid growth in vocabulary and syntax, one youngster may begin to stutter. When a family moves to a new place, one member's stuttering may occur for the first time. Many variables, acting separately or in combination, might trigger the development of stuttering in a child with a neurophysiological predisposition, or inborn inclination, to stutter (Guitar, 2013). Other theories view stuttering as a taught trait resulting from negative external (typically parental) responses to normal childhood dysfluencies (Johnson, 1955). Some people who stutter, on the other hand, may not have inherited any factors that predispose them to stutter. Instead, they might have been exposed to a physical or psychological stress that predisposed them to stuttering or perhaps triggered it. Such events could have occurred before or shortly after birth, and they would be considered congenital factors (Guitar, 2013). Stuttering was often regarded to be mostly a psychological problem. As a result, psychoanalytical techniques and behavioral therapy were used to resolve

any potential neurotic issues (Plänkers, 1999). Learned reactions may play a role in the severity of the condition as it progresses (Guitar, 2013).

2.3 Conventional methods for assessment of Disfluencies

Several methods were used to assess speech characteristics in order to distinguish between stuttering and normal speech. The Iowa scale for assessing the severity of stuttering was one of these techniques (Naylor, 1953). The participants were rated with a scale ranging from 0 to 7. The goal of the individual with stuttering for a certain behavior is required for many of the components on this grading scale.

The most extensively used syllable-based approach for assessing stuttering symptoms is the Stuttering Severity Instrument (SSI) versions 3 and 4 (Riley, 1994; Riley, 2009). To calculate an overall severity score, SSI-3 and SSI-4 incorporate the proportion of stuttered syllables (percent SS) and average duration of the three longest stuttering symptoms, as well as physical concomitants to stuttering noticed at the time of symptom evaluation (Riley, 1994; Riley, 2009). SSI can be used to: (1) diagnose stuttering; (2) track severity changes during and after treatment (Cook et al., 2013; Miller & Guitar, 2009); (3) describe the severity distribution in experimental groups that include persons with stuttering (Howell et al., 2008); and (4) validate other stuttering measures (Howell et al., 2009). Riley's work exemplifies one method of making judgments: disfluency-based analyses.

The improvised version of SSI-3 is SSI-4 (Riley, 2009). Respondents are asked to explain their employment (if employed) or school (if enrolled in school) and to read a short paragraph (or describe pictures if they cannot read). The clinician records the respondent's speech and assigns a score based on stuttering frequency, length, and physical concomitants in four categories. In SSI-4, the following scoring pattern is used: - Frequency is measured as a percentage of syllables stuttering and

converted to a scale of 2–18. The duration is measured to the tenth of a second and converted to a scale of 2–18. Physical concomitants (distracting sounds, face grimaces, head movements, and extremity movements) are scored and converted to scale scores of 0–20. The frequency and duration of stuttering dysfluencies, the score of physical concomitants, and the evaluation of speech naturalness all contribute to a total score. The naturalness of speech is graded on a scale of one to nine, with one indicating highly natural sounding speech and nine indicating highly artificial sounding speech (Riley, 1972). Physical concomitants associated with blocks or attempts to avoid blocking are rated on a five-point scale: 0 - none, 1 - not noticeable unless looking for it, 2 - barely noticeable to casual observers, 3 - distracting, 4 - very distracting, 5 - severe and painful looking (Riley, 2009). This total score is used to assign a verbal descriptor of stuttering severity, ranging from very mild to very severe, based on age-specific population norms (children of preschool age, children of school age, and adults).

2.4 Limitations of conventional methods

SLPs face many constraints when utilizing the SSI-4 to measure the severity of stuttering. Some of these limitations were noted by (Manning, 2009). The significant degree of inter-speaker variability is one of the obstacles in judging fluency. Fluency varies according to the time and place. As a result, the amount and degree of difficulty in each given speaking situation cannot be predicted. Many aspects of fluency condition in young speakers would go unnoticed unless the assessment is conducted in a variety of speaking contexts. Experienced therapists are needed to assess the frequency and duration of stuttering periods perceptually. The perceptual evaluation takes time. Furthermore, the irregularity of stuttering demands ongoing examination throughout multiple assessment or therapy sessions. Limitations

for people with stuttering who take the traditional SSI-4 assessment include their degree of motivation, loss of control due to stuttering, and so on.

2.5 Computerized techniques for assessment of Disfluencies

Yaruss (1999) developed a computer software that counts the frequency of fluent and stuttered syllables, with distinct key strokes denoting different types of stuttering. Work has been undertaken to test the program's reliability and validity. However, this method requires manual entry using specific keys on the keyboard.

TrueTalk, another instrument based on similar concepts, has been employed in Lidcombe's research (Lincoln & Harrison, 1999). TrueTalk Speech Fluency Rater (Synergistic Electronics) is a specialised electronic instrument that measures speech fluency. It has two buttons that show when a fluent syllable is heard and when a stuttering word is heard. A clinician must count syllables and stuttered syllables using the TrueTalk electronic device as part of the Lidcombe program's standard method for determining the frequency of stuttering (Lincoln & Harrison, 1999). Instead of using a recording, the Lidcombe approach enables the clinician to evaluate a child's speech in real time. This reduces the burden on clinicians when working with a person who stutters, as it would be less difficult to make the assessments from a recorded sample than from a live one (Bakker et al., 1995). If the assessments were performed later, the clinician would be free to focus on ancillary behaviors, such as physical concomitants.

SSI-4 comes with a programme called Computerized Scoring of Stuttering Severity version 2 (Bakker & Riley, 2009). It is used to count syllables as well as the frequency and length of stuttering. Clicking the left and right buttons of a mouse, respectively, indicates correct syllables and stuttering syllables (Riley, 2009). ‘Holding down the key’ is used to signal the duration of the stutter and is utilised to

calculate its duration.

All these techniques use computers/ softwares for measurement of frequency and duration of the stuttering events. In all these systems the stuttering events are identified manually by the speech language pathologist and fed to the computer. Total dysfluencies are further computed. Most of the methods discussed above are semiautomatic in nature, which requires manual intervention most of the time during the test. Hence, these methods carry forward most of the limitations encountered in conventional methods.

2.6 Detection of stuttering events using automatic speech processing

Many researchers (Liu et al., 2006; Salesky et al., 2019; Wu & Yan, 2004) have worked on detection of stuttering events using automatic speech processing. These researchers have used mainly one of the two methods:- i. Post processing after the output of automatic speech recognizer or ii. Pre-processing before the automatic speech recognizer. Riad et al. (2020) used log-energy Mel scale filters to detect stuttering events using support vector machine (SVM) and Deep Neural Network (DNN) classifiers. Kaushik et al. (2010) detected stuttering events using formant information and nasality effect They used signal processing techniques to automatically identify and eliminate repetitions as well as filled pauses from it. The algorithms demonstrate notable increases in word recognition accuracy and consequent decreases in substitution, deletion, and insertion errors when tested using Dragon naturally speaking speech recognizer.

In a study on the classification of childhood dysfluencies given by Geetha et al. (2000), artificial neural networks (ANNs) were utilized. They predicted normal, non-fluency, and stuttering with 92% accuracy. They came to a conclusion that the children with disfluencies can be divided into normal non fluency and stuttering

groups based on Disfluency Assessment Procedure for Children (DAPC) based on their observations of the ANN analysis results. And thus ANN can be created as a practical clinical tool to objectify the diagnostic processes.

Ravikumar et al. (2008) developed an automatic detection method based on Support Vector Machine (SVM) to differentiate between fluent and dysfluent speech. The accuracy of the system was 94.35%. Surya and Varghese (2016) proposed three methods to recognize the stuttered speech such as (a) Supervised model for stuttered speech recognition (b) Stuttered speech recognition by stuttering pruning (c) Automated text-to-speech based stuttered speech recognition. The first method involved two phases - testing and training. The audio array is created by converting 'N' audio streams, which is used to extract the MFCC features. The second method involves (i) Convert a speech sample to an audio signal with amplitude and time (ii) Get the highest possible speech amplitude (iii) Use the neural network to calculate a threshold value using the maximum amplitude (iv) Divide the audio samples into discrete, equal-length frames (v) Analyze each frame and replicate it if the frame's maximum value is larger than the threshold value (vi) Pass the signal to a voice recognition module once all frames have been analysed.

In the third method, the words are converted into equivalent texts using this method. Each letter in the speech is identified using sophisticated ANNs. ANNs are trained to predict vowels and consonants terms in a speech using intelligent guessing. This ANN analyses the input speech and generates equivalent texts based on its training experience. The text is then fed into a dictionary, which selects the most closely related word. This method eliminates any stuttering speeches. The full text-based speech is then converted back to speech. To create the text corresponding audio for each utterance, the reversion process employs machine learning techniques once

again. As a result, we can have clear, stammer-free speaking. They have implemented the above mentioned methods and its accuracy was measured. The initial technique used a classifier model to recognize stuttering in speech. Support vector machine model was used in the classifier. They achieved 76% accuracy in correctly classifying the words. The accuracy of the neural network-based speech correction method was 62%. The final technique, which attempted to identify the stuttering events by converting it into corresponding text, obtained an accuracy of 80%. They found that the accuracy can be improved using more training data and considering other features for identifying the stuttering episode.

Mendhakar and Mahesh (2018) attempted to develop a tool to automatically segment audio recordings into silences and chunks of speech corpus and then to assess the same using speech recognition method to obtain a time annotated speech transcript. The features of those speech clusters obtained was analyzed further using HMM models on MATLAB platform. The results of the study showed a high recognition rate (90%) with a time annotation difference of 0.5s from that of PRAAT analysis and MATLAB transcript.

