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

Veda P.

Reg. No: 19SLP036

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

SEPTEMBER 2021

CERTIFICATE

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Kannada 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: **19SLP036**. 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

Director

All India Institute of Speech and Hearing

Manasagangothri, Mysuru - 570006

Mysuru

September 2021

CERTIFICATE

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Kannada 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 & Head,

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

September 2021

DECLARATION

This is to certify that the dissertation entitled "**Development of an automatic speech processing based procedure for assessment of stuttering in Kannada 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**: **19SLP036**

September, 2021

ACKNOWLEDGEMENT

My sincere regards and deepest gratitude to my guide Professor **Ajish. K. Abraham***, for constant support and guidance. Without your support this study would not be possible sir and to my co-guide* **Dr. Sangeetha Mahesh***, who did not lose faith in me and gave me that extra push I needed. They are true definition of "Guru Devo Bhava".*

My parent's and my sister's blessings, support and prayers worked magic and empowered me to reach this stage. They were the two pillars who always stood by me.

Very special thanks to Sumathi, Niranjana,Tanuja, Rini, Shishira, Kriba and Viji who took time out of their busy schedule to help me in the process.

Last but not the least; I would like to extend my gratitude to all the participants for being very co-operative, patient and enthusiastic during the study.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO.
	List of Tables	ii
	List of Figures	iii
I	Introduction	1-6
II	Review of Literature	7-14
III	Method	15-21
IV	Results	22-35
V	Discussion	36-44
VI	Summary & Conclusion	45-48
	References	49-56
	Annexure- A	57

LIST OF TABLES

Table No.	Title of Table	Page No.
4.1	Characteristics of normal participants with respect to Mean, SD	23
4.2	Characteristics of participant with stuttering with respect to Mean, SD	23
4.3	Frequency and duration of filled pauses, repetitions and prolongations of participants with stuttering assessed through online perceptual evaluation	25
4.4	Frequency and duration of filled pauses, repetitions and prolongations of participants with stuttering assessed through automatic speech processing	27
4.5	Comparison between assessment of number of stuttering events through automatic speech processing and perceptual evaluation	28
4.6	Error in assessment of number of stuttering events through automatic speech processing across mild, moderate and severe stuttering	30
4.7	Overall error in assessment of number of stuttering events through automatic speech processing across different stuttering events	31

LIST OF FIGURES

Figure No	Title of Figure	Page No
3.1	Block schematic showing the process for automatic assessment of stuttering parameters	19
4.1	Technical quality based on evaluation by SLPs on a 5 point scale	32
4.2	Technical quality based on evaluation by participants on a 3 point scale	33
4.3	Clinical quality based on evaluation by SLPs on a 5 point scale	34
4.4	Clinical quality based on evaluation by participants on a 3 point scale	35

Chapter I

INTRODUCTION

Fluent speech makes communication more effective. "Dysfluency" is any break in fluent speech. "Stuttering" is the condition in which speech has more dysfluencies than is considered average. Van Riper (1982) defined stuttering as a “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”. Fluency disorder, generally referred to as “stuttering” is characterized by primary (core) and secondary behaviours. Primary behaviours include repetitions of sounds, syllables, or whole word and prolongations of single sound or blocks of airflow or voicing during speech. Abnormal breathing patterns, circumlocutions or word avoidances, and interjections are also observed.

Dysfluencies are seen in typically fluent speakers as well as persons with stuttering. The terms “stutter-like dysfluency” (SLD) and “other like dysfluency” (OLD) have been coined to distinguish between typical and atypical dysfluencies. Prolongations, syllable repetitions, part word repetitions and blocks are considered as SLDs, whereas phrase repetitions, revisions and interjections are considered as OLDs. SLDs are not generally observed in typically fluent speakers. OLDs are often observed in the speech of both people who stutter and typically fluent speakers.

The types of dysfluencies observed in persons with stuttering are:

- Part-word repetitions – "I **w-w-w-**want a drink."
- One-syllable word repetitions – "**Go-go-go** away."
- Prolongation of sounds – "**Sssssss**am is nice."
- Blocks or stops – "I want a c **(Struggle with pause)** ookie."