Gupta et al. (2019) introduced a new method for analysing stuttering speech that included feature extraction with the Weighted Mel Frequency Cepstral Coefficient (WMFCC) and classification with a Bi-directional Long-Short Term Memory neural network. WMFCC outperformed the other feature extraction methods in the testing, achieving an average recognition accuracy of 96.67 percent.

2.7 Limitations of the previous research in automatic detection of stuttering events

It has been observed that most of the researchers have implemented their algorithms on online available datasets and not on real-time data. This results in large

variation in accuracy when analysed on real data. (Gupta et al., 2019). In the context of Indian languages including Indian English, very few researchers have attempted to automatically detect the stuttering events. Audhkhasi et al. (2009) introduced a formant based thresholding system to detect filled pause in Indian English. Kaushik et al. (2010) attempted to detect filled pause and word repetition using the first four formants to identify the stuttering event from a dataset of 60 sentences of Indian English. Veda (2021) made an attempt to detect the stuttering events automatically from Kannada speakers. The findings revealed that all durational fluency parameters had 100% agreement between perceptual evaluation and automatic speech recognition. Also it was discovered that the error in detection could range from 26 to 35 percent. Since each language is different in their spectral and temporal characteristics there is a need of developing an automatic speech processing based procedure for assessment of stuttering in different languages. No such attempts have ever been made in Malayalam language. Thus the present study attempts to detect Stuttering events in Malayalam Speaking Adults through an automatic speech processing based procedure.

CHAPTER III

METHOD

The aim of the study was to create and test a virtual mode approach based on automatic speech processing for assessing stuttering in Malayalam-speaking adults.

3.1 Participants

A total of 45 native literate Malayalam speakers between 18-35 years of age were recruited in the study. The participants were divided into two groups, Group I included 30 normal adults, and Group II included 15 adults with stuttering.

3.1.1 Inclusion criteria for selection of both Group I & II

By completing a systematic interview, those with normal or corrected vision and with no neurological abnormalities, social, emotional, cognitive, or mental issues were included. Participants with a smart phone that costs between Rs 10,000 and Rs 20,000 with the Zoom app installed were considered.

3.1.2 Inclusion criteria for selection of normal participants (Group I)

Those who have normal speech and language skills as well as hearing sensitivity were included. This was ensured by adopting the "WHO ten question disability screening checklist" (Singhi et al., 2007) which screens disabilities related to developmental milestones, vision, hearing, comprehension, movements, seizures, learning, lack of speech, unclear speech, and slowness to rule out any previous speech-language, sensory, motor, or cognitive difficulties.

3.1.3 Inclusion criteria for selection of participants with stuttering (Group II)

Those individuals with mild to severe stuttering since childhood and diagnosed by a Speech Language Pathologists using the SSI-4 (Riley, 2009), were recruited, prior to treatment.

3.2 Material

The standardized passage (Annexure-A) in Malayalam (Savithri & Jayaram, 2004) consisting of 100 words (384 syllables) was used for assessing stuttering from the reading task through automatic speech processing and through perceptual evaluation. Spontaneous speech was also assessed using SSI-4 through virtual mode.

The text of Brahmin passage was prepared in 'kartika' font in three font sizes (18,20,22). A single space was used between the words and 1.5 line spacing was used between the lines. The text was arranged in six paragraphs. Only one paragraph was displayed at a time to the participant. Based on the validation by three SLPs, the passage with font size of 22 and 1.5 line spacing was chosen for the study.

3.3 Procedure

3.3.1 Recording of read passage through Zoom app

Before beginning the recording, the researcher created a video conference through zoom app with the participants (both Groups I and II), to ensure that they were comfortable and provided clear instructions to them. Each participant was advised to sit comfortably in a quiet room with their cell phone at least 10 cm away from their lips. They were instructed to read the standardized passage displayed on their mobile phones. The conference was recorded using the Zoom app by the researcher. All of the participants were instructed to read the standardized Malayalam passage aloud at a comfortable loudness. For subsequent analysis, the recorded samples were saved to a laptop or PC. Audio samples were retrieved from the laptop and converted to '.wav' format utilizing an online audio-video recording conversion service. The recorded samples were then stored in.wav format.

3.3.2 Online assessment through Zoom platform using SSI-4

The SSI-4 was used to evaluate each participant in Group II via the Zoom platform. The researcher initially developed a rapport with the participant and made them feel at ease. The participant was asked to read the Malayalam standardised passage that was delivered via Zoom utilising the screen share feature. The researcher made sure that the participant was satisfied with the visual quality and readability of the paragraph before collecting data. Three tasks were included in the SSI-4 assessment protocol: a) Job task- Conversation, spontaneous speech and narration, b) Reading task - The standardized Malayalam passage (Annexure-A). The researcher used the symbols (.) and (/) to indicate whether a syllable was fluent or not. The researcher observed secondary behaviours of physical concomitants (distracting sounds, facial grimaces, head movements, and extremity movements) while the subject was reading or executing a job task (Riley, 2009). The researcher recorded the speech and scored the individual in four categories: stuttering frequency, stuttering duration, and physical concomitants to find out the severity of stuttering.

3.3.3 Inter judge reliability

Two other judges also evaluated the dysfluency parameters such as filled pause, Word repetition, part word repetition, prolongation, and block of all the 15 persons with stuttering. These judges were SLPs with more than 5 years of expertise in stuttering evaluation. SLPs were not informed about the study's goal. The researcher's assessment and agreement with the independent judges were compared using cronbach’s alpha values to assess inter judge dependability.

3.3.4 Assessment of technical and clinical quality of the online session by the participant

Each participant of group II rated their satisfaction with the technical and clinical quality of the online session on a three-point scale: '3'- highly satisfied, '2'- somewhat satisfied, '1'- not at all satisfied" at the end of each session. The following factors were considered in this evaluation: image and sound quality, as well as the quality of contact between the participant and the researcher (Sicotte et al., 2003).

3.3.5 Rating of technical and clinical quality of the online session by three SLPs

Three SLPs judged the quality of each recorded session on a six-item, five-point scale where, '1' is highly dissatisfied and '5' is highly satisfied. The rating was done based on the technical quality as well as clinical quality. Three aspects of technical quality were evaluated: sound quality, signal reception latency, and image quality. Clinical quality was also examined using three criteria: degree of control over the participant throughout the session, achievement of clinical goals, and participant compliance with the researcher's instructions (Sicotte et al., 2003).

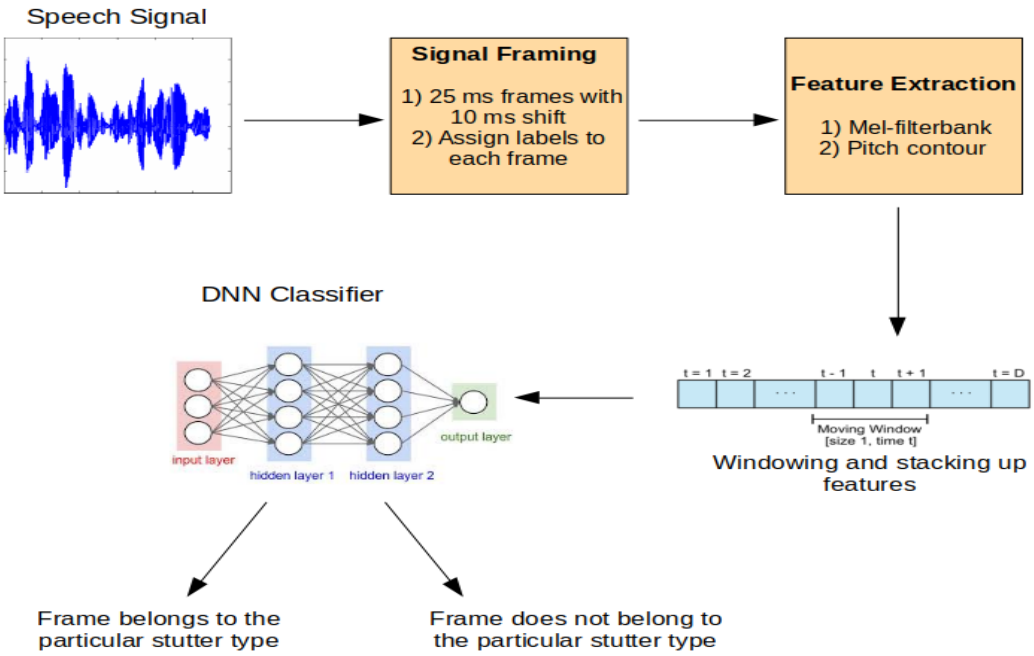
3.3.6 Assessment of stuttering through automatic speech processing

Figure 3.1 depicts the automatic speech processing procedure for recognizing dysfluencies in stuttering speech from recorded audio. A Deep Neural Network (DNN) that had been trained was employed. This model determines whether or not a specific type of dysfluency is present in a 10ms speech frame. The time stamps for dysfluencies can be determined in an audio file using this method. In a 25 ms window with a 10 ms shift, Mel-filter bank features were retrieved. Using the Voice Activity Detection (VAD) information, the Mel-filter bank characteristics were then

normalised (mean variance). Fundamental frequency was retrieved in a window of 50 ms and sampled every 1.8 ms for intonation.

Figure 3.1

Block schematic showing the process for automatic assessment of stuttering parameters



The final feature vector for each frame was created by stacking features from three frames before and three frames after the centre frame. After that, a binary classifier was trained to determine whether or not a specific form of dysfluency is present in the 10 ms audio frame. DNN classifier with 2 hidden layers was considered for disfluency detection. The number of hidden units in each layer were 100 and 50, respectively, with sigmoid activation function after each layer. The Adam optimizer was used. Hyperparameter tuning was performed as well for training the DNN. An optimal learning rate of 0.001 and an optimal batch size of 32 was used here. The algorithm evaluates the frequency and duration of the stuttering moments after determining the kind of dysfluency and the dysfluent syllable.

3.3.7 Parameters assessed

- Frequency: Frequency of stuttering is assessed as the percentage of words stuttered which is equal to number of stuttered moments per 100 words.
- Duration: Determined by measuring the time interval of three of the longest stuttering instances and then calculating their mean duration.

3.4 Analysis

- Speech recordings from Group II participants were run through a Matlab-based code (created by a Matlab expert) to automatically determine the frequency and duration of prolongations, filled pauses, blocks and repetitions. The normative data for this was provided by the recordings of speech from participants in Group I.
- The researcher studied the recorded speech of Group II participants to determine the frequency and duration of prolongations, filled pauses, blocks and repetitions. Three experienced SLPs validated the researcher's findings.
- Zoom recording provided the audio data, which was converted to a .wav file. The PRAAT software was then used to extract the wave file. The Each syllable's beginning and ending points were documented. The researcher utilised the following codes for perceptual evaluation to determine if the syllable is fluent or dysfluent, '0' - Clean speech, '1'- Filled pause/Interjection, '2'- Word repetition, '3' - Part word repetition, '4' – Prolongation &'5' - Block. The frequency and duration of prolongations, filled pauses, and repetitions estimated using Matlab-based code were compared to the results of the researcher's perceptual evaluation, and the difference was analyzed.
- Results of the perceptual evaluation and automatic speech recognition results were statistically analysed using the Statistical Package for the Social Sciences (SPSS)

software (Version 21.0). Descriptive statistics was carried out to obtain the mean, median, and standard deviation for both the groups. The normality was tested using the Shapiro Wilk test. A non-parametric analysis was used because the data was not normal. Then Wilcoxon Signed Ranks Test was used to compare the results of perceptual judgement with automatic speech recognition.

- The accuracy of fluency parameters derived through automatic speech processing was investigated at various levels of severity.

3.5 Ethical consideration

All participants were informed of the study's procedures and purpose, as well as the research's goal and objectives. Before the study, their safety and privacy were assured, and a written consent was obtained.

CHAPTER IV

RESULTS

The purpose of the current study was to develop an automatic speech processing based procedure in virtual mode for assessing stuttering in individuals who speak Malayalam. Each of the 15 adults who stutter had their frequency, duration, and severity of stuttering occurrences as well as their severity of stuttering assessed using the SSI-4 through virtual mode. The frequency and length of filled pauses, repetitions, blocks and prolongations were also assessed using automatic speech processing of the recorded standard Malayalam passage (Annexure-A). The normative speech for automatic speech processing was also the standard Malayalam passage (Annexure-A) spoken by 30 normal individuals which is recorded using the Zoom application. The objective was to compare the values of the dysfluency characteristics obtained through online perceptual assessment to those obtained through automatic speech processing.

4.1 Characteristics of participants

4.1.1 Characteristics of normal participants

Thirty Malayalam speakers who were adults, literate, and with normal hearing, vision, communication, and cognition took part in the study. Table 4.1 displays these participants' characteristics.

Table 4.1

Details of normal participants

Sex	No of participants	Mean age
Male	15	22.06
Female	15	23.33
Overall	30	22.70

4.1.2 Characteristics of participants with stuttering

In the current study, 15 native Malayalam speaking adult, literate persons with mild to severe stuttering who had normal language skills and no history of hearing, vision, or other communication issues participated. Additionally, everyone who participated knew how to use Zoom for video calls. Participants who displayed indications of psychiatric, social, emotional, or neurological issues were not allowed to participate in the study. Table 4.2 outlines the characteristics of these participants.

Table 4.2

Details of participants with stuttering

Severity	Male	Female	Mean age
Very mild	2	0	24.00
Mild	4	0	26.25
Moderate	4	0	27.25
Severe	4	1	27.00

4.2 Online assessment of the severity of stuttering through perceptual evaluation using SSI – 4

4.2.1 Severity assessment

With the help of the Zoom platform in virtual mode and the SSI-4, the researcher evaluated the severity of stuttering in fifteen persons who stutter (Table 4.3). Two were observed to have very mild and four to have mild stuttering diagnoses. Four participants had moderate, while five participants had severe stuttering. Three SLPs confirmed the researcher's diagnosis, and the validation results were wholly consistent with the researcher's diagnosis.

Table 4.3

SSI-4 scores of each participant, percentile ranks and severity

Participant	SSI-4 scores	Percentile rank	Severity
1	26	41-60	Moderate
2	33	78-88	Severe
3	34	78-88	Severe
4	32	78-88	Severe
5	26	41-60	Moderate
6	10	1-4	Very mild
7	34	78-88	Severe
8	14	5-11	Very mild
9	18	12-23	Mild
10	20	12-23	Mild
11	27	41-60	Moderate
12	26	41-60	Moderate
13	20	12-23	Mild
14	18	12-23	Mild
15	32	78-88	Severe

4.3 Assessment of frequency and duration of dysfluency parameters

Through online communication, the researcher assessed each participant's stuttering severity using the SSI-4 for spontaneous speech and reading passage by measuring the frequency and duration of prolongations, filled pauses, blocks and repetitions from the recorded passage. Each participant's stuttering episodes were coded as indicated below:

0 - Clean speech,

1- Filled pause/Interjection,

2-Word repetition,

3 - Part word repetition,

4 - Prolongation,

5- Block

Table 4.4

Frequency and duration of filled pauses, repetitions, prolongations and blocks of participants with stuttering assessed through online perceptual evaluation

Participant	Stuttering severity	Filled pauses		Part word repetitions		Prolongations		Blocks		Word repetitions		Duration in seconds
		No.	Freq	No.	Freq	No.	Freq	No.	Freq	No.	Freq	
1	Moderate	0	0	2	2	3	3	5	5	6	6	8
2	Severe	0	0	0	0	0	0	11	11	0	0	8
3	Severe	0	0	10	10	13	13	26	26	3	3	8
4	Severe	0	0	13	13	7	7	12	12	3	3	8
5	Moderate	1	1	10	10	6	6	17	17	0	0	6
6	Very mild	0	0	0	0	0	0	2	2	2	2	4
7	Severe	7	7	1	1	0	0	22	22	4	4	8
8	Very mild	2	2	0	0	4	4	12	12	0	0	2
9	Mild	1	1	4	4	0	0	3	3	0	0	6
10	Mild	5	5	4	4	1	1	0	0	5	5	4
11	Moderate	1	1	8	8	1	1	36	36	3	3	8
12	Moderate	3	3	0	0	0	0	8	8	6	6	6
13	Mild	0	0	1	1	0	0	10	10	9	9	4
14	Mild	0	0	7	7	0	0	1	1	3	3	4
15	Severe	0	0	5	5	5	5	13	13	3	3	8

For each of the 15 individuals, the findings of a perceptual evaluation using the SSI-4 are presented in Table 4.4. This assessment measures the frequency and duration of stuttering events. The frequency and number of the following stuttering parameters-prolongations, blocks, part-word repetitions, word repetitions and filled pauses were assessed. The number of stuttering moments per 100 words, or the

percentage of syllables that stutter, is expressed as stuttering frequency. Participant number 7 with severe stuttering had the highest frequency (7) of filled pauses, and participant number 13 with mild stuttering had the highest frequency (9) of multi-syllable word repetition. The participant with severe stuttering, No. 4, displayed the highest frequency (13) of part-word repetitions. The participant with severe stuttering, No. 3, had the highest frequency (13) of prolongations. The participant with the moderate stuttering, participant no. 11, had the highest frequency (36) of blocks. The technique described in method section was used to calculate frequency. The lengths were expressed in seconds. The Average length of the 3 longest stuttering events was considered for the duration measurement. The SSI-4 duration scale was applied.

The researcher's report of stuttering instances was compared against the opinions of two separate judges who were not informed of the experiment's goal. Between the two sets, there was a 90% correlation.

4.4 Assessment through automatic speech processing

The fluency parameters for each of the 15 stuttering participants, as determined by automatic speech recognition from the recorded standard passage, are shown in Table 4.5. Similar to perceptual examination, participant no.7 with severe stuttering displayed highest frequency (3) of filled pauses, participant no. 4 with severe stuttering displayed the highest frequency (6) of part-word repetitions, Participant no.3 with severe stuttering displayed highest frequency (9) of prolongation, participant no.11 with moderate stuttering displayed highest frequency (16) of blocks and participant number 10 with mild stuttering displayed the highest frequency (3) of word repetitions, , where all of them were in agreement with the findings of the perceptual evaluation except word repetition.