Yairi et al. (1996) observed that compared to females, stuttering is three times more common in males. Stuttering is found in approximately 2% of children aged 3–17 years (Zablotsky et al., 2019). 0.78% of adults in the age range of 21 to 50 years and 0.37% of those aged 51 and above are reported to have stuttering (Craig et al., 2002).

1.1 Conventional method for assessment of stuttering

Stuttering Severity Instrument - 4th editions (SSI-4) is a reliable and widely used tool for assessment of stuttering. SSI-4 assesses four areas of speech behavior: frequency, duration, physical concomitants and naturalness of the individual's speech (Riley, 2009). In SSI-4, the individual who is being assessed will be asked to read a structured passage in his native language.

From the spoken passage, the following are assessed through perceptual evaluation:-

- Syllable & word count of stuttering frequency.
- Frequency and duration of dysfluencies. (Typical dysfluencies:- hesitations, filler words, revisions, phrase & word repetitions, atypical dysfluencies:- blocks, prolongations, sound syllable and word repetitions)
- Disruption in forward flow
- Struggle behavior(s)
- Reduced or no eye contact
- Avoidance behaviors

1.2 Limitations of the conventional method

- It is difficult for the clinician to manually count the stuttered syllables when the person is reading the standard text as well as during spontaneous speech.
- Identifying the syllable boundary is difficult if the client has misarticulation.

- Only an experienced clinician can precisely identify the stuttered events.
- The factors that influence naturalness of speech are unknown.
- It takes into account overt features of expression only.
- There is no explanation for the covert features.
- Considerable amount of time is taken for conducting the conventional evaluation and hence, is taxing for the client.
- Inter-judge variability.

1.3 Online techniques for assessment of stuttering

Researchers have explored the use of telepractice across various communication disorders (Carey et al., 2012; Lewis et al., 2008; Irani & Gabel, 2011; O'Brian et al., 2008; Sicotte et al., 2003; Wilson et al,. 2004).

The merits and feasibility of telepractice have been explored by Blaisier et al., 2002. Barriers to telepractice such as connectivity issues, issues related to privacy of the client and professional licensing issues while providing the service to a client of another country, have also been reported in the literature (Cohn, 2012; Cohn et al., 2011; Denton, 2003).

With increased access to faster internet connection and availability of free of cost videoconferencing platforms, telepractice has become a viable option for many professionals. Telepractice links clinician to client/patient or clinician to clinician for assessment, intervention, and/or consultation (American Speech-Language-Hearing Association [ASHA], 2018).

1.4 Assessment of stuttering using automatic speech processing

Many researchers have attempted to assess fluency disorders by detecting the stuttering events using automatic speech processing. The approaches used by these researchers include:- Mel Frequency Cepstral Coefficient (MFCC) based feature extraction and identifying fluency disorders using Linear Discriminant Analysis (LDA) based classifier and k-nearest neighbors (k-NN) (Chee et al., 2009a); Linear Predication Cepstral Coefficients (LPCC) based feature extraction and identifying fluency disorders using LDA & k-NN (Chee et al., 2009b); LP-Hilbert Envelope Based MFCC features for the detection of prolongations, repetitions and interjections (Mahesha & Vinod, 2016).

In automatic speech processing, segmenting the speech signal followed by classification into fluent or dysfluent is the standard procedure with which stuttering events are detected. Two approaches are followed in this: 1. Segmenting audio files in windows of fixed length and then classifying those segments as one of the dysfluencies (Mendhakar & Mahesh, 2018) and 2. Segmenting audio files in some continuous frames and then finding similarity on those frames to predict the type of dysfluency. (Esmaili et al., 2016; Esmaili et al., 2017)

Non-linear features like wavelet packet transform with sample entropy (Hariharan et al., 2013) have also been used for detection of dysfluencies as voice production (in stuttering) is non-linear in nature due to the involuntary silent pauses, repetition and prolongation of words. Recently, deep learning architectures are also being used with both text and signal level features (Alharbi et al., 2018; Santoso et al., 2019; Kourkounakis et al., 2020). Kourkounakis et al. (2020) used spectrogram features to classify a 4-second stutter file into one of the 6 types of dysfluencies.