Table 4.5

Frequency and duration of filled pauses, repetitions, blocks and prolongations of participants with stuttering assessed through automatic speech processing

Parti-cipant	Stuttering severity	Frequency					Duration in seconds
		Filled pauses	Part word repetitions	Prolongations	Blocks	Word repetition	
1	Moderate	0	1	2	2	2	8
2	Severe	0	0	0	3	0	8
3	Severe	0	4	9	12	1	8
4	Severe	0	6	4	7	2	8
5	Moderate	0.40	5	2	7	0	6
6	Very mild	0	0	0	1	1	4
7	Severe	3	0.24	0	8	1	8
8	Very mild	1	0	2	8	0	2
9	Mild	0.32	2	0	1	0	6
10	Mild	2	1	0.37	0	3	4
11	Moderate	0.43	2	0.28	16	1	8
12	Moderate	1	0	0	4	2	6
13	Mild	0	0.41	0	2	2	4
14	Mild	0	2	0	0.36	1	4
15	Severe	0	1	1	5	1	8

4.5 Comparison of results of automatic speech processing and online perceptual evaluation

Wilcoxon Signed Ranks Test was used to compare the outcomes of automatic speech processing and perceptual evaluation. Results indicate that there was no significant difference among various types of dysfluencies in mild severity. The z value obtained for each of the dysfluency type were as follows: prolongation (z= 1.000), part-word repetition (z= 1.826) Filled pauses (z = 1.342) word repetition (z= 1.604), and block (z= 1.604) ($p > 0.05$). Similarly for moderate severity also there was no significant difference seen among different types of dysfluencies such as filled

pause (z= 1.604), part word repetition (z=1.604), prolongation (z=1.604), word repetition (z=1.604), and block (z= 1.826) ($p>0.05$). For the severe group also similar pattern of results were observed and thus there was no significant difference among various dysfluencies such as filled pause (z=1.000), part word repetition (z= 1.826), prolongation (z= 1.604), word repetition (z= 1.826) ($p>0.05$) and block (z= 2.023) as the p value was not significantly low ($p>0.05$).

Each participant's accuracy for each of the fluency parameters obtained by automatic speech processing was computed using a formula. The formula used for finding out the percentage of accuracy in part-word repetition is given below:

Percentage of accuracy in part-word repetition = {(number of part-word repetition in perceptual assessment -number of part-word repetition in automatic speech processing) / number of part-word repetition in perceptual assessment} x 100

The same formula was used for finding out the percentage of accuracy for other fluency parameters also.

The accuracy of fluency parameters generated through automatic speech processing are displayed in Table 4.6, for stuttering at different severity levels, such as mild, moderate, and severe. The subject with the highest accuracy (66%) for filled pause was participant 4 (severe). whereas the participant number 1 (moderate) was found to have the highest accuracy (59%) in part word repetition. The subject with the highest accuracy (66%) for prolongations was participant 3 (severe). Whereas the participant number 5 (moderate) was found to have the highest accuracy (68%) for word repetition and the subject with the highest accuracy (65%) for block was participant 8 (very mild).

Table 4.6

Accuracy in assessment of number of stuttering events through automatic speech processing across participants

Participant	Stuttering severity	Accuracy				
		Filled pauses	Prolongations	Part word repetitions	Blocks	Word repetitions
1	Moderate	51%	56%	59%	35%	36%
2	Severe	17%	27%	24%	24%	19%
3	Severe	49%	66%	43%	46%	38%
4	Severe	66%	63%	49%	55%	55%
5	Moderate	40%	36%	51%	43%	68%
6	Very mild	38%	53%	55%	47%	64%
7	Severe	39%	41%	25%	37%	26%
8	Very mild	57%	53%	33%	65%	51%
9	Mild	33%	31%	38%	22%	51%
10	Mild	42%	38%	22%	44%	61%
11	Moderate	43%	29%	29%	45%	41%
12	Moderate	38%	50%	21%	48%	41%
13	Mild	41%	48%	41%	24%	27%
14	Mild	32%	44%	24%	37%	26%
15	Severe	31%	23%	23%	36%	34%

Table 4.7

Overall accuracy in assessment of number of stuttering events through automatic speech processing across different stuttering events

Stuttering event	Identified through SSI-4	Identified through automatic speech processing	Percentage of accuracy
Filled pauses	20	8	41%
Prolongations	40	18	44%
Part word repetitions	65	23	36%
Blocks	178	73	41%
Word repetitions	47	20	43%

The following formula was used to calculate the overall accuracy in each of the fluency parameter derived through automatic speech processing;

Overall percentage of accuracy in part-word repetition = {(Total number of part-word repetition in perceptual assessment –total number of part-word repetition in automatic speech processing) / number of part-word repetition in perceptual assessment} x 100

Table 4.7 shows the overall accuracy in fluency parameters derived through automatic speech processing. The highest accuracy (44%) was seen in prolongations, whereas the lowest accuracy (36%) was observed in part-word repetitions. Occurrences of filled pauses and word repetition and blocks were less compared to prolongation.

4.6 Technical and clinical quality of the online assessment of persons with stuttering

4.6.1 Technical quality based on assessment by the SLPs

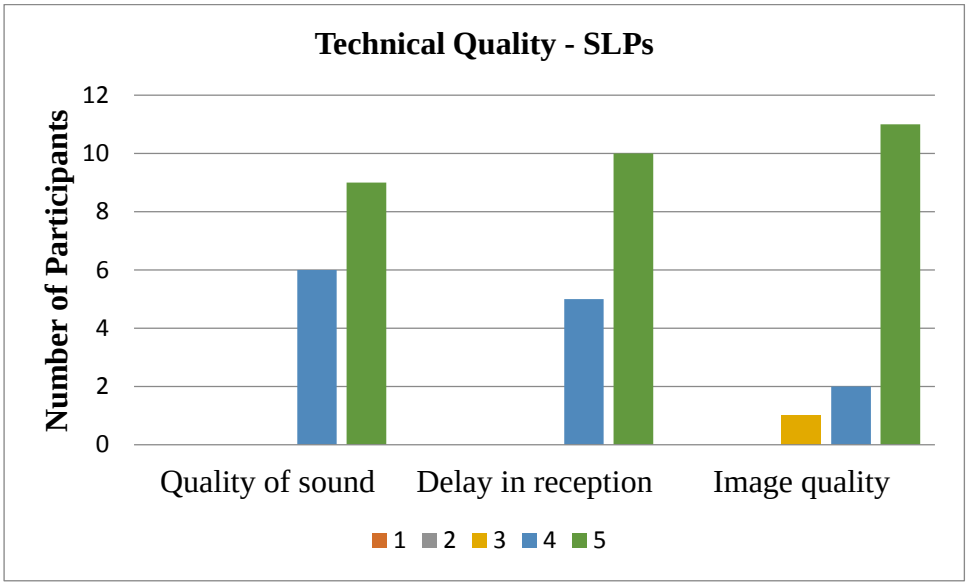
Figure 4.1 shows the results of the technical quality assessment performed by three SLPs after viewing the session's recorded video for each participant. A five-point rating scale was used for the evaluation, with a score of ‘1’ denoting high dissatisfaction and a score of ‘5’ denoting high satisfaction.

Three criteria were used to assess technical quality: image quality, reception delay, and sound quality.

Nine participants reported being highly satisfied with the sound quality (rating of 5) while all other participants reported being satisfied (rating of 3 or higher). Ten individuals rated the delay in reception as highly satisfied (five), while the assessment for all other participants was satisfactory (three or higher). Eleven participants received a rating of highly satisfied (five), while other three participants received a rating of satisfactory (three or higher).

Figure 4.1

Technical quality based on evaluation by SLPs on a 5 point scale (1 = highly dissatisfied, 2 = somewhat satisfied, 3 = neutral, 4 = satisfied and 5 =highly satisfied)

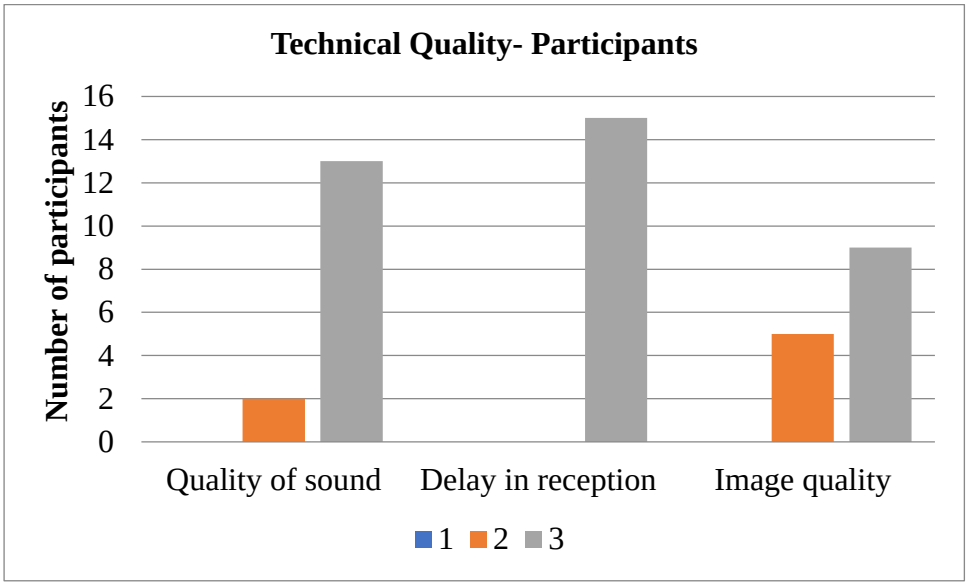


4.7 Technical quality based on assessment by the participant

At the end of the online session, participants with stuttering evaluated the technical quality, as shown in Figure 4.2. It was based on a three-point rating scale, with a score of ‘3’ representing a high level of satisfaction, a score of ‘2’ representing a moderate level of satisfaction, and a score of ‘1’ representing a dissatisfaction. Thirteen out of the fifteen participants rated the sound quality as "highly satisfied" (3) two participants as "somewhat satisfied" (2). Nine individuals evaluated the image quality as "highly satisfied" (3), while the remaining five participants gave it a "somewhat satisfied" rating (2). And all the participants evaluated the delay in perception as highly satisfied (3).