The above studies indicate that automatic speech processing can be effectively used for assessment of stuttering. However, very few studies have been done in Indian languages and no studies have been reported in Kannada. MFCCs shimmer and formant features have been used to detect repetition in Hindi stuttered speech in the study by Ramteke et al. (2016). Most of the studies cited above have dealt with detection of dysfluency in English and other European languages.

1.5 Need for the study

Review of the past studies indicates that telepractice is emerging as a feasible service delivery method, with many beneficiaries reporting satisfaction. These studies established the effectiveness of telepractice. However, very few studies have been reported on tele-assessment of fluency disorders. Moreover, no research is conducted on online assessment of stuttering in Kannada speaking adults using automatic speech recognition techniques. In the present pandemic situation, it is difficult to have the conventional face to face assessment. The possibility of conducting online assessment needs to be explored.

Stuttering severity assessment is usually done through perceptual evaluation by speech language pathologists (SLPs), which is time consuming and requires lot of effort from the SLP. While attempts towards automated assessment in English were made by several researchers, very few attempts have been made in Indian languages. There is a gap between the demand and availability of SLPs in the country. A fully automated system for assessment of stuttering will save the time for assessment, reduce the effort of the SLP and eliminate manual errors and biasing. Hence, the present study attempts to develop a technique for automatic assessment of fluency parameters in Kannada speaking adults through virtual mode.

1.6 Aim of the study

The aim of the present study was to develop an online technique using automatic speech processing for assessment of stuttering in Kannada speaking adults.

1.7 Objectives of the study

- To record through Zoom app, the standard Kannada passage spoken by 30 normal adults, which was used as normative speech for automatic speech processing.
- To determine online, the severity of stuttering for each of the 15 adults with stuttering using SSI-4 including the fluency parameters such as frequency and duration of filled pauses, repetitions and prolongations through perceptual evaluation.
- To assess through automatic speech processing, the frequency and duration of blocks, repetitions and prolongations, from the standard Kannada passage spoken by 15 adults with stuttering, recorded through Zoom app.
- To find the correlation between the values of fluency parameters derived through automatic speech processing and the values obtained through perceptual assessment.
- To assess the accuracy of fluency parameters derived through automatic speech processing across different levels of severity.
- To assess the technical and clinical quality of the online assessment based on the evaluation by the participant and three SLPs.

Chapter II

REVIEW OF LITERATURE

Assessment of stuttering is done mainly to:- a) diagnose the presence of stuttering and to identify the subgroup of stuttering, b) obtain base measurements, c) observe and record improvements, d) plan effective therapy, e) document the outcome of therapy, f) evaluate the therapy protocol and make modifications, g) provide outcome measures and h) identify the nature of stuttering (Hayhow, 1983).

2.1 Conventional methods for assessment of stuttering

Many protocols were practised to examine the characteristics of speech to discriminate stuttered speech from normal speech. One of such protocols was the Iowa scale for rating severity of stuttering (Naylor, 1953). A rating scale ranging from zero to seven was employed. Many of the items on this rating scale require judgement of the motivation of the person with stuttering for a given behaviour. Lewis and Sherman (1951) used a different approach that was the "Sherman- Lewis Scale" which used 9 tape-recorded samples of stuttering, ranked from mild to severe, as reference, to judge the severity of stuttering. Parameters such as words stuttered per minute and number of blocks per words were also utilized for severity measurements (Minifie & Cooker, 1964).

SSI-3: Stuttering Severity Instrument – Third Edition (Riley, 1972) provides a single numerical score which will provide severity evaluation and a benchmark for monitoring clinical change. It has 3 parameters such as: frequency (0-18), duration (0-7), and physical concomitants (0-20). Mirawdeli (2014) established the reliability of SSI-3 in screening children, as it successfully distinguishes fluent from dysfluent children. It was also observed that SSI-3 has high specificity and low sensitivity.

SSI-4 (Riley, 2009), is an improvised version of SSI-3. In this test, respondents are asked to describe their job (if employed), or school (if school going), and read a short passage (or describe pictures if they cannot read). The clinician records the speech and scores the respondent on stuttering frequency, stuttering duration, and physical concomitants across four categories. The scoring pattern used in SSI-4 is as follows:- Frequency is expressed in percentage of syllables stuttered and is converted to scale scores of 2–18. Duration is timed to the nearest tenth of a second and is also converted to scale scores of 2–18. The four types of physical concomitants (distracting sounds, facial grimaces, head movements and movements of the extremities) are scored and converted to scale scores of 0–20. A total score is obtained from the frequency and duration of stuttering dysfluencies, the score of physical concomitants, and the estimation of speech naturalness. Speech naturalness is rated from 1 to 9, where 1 indicates highly natural sounding speech and 9 indicates highly unnatural sounding speech (Riley, 1972). Physical concomitants associated with the blocks or with attempts to avoid blocking is scored on the following scale: ‘0’- none, ‘1’- not notice- able unless looking for it, ‘2’ - barely noticeable to casual observer, ‘3’- distracting, ‘4’- very distracting, ‘5’- severe and painful looking (Riley, 2009). This total score is ranked according to age-specific population norms (Pre-school age children, school age children and adults) and is used to assign a verbal descriptor of stuttering severity, ranging from very mild to very severe.

A very high inter-judge reliability along with poor absolute intra-judge reliability for the SSI-4 scores was reported by Tahmasebi (2018). Davidow (2017) calculated intra- and inter-judge reliability of SSI-4 across the sub scores (frequency, duration and physical concomitants) and found that the total score agreed with the

SSI-4 manual's reported reliability. The above discussed studies indicate that SSI-4 can be considered as a reliable test for perceptual assessment of stuttering.

2.2 Limitations of the conventional method

While assessing the stuttering severity using SSI-4, SLPs experience several limitations. Manning et al. (2009) highlighted some of these limitations. One of the challenges in assessing fluency is the high degree of inter-speaker variability. The level of fluency varies depending on the time and place. As a result, in any speaking circumstance, the amount and degree of difficulty cannot be predicted. Many features of fluency condition would go unreported in young speakers unless the assessment is undertaken in a range of speaking situations. For perceptual evaluation of frequency and duration of stuttering moments experienced clinicians are required. The perceptual evaluation is time consuming. Moreover, the inconsistency of stuttering necessitates continuous assessment that takes place over several assessment or treatment sessions. Limitations pertaining to persons with stuttering who undergo the conventional assessment using SSI-4 include their level of motivation, loss of control because of stuttering, etc.

2.3 Techniques for online assessment of stuttering

2.3.1. Telepractice

During the last 15 years, telepractice was effectively used for assessment and treatment of various communication disorders, such as neurogenic communication disorders (Armfield et al., 2012); voice disorders (Halpern, et al., 2012); speech-language disorders (Grogan-Johnson et al., 2009) and fluency disorders (Carey & Packman, 2012). Merits of telepractice mode and the barriers have also been discussed in the literature (Blaisier et al., 2013; Cohn et al., 2011).

2.3.2 Telepractice for Stuttering

2.3.2.1 Telepractice for treatment of stuttering

In a case study conducted by Kully (2000), adult with stuttering was selected to telepractice treatment, after 3-week intensive in-person treatment. The telepractice targeted practicing the fluency techniques and solving other real life problems. An overall satisfaction was reported by the client. In another study conducted by Irani and Gabel (2011), the participant (adult with stuttering) was first assigned an in-person therapy program, intensively for a 3-week period. Then, the participant was enrolled for treatment via telepractice, twice a week for 6 months followed by weekly therapy for an additional 6 months. A decline was observed in the percentage of syllables stuttered. Further, the participant reported a positive change in his attitudes and self-image. Thus, hybrid treatment programs can help in treatment of stuttering.

McGill et al. (2018) reviewed seven studies, one of them on an adult, undergoing telepractice for stuttering treatment where it was reported that the telepratice delivered almost similar levels of success as the traditional face-to-face therapy. Moderate to high satisfaction was rated by the clinicians who participated in these studies. Further research is required for validating, correcting the glitches and improving the features of the telepractice (Cullen & Webb, 2019). McGill et al. (2018) reviewed the use of telepractice for assessment and treatment of stuttering. The review concluded that live-stream video telepractice can be effectively used for treating stuttering.