Figure 4.2

Technical quality based on evaluation by participants on a 3 point scale (3 = highly satisfied, 2 = somewhat satisfied, 1 = not at all satisfied).

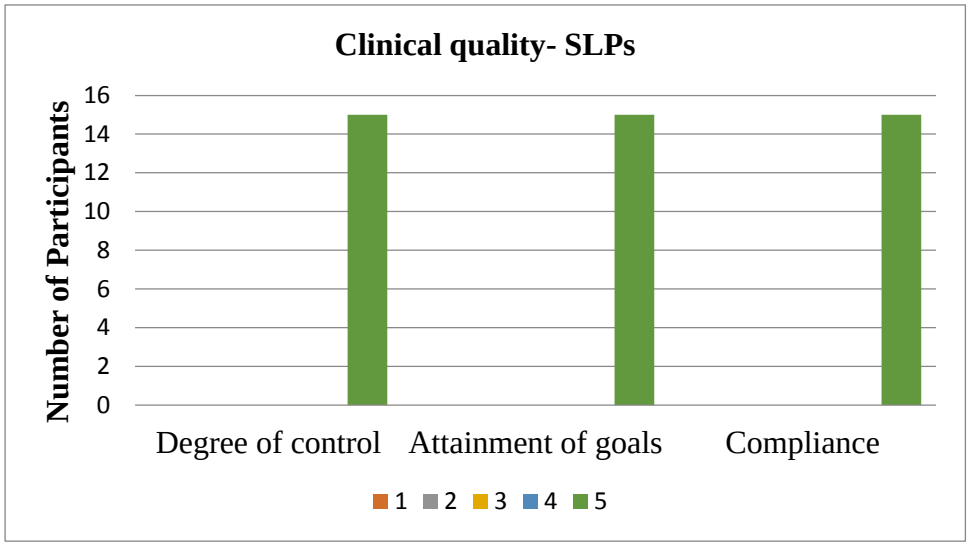


4.6.2 Clinical quality based on assessment by the SLPs

Figure 4.3 shows the results of the analysis, of the session videos of each participant, by three SLPs. The evaluation was based on a five-point rating scale, where "1" denotes a high degree of dissatisfaction and "5" denotes a high level of satisfaction with regard to degree of control, target attainment, and compliance, all the sessions of the fifteen participants were rated "very satisfied" (5) for all the 3 aspects - objective attainment, degree of control and compliance.

Figure 4.3

Clinical quality based on evaluation by SLPs on a 5 point scale (1 = highly dissatisfied 2= somewhat satisfied 3 = neutral 4= satisfied and 5 = highly satisfied)

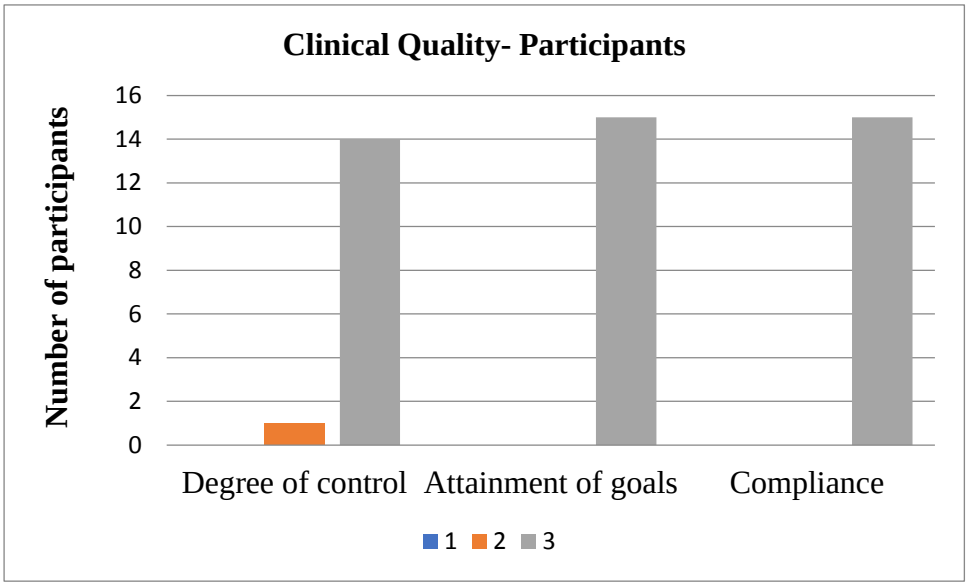


4.6.4 Clinical quality based on assessment by the participant

Results of the clinical quality assessment completed by the participants at the end of the online session is depicted in Figure 4.4. Regarding degree of control, target accomplishment, and compliance for all fifteen participants with stuttering, assessment was done on a three-point rating scale, where "3" indicates "very satisfied," "2" indicates "somewhat satisfied," and "1" indicates "not at all satisfied. All of the fifteen participants scored "highly satisfied," for attainment of goals and compliance, fourteen of them rated the degree of control as highly satisfied and one participant rated as somewhat satisfied.

Figure 4.4

Clinical quality based on evaluation by participants on a 3 point scale (3 = highly satisfied, 2 = somewhat satisfied, 1 = not at all satisfied)



Overall, the results of the current study show that there is no significant difference between the assessment of stuttering events through perceptual evaluation and automatic speech processing. The technical and clinical quality of the online method of evaluation was found to be satisfactory for all the participants as well as per the rating given by the speech language pathologists.

CHAPTER V

DISCUSSION

The following queries were addressed in the current study:

1. Is it feasible to determine the frequency and duration of filled pauses, repetitions, prolongations, and blocks using automatic processing of recorded speech?

2. Is there a concurrence between the values of fluency characteristics obtained through online perceptual assessment and those obtained through automatic processing of recorded speech?

3. Does the accuracy of the fluency parameters obtained through automatic speech processing depend on levels of severity of stuttering?

4. Is the online assessment's clinical and technical quality acceptable to the clinician as well as the client?

5.1 Selection of participants and their characteristics

In Group I, a total of 30 literate, native Malayalam speakers who were adults (mean age = 22.7 and SD = 2.23) took part in the study. Through an interview and screening/testing procedures, 15 male and 15 female participants were chosen. The individuals had no issues with hearing, otology, vision, communication, or cognition as well as no neurological issues. They also had good comprehension of speech. In Group II, 15 native Malayalam-speaking adult, literate participants (14 male and 1 female; mean age = 26.46; SD = 3.40) who had mild to severe stuttering, according to certified speech language pathologists, took part in the study. Through a systematic interview, it was ascertained that all of the participants knew how to utilize the Zoom application for video calls. The researcher also confirmed that none of the subjects displayed signs of any neurological, social, emotional, or psychiatric disorders through systemic interview.

5.2 Feasibility to assess the frequency and duration of fluency parameters through automatic speech processing

In general, majority of the methods for detecting speech disfluency fall into one of the following categories: one which employs text-based features along with speech for disfluency detection (Lu et al., 2019; Zayats et al., 2016) or the other one by using signal level methods to identify them (Hamzah & Jamil, 2019; Salesky et al., 2019). While the former approach is more effective and produces encouraging results, it is computationally more expensive and prone to errors (Mehrotra et al., 2021). The signal based methods were utilized through a Matlab-based programme in the present study.

The online recorded passage, spoken by participants of Group II, was run through the Matlab-based programme to obtain the frequency and duration of the fluency characteristics. The system correctly identified many of the stuttering episodes and there was good agreement in terms of duration of the fluency parameters. However, the maximum accuracy towards frequency detection was reported to be 68%. Riad et al. (2020) detected stuttering events using log-energy Mel scale filters, support vector machine (SVM) and deep neural network classifiers (DNN). Similarly, the current study used DNN classifiers based models to detect the stuttering events. The overall accuracy of 36% to 44% (Table 4.7) obtained with the limited number of samples, indicate the feasibility to use automatic speech processing to assess the frequency and duration of fluency parameters.

5.3 Concurrence between the values of fluency parameters derived through automatic speech processing and the values obtained through perceptual assessment

The frequency and duration of the dysfluency parameters showed no significant difference between perceptual assessment and automatic speech processing. This was similar to the results of the study done by Veda (2021) on Kannada speakers, where they found no significant difference between perceptual evaluation and automatic speech recognition.

Calculation of the accuracy percentage by keeping perceptual evaluation as the reference, allowed comparisons to be made between perceptual assessment and automatic speech processing. The accuracy percentage was computed for each of the stuttering occurrences, including pauses, repetitions (part words and words), prolongations and blocks. It was observed that the accuracy ranged from 36% to 44%. Ravikumar et al. (2009) used SVM based automatic detection method where they did the assessment on 15 adults who stutter to discriminate fluent and non fluent speech and obtained 93.45% accuracy. Surya and Varghese (2016) identified stuttering events by converting it into corresponding text using support vector machine and neural network based speech correction. 76 percent accuracy was reported using SVM and 62% accuracy using DNN based speech correction method. In the present study, the accuracy was observed to be 41%, 44%, 36%, and 41%, respectively for filled pauses, prolongations, part-word repetitions, word repetitions, and blocks.