Sicotte et al. (2003) reported that when the speech therapy was delivered through telepractice for persons with stuttering, the satisfaction of both patient and clinician were high. Participants opined that the intervention is effective. Tele assessment is more demanding for the clinician, particularly when it comes to dealing with young children, and for parents, who must take an active role during treatment.

2.3.2.2 Telepractice for assessment of stuttering

ASHA states that telepractice is a “telecommunications technology for delivery of professional services at a distance by linking clinician to client, or clinician to clinician for assessment, intervention, and/or consultation” (Gabel et al., 2013). The feasibility of utilizing telepractice to assess stuttering needs to be explored. No studies have been reported so far on tele-assessment of stuttering in India.

2.4 Studies on assessment of stuttering using automatic speech processing

Different techniques have been tried out to detect stuttering through processing of recorded speech samples (Arjun et al., 2020). Use of Artificial Neural Networks (ANN), Hidden Markov Models (HMM), and Support Vector Machine (SVM) have been used to identify stuttering. Mahesha et al. (2013) used MFCC, LPC and LPCC features and Multiclass SVM classifier to identify syllable repetition, word repetition and prolongation. An accuracy of 75%, 88% and 92% is obtained respectively when LPC, MFCC and LPCC features are considered.

Geetha et al. (2000) conducted a research on classification of childhood dysfluencies using Artificial Neural Networks (ANNs). They found that ANNs could distinguish normal non-fluency and stuttering with 92% accuracy. Ravikumar et al. (2008) developed an automatic detection method for syllable repetition in read speech, through segmentation, feature extraction, score matching and decision logic. 12 MFCC features were employed in a feature extraction algorithm and the recognition system was based on neural networks.

Świetlicka et al. (2009) suggested Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks to identify and distinguish fluent and non-fluent speech samples. The study showed a classification accuracy between 88.1% and 94.9% for all networks. The study also showed that ANNs can be used as a tool in speech analysis both of the fluent and non-fluent speaker.

Gupta et al. (2021) proposed a new technique for assessing stuttered speech, consisting of feature extraction using the Weighted Mel Frequency Cepstral Coefficient (WMFCC) and classification using a Bi-directional Long-Short Term Memory neural network. The results of the experiment revealed that WMFCC outperforms the other feature extraction methods, with an average recognition accuracy of 96.67%.

Ghonem et al. (2017) proposed a new system employing i-vector methodology for classifying speech samples as normal, repetition or prolongation, using MFCC feature extraction technique. The system showed an accuracy of 52.43%, 69.56%, 40%, and 50% respectively for normal, repetition, prolongation and rep-pro classes. Szczurowska et al. (2009) introduced a two stage Artificial Neural Network (ANN) based technique for classifying repetition and prolongation of syllables. This architecture achieved the average accuracy of 91%.

Studies reviewed above indicate that automatic speech processing can be effectively utilized for assessment of stuttering. However, lack of studies in Indian languages deprives the opportunity to use this technology. Hence, development of such systems in Indian languages is essential.

2.5 Limitations of the studies on automatic assessment of stuttering

John (2013) reported thatan effective ASR system poses a number of challenges. Benzeghiba (2007) suggested that while performing automatic speech recognition, speech variability is generally seen. Some of the factors affecting the speech realization: regional, sociolinguistic, or related to the environment or the speaker herself. Khara (2018) observed that MFCC give better accuracy than LPCC with noiseless speech data, whereas, LPCC gives better accuracy compared to LPC. Performance of RASTA (Relative Spectral Filtering) is good with noisy speech signals, but, shall be used with PLP (Perceptual Linear Prediction) for better accuracy. PLP works by converting frames of windowed speech into the frequency domain. It is found to be accurate than MFCC and LPCC, but it depends on spectral balance.

Kumar (2008) found the number of repetitions from the speech samples using MFCC feature extraction algorithm, Chee et al. (2009) opined that LPCC can be used for identification of repetitions and prolongations in stuttered speech with the average accuracy of 88.05%.