With reference to the Table 4.7, though 178 events of blocks were identified through perceptual evaluation using SSI-4, only 73 of them could be detected in automatic speech recognition. The ASR used in the study was based on a deep neural network (DNN) that contains multiple layers of computing units (usually convolution

units) between the input and output layers. All DNN's consist of the following components: neurons, synapses, weights, biases, and functions. DNN perform very well and give very high accuracy when trained on large amounts of data. These networks are trained using supervised machine learning techniques and using methods like back propagation to compute the weights of the network. Once trained on a large dataset, these models perform with very high accuracy, especially for classification problems. In the present study, the samples pertaining to only 15 Kannada speakers could be analysed due to the limited time. Only a total number of 18 prolongations, 73 blocks, 23 part word repetitions, 20 word repetitions, and 8 filled pauses were available to the DNN.

Block is a stuttering event in which sound or air flow persists while articulator movement is halted which are indicated by silence region. In many instances in the present study, the ASR was considering this silence as the usual silence between the words and hence, some of the blocks were missed. This can be improved by training the ASR with more samples. Moreover, in perceptual evaluation, physical concomitants are also observed while identifying the blocks. As the physical concomitants were not taken into consideration in ASR, some of the blocks must have missed.

In part-word repetition and word repetition, the ASR might have wrongly identified the repeated word or segment as a true word. This error in identification was more evident in participants with severe stuttering. This also could be improved by improving the performance of ASR by training it with more samples.

Tottie (2014) observed that, in British English, the mainly observed filler words in filled pauses were either 'umm' or 'uhh'. In the present study, the ASR could recognize most of the 'uhh' filler words, whereas the recognition was poor in

instances of ‘umm’. This may be because the ASR was wrongly identifying ‘umm’ as noises, as the energy of the filler word ‘umm’ is very less compared to filler word ‘uhh’. In the present study there were 13 instances of ‘uhh’ and 7 instances of ‘umm’ which identified through perceptual evaluation. If the ASR is trained with more samples of filled pauses with filler words ‘umm’, it will help to recognize ‘umm’ also as a filled pause and thereby improve the accuracy.

In prolongations, 16 events of prolongation of consonants and 24 events of prolongation of vowels were identified through perceptual evaluation. In prolongations of consonants, (for eg: /n̪eː…ɾiʈeːɳʈi/, kʂaɲiː…t͡ʃu/ etc), the ASR was not recognizing them as prolongations, may be due to the same reason as observed in the case of filled pauses. The consonant prolongation is wrongly recognized by the ASR as noises. This may also be improved, if more samples are available to train the ASR.

5.4 Dependence of accuracy of fluency parameters on severity level of stuttering

In the current study, the statistical analysis also showed that there was no significant difference between the detection of stuttering events even among different severity groups such as mild, moderate and severe. This was also in agreement with Veda (2021).

5.5 Technical and clinical quality of the online assessment

The participants and three skilled SLPs evaluated the technical and clinical quality of the online sessions. For quality assessment, the SLPs watched videos of the sessions that were captured using Zoom. The SLPs' evaluation of technical quality was based on three factors: sound quality, reception latency, and image quality. The rating was given on a five-point scale, with a score of ‘5’ denoting high satisfaction and a score of ‘1’ denoting high dissatisfaction. For all participants, the mean rating

for each of the three factors was "3" or higher. This shows that the sound and image quality was acceptable, and no unsatisfactory signal reception delays were noted. The SLPs also assessed three components of clinical quality, including degree of control, goal attainment, and compliance. For all participants, the mean ratings were "3" or higher for each of the three elements. This demonstrates the researcher's high level of control, accomplishment of the intended aims, and compliance. Sicotte et al. (2003) reported similar findings after having evaluated 6 adult participants for technical and clinical quality on 6 items using the same rating system as this study. The SLPs rated the technical quality as being moderately good. On a five-point scale, 50% of the sessions received ratings of 3 or higher, and 43% of sessions received ratings of 4 or higher for technical quality. The SLPs in the current study also rated the technical quality as being good, and overall, the sound quality of 40% of the participants received ratings of 4, 60% of the participants were rated as highly satisfied (5). 33% received ratings of 4 in ‘delay in reception’ and 67% received a rating of 5. For image quality, 7% received a rating of 3, 14% received a rating of 4 and 79 percent received ratings of 5 on a five-point rating scale. Image quality was deemed to be the least successful of the three technical quality indicators in (Sicotte et al., 2003) , with 63 percent of the ratings falling in the middle of the scale. The SLPs gave a more favorable assessment of clinical quality, with 81 percent of their assessments falling on the positive side of the scale, meaning that they were satisfied 53 percent of the time and highly satisfied 28 percent of the time. In the current study, the clinical quality was rated as good by the SLPs. Overall, 100 percent of the clinical parameters received evaluations of 5.

The technical and clinical quality was as well evaluated by the 15 stuttering participants. The same three criteria-sound quality, reception delay, and image

quality-that the SLPs used to assess technical quality were employed by the participants as well. A three-point rating scale, "3" was deemed to be highly satisfied, "2" to be somewhat satisfied, and "1" to be completely unsatisfied, was used for rating for all three aspects, every participant gave a rating of "2" or above. This demonstrates that the participants' satisfaction with the sound and image quality was either "somewhat satisfied" or "very satisfied." Not even one participant gave the reception delay a "not satisfied" rating.

The participants utilized the same three criteria (degree of control, goal attainment, and compliance) as the SLPs used when evaluating clinical quality. A three-point rating scale was employed rather than a five-point system. For each of the three components, every participant gave a score of "2" or above. This shows that the participants had a positive experience with the researcher's online session. Sicotte et al. (2003) also achieved comparable outcomes, where they utilized a rating scale comparable to the one used in the current study and the same patient satisfaction measures to evaluate the technical and clinical quality of the intervention. Technical quality was rated at the highest level by all participants with the exception of one, according to Sicotte et al. (2003).

The participant's satisfaction with the technical quality was evaluated in the present study using a rating scale that was comparable to the one used by Sicotte et al. (2003). All fifteen individuals evaluated both on the technical and clinical quality aspects. Nine participants out of 15 assessed the technical quality of the image quality as highly satisfied, 5 participants rated as slightly satisfied. Reception delay as rated as "very satisfied," by all the 15 participants. Thirteen participants rated the sound quality as "very satisfied," while the remaining two participants gave it a "slightly satisfied" rating. In terms of clinical quality, for target attainment and compliance, all

the fifteen participants scored "very satisfied". One participant gave a rating of "somewhat satisfied," while fourteen individuals gave a rating of "very satisfied" for the level of control.

5.6 Feasibility to use automatic speech processing based assessment as an online technique for assessment of stuttering

When the findings of the perceptual evaluation and the results of the automatic speech processing-based assessment were compared, significant difference was not observed in the duration and frequency of part-word repetition, filled pauses, prolongation, word repetition and blocks. These findings demonstrate that, with improvements in the recognition accuracy, automatic speech processing can be used for online assessment of stuttering. The noise in the online recorded samples was one of the shortcomings of the present study. Also, automatic speech recognition systems require more number of samples to be trained for recognition of dysfluency parameters. In the current study, the small sample size contributed to the low accuracy. The accuracy for detection of prolongation was found to be 44%. The number of prolongation events presented to the ASR was only 40. Accuracy for word repetition was 43% as the number of word repetition events presented to the ASR were only 47. Filled pauses and blocks are detected with an accuracy rate of 41% each, where 20 filled pauses and 178 block were present in the samples presented to ASR. The part word repetitions were having the least accuracy rate (36%) as the ASR could get a sample size of only 65. Because of the small sample size of the group II, ASR couldn’t be trained to detect the dysfluencies accurately. Further improvements in accuracy can be obtained by using the same experimental setup with more extensive datasets so that the variance in samples of each disfluency can be captured effectively. The SLPs and participants' evaluations of the online sessions' technical

and clinical quality offer positive feedback. These information lead to the conclusion that automatic speech processing-based assessment can be used for online stuttering assessment.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Several researchers have tried to use automatic speech processing to evaluate fluency disorders. However, no studies have been reported yet to derive fluency parameters by automatic speech processing, for stuttering assessment in individuals who speak Malayalam. The traditional face-to-face evaluation is challenging in the pandemic circumstances such as COVID-19. In the current study, the feasibility for conducting an online assessment of fluency metrics in individuals who speak Malayalam, using automatic speech recognition was investigated.

The study included 45 literate, Malayalam-speaking participants ranging in age from 18 to 35 years. They were categorized into two groups: Group I, which included 30 persons who were normal, and Group II, which included 15 persons with stuttering. The widely used video conferencing software, Zoom, was employed by the researcher to conduct the evaluation in virtual mode. The reading material for the automatic speech processing test was the standardized passage in Malayalam (Annexure-A). Using SSI-4, the researcher evaluated the severity of stuttering in fifteen adults with stuttering, and found that two participants had very mild stuttering, four had mild stuttering, four had moderate stuttering, and five had severe stuttering. Through perceptual evaluation, the researcher also assessed the frequency and duration of prolongations, blocks, filled pauses, and repetitions from the recorded passage. For all fifteen participants who stutter, the fluency parameters were also evaluated using automatic speech recognition from the recorded standard passage. Each participant in Group II was asked to rate their satisfaction with the technical and clinical quality of the online session at the end of each session. Three speech-language pathologists evaluated each recorded session for its technical and clinical quality.

The results showed no significant difference between automatic speech recognition and perceptual evaluation for filled pauses, prolongations, blocks and word repetition. According to the evaluation of three SLPs and the subjects, the overall technical quality and clinical quality were satisfactory.

6.1 Important results of the study

The major findings of the study are:

- In terms of the frequency of filled pauses, prolongation, blocks and word repetitions, the present study demonstrated reasonably good agreement between perceptual evaluation and automatic speech recognition.
- The duration of the fluency parameters had complete agreement between perceptual assessment and automatic speech recognition.
- The severity of stuttering did not influence the accuracy of detection using automatic speech processing.
- Online assessment of stuttering can be conducted with acceptable quality.

6.2 Implications of the study

According to the study, automatic speech processing can be utilized to efficiently assess fluency issues online.

The study has also shown that the clinicians and persons with stuttering are satisfied with the technological and clinical quality of online evaluation. Thus, the study has demonstrated the viability of using telepractice for the evaluation of fluency disorders.

6.3 Limitations of the present study

- Only the reading task was used to assess fluency parameters using automatic speech processing.

- There was no gender balance among the participants, and the study only focused on adults.
- The sample size taken into consideration for the study is quite small. Accuracy could be improved if the ASR is trained with more number of samples.
- For the purpose of assessing stuttering, ASR does not take participants' physical concomitants into consideration.

6.4 Future recommendations

- A larger sample size that represents each severity category may be used in the investigation.
- The spontaneous speech also may be considered for automatic assessment of stuttering.
- The study can be replicated in children with stuttering.
- The study may be repeated with a group of participants who are equally divided across the genders.
- The study may be repeated in other Indian languages.

6.5 Significance of the results of the study

On tele-assessment of fluency disorders, only a few studies have been conducted. Furthermore, no studies have been done utilizing automatic speech recognition techniques to assess stuttering in adult Malayalam speakers through online. The results of the study will encourage tele-practice for stuttering assessment, especially in the epidemic circumstances like COVID- 19, where it is challenging to do a traditional face-to-face assessment. The study has proven the feasibility of doing online assessments of fluency disorders. The current study also demonstrated that the speech language pathologists as well as the participants are

satisfied with the technical and clinical quality of the online mode utilized in the study.

REFERENCES

Alharbi, F., & Farrahi, K. (2018). A Convolutional Neural Network for Smoking Activity Recognition. *2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom)*, 1–6. https://doi.org/10.1109/HealthCom.2018.8531148

Audhkhasi, K., Kandhway, K., Deshmukh, O. D., & Verma, A. (2009). Formant-based technique for automatic filled-pause detection in spontaneous spoken english. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4857–4860. https://doi.org/10.1109/ICASSP.2009.4960719

Bakker, K., Brutten, G. J., & McQuain, J. (1995). A preliminary assessment of the validity of three instrument-based measures for speech rate determination. *Journal of Fluency Disorders*, *20*(1), 63–75. https://doi.org/10.1016/0094-730X(94)00009-I

Bakker, K., & Riley, G. D. (2009). Computerized scoring of stuttering severity, v 2.0. *Austin, TX: Pro-Ed*.

Bayerl, S. P., Hönig, F., Reister, J., & Riedhammer, K. (2020). *Towards Automated Assessment of Stuttering and Stuttering Therapy* (pp. 386–396). https://doi.org/10.1007/978-3-030-58323-1_42

Büchel, C., & Sommer, M. (2004). What Causes Stuttering? *PLoS Biology*, *2*(2), e46. https://doi.org/10.1371/journal.pbio.0020046

Chee, L. S., Ai, O. C., Hariharan, M., & Yaacob, S. (2009). Automatic detection of prolongations and repetitions using LPCC. *2009 International Conference for Technical Postgraduates (TECHPOS)*, 1–4. https://doi.org/10.1109/TECHPOS.2009.5412080

Ciabarra, A. M. (2000). Subcortical infarction resulting in acquired stuttering. *Journal of Neurology, Neurosurgery & Psychiatry*, *69*(4), 546–549. https://doi.org/10.1136/jnnp.69.4.546

Cook, S., Donlan, C., & Howell, P. (2013). Stuttering severity, psychosocial impact and lexical diversity as predictors of outcome for treatment of stuttering. *Journal of Fluency Disorders*, *38*(2), 124–133. https://doi.org/10.1016/j.jfludis.2012.08.001

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002). Epidemiology of Stuttering in the Community Across the Entire Life Span. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1097–1105. https://doi.org/10.1044/1092-4388(2002/088)

Geetha, Y. V., Pratibha, K., Ashok, R., & Ravindra, S. K. (2000). Classification of childhood disfluencies using neural networks. *Journal of Fluency Disorders*, *25*(2), 99–117. https://doi.org/10.1016/S0094-730X(99)00029-7

Grant, A. C., Biousse, V., Cook, A. A., & Newman, N. J. (1999). Stroke-Associated Stuttering. *Archives of Neurology*, *56*(5), 624. https://doi.org/10.1001/archneur.56.5.624

Guitar, B. (2013). *Stuttering: An integrated approach to its nature and treatment*. Lippincott Williams & Wilkins.

Gupta, S., Shukla, R. S., & Shukla, R. K. (2019). Literature survey and review of techniques used for automatic assessment of Stuttered Speech. *Int. J. Manag. Technol. Eng*, *9*, 229–240.

Hamzah, R., & Jamil, N. (2019). Investigation of speech disfluencies classification on different threshold selection techniques using energy feature extraction. *Malaysian Journal of Computing (MJoC)*, *4*(1), 178–192.

Hariharan, M., Fook, C. Y., Sindhu, R., Adom, A. H., & Yaacob, S. (2013). Objective evaluation of speech dysfluencies using wavelet packet transform with sample entropy. *Digital Signal Processing*, *23*(3), 952–959. https://doi.org/10.1016/j.dsp.2012.12.003

Howell, P., Davis, S., & Williams, R. (2008). Late Childhood Stuttering. *Journal of Speech, Language, and Hearing Research*, *51*(3), 669–687. https://doi.org/10.1044/1092-4388(2008/048)

Howell, P., Davis, S., & Williams, R. (2009). The effects of bilingualism on stuttering during late childhood. *Archives of Disease in Childhood*, *94*(1), 42–46. https://doi.org/10.1136/adc.2007.134114

Johnson W (1955) A study on the onset and development of stuttering. In: Johnson W, Leutenegger RR, editors. Stuttering in children and adults: Thirty years of research at the University of Iowa. Minneapolis: University of Minnesota Press. pp. 37–73.

Kaushik, M., Trinkle, M., & Hashemi-Sakhtsari, A. (2010). Automatic detection and removal of disfluencies from spontaneous speech. *Proceedings of the Australasian International Conference on Speech Science and Technology (SST), 70*.

Kourkounakis, T., Hajavi, A., & Etemad, A. (2020). Detecting Multiple Speech Disfluencies Using a Deep Residual Network with Bidirectional Long Short-Term Memory. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6089–6093. https://doi.org/10.1109/ICASSP40776.2020.9053893

Lickley, R. J. (1994). *Detecting disfluency in spontaneous speech* (Doctoral dissertation, University of Edinburgh).

Lincoln, M., & Harrison, E. (1999). The lidcombe program. In *The handbook of early stuttering intervention: Early stuttering intervention* (pp. 103–116). Singular Publishing.

Liu, Y., Shriberg, E., Stolcke, A., Hillard, D., Ostendorf, M., & Harper, M. (2006). Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(5), 1526–1540.

Lu, Y., Gales, M. J. F., Knill, K., Manakul, P., & Wang, Y. (2019). Disfluency Detection for Spoken Learner English. *SLaTE*, 74–78.

Manning, W. H. (2009). Clinical decision making in fluency disorders (HS Sim, MJ Shin, & EJ Lee, Trans.). *Seoul: Cengage Learning Korea Limited*.

Mehrotra, U., Garg, S., Krishna, G., & Vuppala, A. K. (2021). Detecting Multiple Disfluencies from Speech using Pre-linguistic Automatic Syllabification with Acoustic and Prosody Features. *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 761–768.

Mendhakar, A. M., & Mahesh, S. (2018). Automatic Annotation of Reading using Speech Recognition: A Pilot study, *Research and Reviews: A Journal of BioInformatics, 5*(2), 25-29.

Miller, B., & Guitar, B. (2009). Long-Term Outcome of the Lidcombe Program for Early Stuttering Intervention. *American Journal of Speech-Language Pathology*, *18*(1), 42–49. https://doi.org/10.1044/1058-0360(2008/06-0069)

Naylor, R. v. (1953). A comparative study of methods of estimating the severity of stuttering. *Journal of Speech and Hearing Disorders*, *18*(1), 30–37.

Plänkers, T. (1999). Speaking in the claustrum: The psychodynamics of stuttering. *International Journal of Psycho-Analysis*, *80*(2), 239–256.

Ravikumar, K. M., Reddy, B., Rajagopal, R., & Nagaraj, H. (2008). Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies. *Proceedings of World Academy Science, Engineering and Technology*, *36*, 270–273.

Ravikumar, K. M., Rajagopal, R., & Nagaraj, H. C. (2009). An approach for objective assessment of stuttered speech using MFCC. In *The international congress for global science and technology* (p. 19).

Riad, R., Bachoud-Lévi, A.-C., Rudzicz, F., & Dupoux, E. (2020). Identification of primary and collateral tracks in stuttered speech. *ArXiv Preprint ArXiv:2003.01018*.

Riley, G. D. (1972). A Stuttering Severity Instrument for Children and Adults. *Journal of Speech and Hearing Disorders*, *37*(3), 314–322. https://doi.org/10.1044/jshd.3703.314.

Riley, G. D. (1994). *Stuttering severity instrument for children and adults*. Pro-ed.

Riley, G. D. (2009). *Stuttering Severity Instrument-Fourth Edition*. Austin, TX: Pro-Ed

Salesky, E., Sperber, M., & Waibel, A. (2019). Fluent translations from disfluent speech in end-to-end speech translation. *ArXiv Preprint ArXiv:1906.00556*.

Savithri, S. R., & Jayaram, M. (2004). Rate of speech/ reading in Dravidian Languages, [ARF Project]. AIISH, Mysore.

Sicotte, C., Lehoux, P., Fortier-Blanc, J., & Leblanc, Y. (2003). Feasibility and outcome evaluation of a telemedicine application in speech–language

pathology. *Journal of Telemedicine and Telecare*, *9*(5), 253–258. https://doi.org/10.1258/135763303769211256

Singhi, P., Kumar, M., Malhi, P., & Kumar, R. (2007). Utility of the WHO Ten Questions Screen for Disability Detection in a Rural Community the North Indian Experience. *Journal of Tropical Pediatrics*, *53*(6), 383–387. https://doi.org/10.1093/tropej/fmm047

Surya, A. A., & Varghese, S. M. (2016). Automatic speech recognition system for stuttering disabled persons. *International Journal of Control Theory and Applications*, *9*(43), 16–20.

Synergistic Electronics (TrueTalk Speech Fluency Rater). (n.d.). Retrieved August 5, 2022, from http://www.synelec.com.au/synergy/

Van Riper, C. (1982). *The nature of stuttering*. Prentice Hall.

Veda, P. (2021). Development of an Automatic Speech Processing Based Procedure for Assessment of Stuttering in Kannada Speaking Adults through Virtual Mode [Unpublished master's dissertation]. University of Mysore.

Wingate, M. E. (1964). A Standard Definition of Stuttering. *Journal of Speech and Hearing Disorders*, *29*(4), 484–489. https://doi.org/10.1044/jshd.2904.484

Wu, C.-H., & Yan, G.-L. (2004). Acoustic feature analysis and discriminative modeling of filled pauses for spontaneous speech recognition. In *Real World Speech Processing* (pp. 17–30). Springer.

Yairi, E., & Ambrose, N. G. (1999). Early Childhood Stuttering I. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1097–1112. https://doi.org/10.1044/jslhr.4205.1097

Yaruss, J. S. (1999). Disfluency frequency counter (computer software). *Pittsburgh: Author*.

Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., Danielson, M. L., Bitsko, R. H., Blumberg, S. J., Kogan, M. D., & Boyle, C. A. (2019). Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009–2017. *Pediatrics*, *144*(4). https://doi.org/10.1542/peds.2019-0811

Zayats, V., Ostendorf, M., & Hajishirzi, H. (2016). Disfluency detection using a bidirectional LSTM. *ArXiv Preprint ArXiv:1604.03209*.

Annexure-A

MALAYALAM PASSAGE (Savithri & Jayaram, 2004)

ഒരിടത്ത് ഒരു ബ്രാഹ്മണൻ ഉണ്ടായിരുന്നു. അന്ധവിശ്വാസങ്ങൾക്കു അടിമയായ അയാൾക്ക് ജീവിതത്തിൽ ഒരുപാട് പ്രശ്നങ്ങൾ നേരിടേണ്ടി വന്നു. തടിയനായിരുന്നതിനാൽ ജനങ്ങൾ അവനെ പൊണ്ണത്തടിയാ എന്ന് വിളിച്ചു.

ആര് എന്ത് വിശേഷാവസരത്തിന് ക്ഷണിച്ചാലും മുൻപേ ഇവൻ ഹാജാരാകുമായിരുന്നു.
ഒരു ദിവസം ധനപതി എന്ന ബാല്യകാല സുഹൃത്ത് ബ്രാഹ്മണനെ തൻ്റെ മകളുടെ ജന്മദിനാഘോഷത്തിനായി ക്ഷണിച്ചു.

എന്നാൽ സ്നേഹിതൻ്റെ വീട് ബ്രാഹ്മണൻ്റെ വീട്ടിൽ നിന്നും ആറു കിലോ മീറ്റർ അകലെ ആയിരുന്നു. നടന്നു പോയാൽ ആരോഗ്യത്തിന് നല്ലത്, കൂടാതെ കൂടുതൽ വിശന്നാൽ ഭക്ഷണം കൂടുതലും കഴിക്കാം.

ബ്രാഹ്മണൻ തീരുമാനിച്ചു.
ജന്മദിനാഘോഷത്തിൻ്റെ ദിവസം എത്തി. ഒരുക്കങ്ങളെല്ലാം വേഗത്തിൽ നടത്തി വീടിനു പുറത്തേക്കിറങ്ങിയപ്പോൾ ഒരു കുഷ്ഠരോഗി മുൻപിൽ പ്രത്യക്ഷപ്പെട്ടു.

ശകുനം ശരിയല്ലെന്ന് പറഞ്ഞ് ബ്രാഹ്മണൻ വീടിനുള്ളിലേക്ക് തന്നെ കയറിപോയി. ഇതുപോലെ മൂന്നു പ്രാവശ്യം കൂടി സംഭവിച്ചു. പിന്നീട് അയാൾ വേഗത്തിൽ നടന്ന് സ്നേഹിതൻ്റെ വീട്ടിലെത്തി.

അപ്പോൾ അവിടെ എല്ലാവരും ഭക്ഷണം കഴിച്ച് താംബൂലം ചവച്ച് വിശ്രമിക്കുകയായിരുന്നു.
വൈകി വന്ന ബ്രാഹ്മണനെ കണ്ട്, ധനപതി "എന്തേ ഇത്ര വൈകിയത്? ഭക്ഷണത്തിനുള്ള സമയം കഴിഞ്ഞു പോയല്ലോ"എന്ന് പറഞ്ഞുകൊണ്ട് രണ്ടു വാഴപ്പഴവും പാലും വരുത്തിക്കൊടുത്തു.

MALAYALAM PASSAGE - IPA (Savithri & Jayaram, 2004)

/oɾiʈat̪t̪ə oɾu braːɦmaɳan uɳʈaːjiɾun̪n̪u/ /an̪d̪ɦaʋiʃʋaːsaŋŋaɭkku aʈimajaːja ajaːɭkkə ɟiːʋit̪at̪t̪il oɾupaːʈə praʃnaŋŋaɭ n̪eːɾiʈeːɳʈi ʋan̪n̪u/ /t̪aʈijanaːjiɾun̪n̪at̪inaːl ɟanaŋŋaɭ aʋane poɳɳat̪t̪aʈijaː en̪n̪ə ʋiɭit͡ʃt͡ʃu/

/aːɾə en̪t̪ə ʋiʃeːʂaːʋasaɾat̪t̪inə kʂaɳit͡ʃt͡ʃaːlum munpeː iʋan ɦaːɟaːɾaːkumaːjiɾun̪n̪u/ /oɾu d̪iʋasam d̪ɦanapat̪I en̪n̪a baːljakaːlasuɦɾit̪t̪ə braːɦmaɳane t̪ant̪e makaɭuʈe ɟanmad̪inaːgɦoːʂat̪t̪inaːji kʂaɳit͡ʃt͡ʃu/

/en̪n̪aːl sneːɦit̪ant̪e ʋiːʈə braːɦmaɳant̪e ʋiːʈʈiln̪in̪n̪um aːrukiloːmiːt̪t̪ar akaleaːjiɾun̪n̪u/ /n̪aʈan̪n̪u poːjaːl aːɾoːgjat̪t̪inə n̪allat̪ kuːʈaːt̪e kuːʈut̪al ʋiʃan̪n̪aːl bɦakʂaɳam kuːʈut̪alum kaɻikkaːm/

/braːɦmaɳan t̪iːɾumaːnit͡ʃt͡ʃu/ /ɟanmad̪inaːgɦoːʂat̪t̪int̪e d̪iʋasamet̪t̪I /oɾukkaŋŋaɭellaːm ʋeːgat̪t̪il n̪aʈat̪t̪I ʋiːʈinu purat̪t̪eːkkiraŋŋijappoːɭ oɾu kuʂʈʰaɾoːgi munpil prat̪jakʂappeʈʈu/

/ʃakunam ʃaɾijallen̪n̪ə paraɲɲə braːɦmaɳan ʋiːʈinuɭaɭileːkkə t̪an̪n̪e kajaripoːji/ /it̪upoːle muːn̪n̪u praːʋaʃjam kuːʈi sambɦaʋit͡ʃt͡ʃu //pin̪n̪iːʈə ajaːɭ ʋeːgat̪t̪il n̪aʈan̪n̪ə sneːɦit̪ant̪e ʋiːʈʈilet̪t̪i/

/appoːɭ aʋiʈe ellaːʋaɾum bɦakʂaɳam kaɻit͡ʃt͡ʃə t̪aːmbuːlam t͡ʃaʋat͡ʃt͡ʃə ʋiʃramikkukajaːjiɾun̪n̪u/ ʋai̯ki ʋan̪n̪a braːɦmaɳane kaɳʈə d̪ɦanapat̪I en̪t̪eː it̪ra ʋai̯kijat̪ə bɦakʂaɳat̪t̪inuɭɭa samajam kaɻiɲɲupoːjalloː en̪n̪ə paraɲɲukoɳʈə ɾaɳʈu ʋaːɻappaɻaʋum paːlum ʋaɾut̪t̪ikkoʈut̪t̪u/