2.6 Need for automatic assessment

The above studies reveal that automatic speech processing systems with better accuracy and efficiency are required to find out the frequency and duration of stuttering events in Indian languages. Waghmare (2017) reported that speech recognition-based applications are becoming increasingly common, and they are now being used in a variety of settings. Until now, very little work has been done on stuttered speech recognition for Indian languages. Such systems for Indian languages are required, as they will lead to development of online assessment systems, which will address the shortage of manpower in India for assessment of stuttering. Moreover, such systems will be very helpful to provide tele-services during the present pandemic situation. Such systems will also reduce the time and effort of SLP in assessment. The present study is an attempt to develop an automatic system for finding out the frequency and duration of stuttering events through automatic speech processing.

Chapter III

METHOD

The aim of the study was to develop an automatic speech processing based technique for assessment of stuttering in Kannada speaking adults through virtual mode.

3.1 Participants

A total of 45 literate, Kannada speaking participants within the age range of 18-35 years were recruited in the study. They were divided into two groups, Group I included 30 normal participants, and Group II included 15 participants with stuttering.

Inclusion criteria for selection of normal participants (Group I)

- Participants having mobile phones within the price range of Rs. 10,000-20,000/-.
- Participants with the Zoom app installed in their mobile phones.
- Participants with normal speech and language, hearing, vision, communication, intact cognition and without any neurological, social, emotional, cognitive or psychiatric disturbances.

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

- Participants who have been diagnosed mild to severe stuttering by SLPs.
- Participants having mobile phones within the price range of Rs.10, 000-20,000/-.
- Participants with the Zoom app installed in their mobile phones.
- Participants with normal hearing, vision, communication, intact cognition and without any neurological, social, emotional, cognitive or psychiatric disturbances.

3.2 Material

The standardized passage (Annexure A) in Kannada (Bengaluru and Krishna nadi passage) was used as the reading material to assess stuttering through automatic speech processing and through perceptual evaluation. Spontaneous speech was also used in virtual mode assessment using SSI-4.

The text of the Bengaluru and Krishna nadi passage was prepared in Nudi 0.5e font in three font sizes (18, 20 and 22), in bold letters. Double space was used between words and double line spacing was used between lines. The text was arranged in two 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 20, in bold letters and double line spacing was chosen for the study.

3.3 Procedure

3.3.1 Recording of read passage through Zoom app

Before the recording, the researcher established video conference through Zoom app and built up rapport with the participant (both Group I and II), ensured that they are comfortable and gave clear instructions. Each participant was instructed to be seated comfortably in a noise-free room, keeping their mobile phone at a distance of 10 cm away from the mouth. They were requested to read the standardized passage which was displayed on the mobile phone. The researcher recorded the meeting through Zoom app. All the participants were asked to read the standardized Kannada passage at a comfortable loudness level. The recorded samples were stored in laptop or PC for further analysis. Recorded files were extracted from the laptop and audio samples were converted to .wav format using the online conversion of audio-video recordings. The recorded samples in the .wav format were then saved.

3.3.2 Online assessment through Zoom platform using SSI-4

Each participant from Group II was assessed online through Zoom platform using SSI-4. Initially, researcher built rapport with the participant and made them feel comfortable. The participant was instructed to read the Kannada standardized passage presented via Zoom using screen share option. Before collecting data, the researcher obtained feedback from the participant to ascertain the visual quality and readability of the passage.

Assessment protocol for SSI-4 involves 3 tasks: a) Job task- Conversation, spontaneous speech and narration, b) Reading task - The standardized Bengaluru and Krishna nadi passage in Kannada (Annexure-A) and c) Picture description task for non-readers. The researcher engaged the participant with stuttering in job task, reading task or picture description. The researcher marked fluent word as (.) and non-fluent word as (/). While the participant is reading or doing job task, the researcher observed secondary behaviors and scores were given to secondary behavior of physical concomitants (Distracting sounds, facial grimaces, head movements and movements of the extremities) according to the Manual of SSI-4 (Riley 2009). The researcher recorded the speech and assessed the score of the participant on stuttering frequency, stuttering duration, and physical concomitants across 4 categories.

3.3.3 Inter judge reliability

Assessment of fluency parameters was also done by two independent judges. These judges were SLPs who has experience of more than 5 years in assessment of stuttering. The purpose of the study was not revealed to the SLPs. Inter judge reliability was assessed by comparing the judgment and agreement between the researcher and the independent judges.

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

At the end of each session, satisfaction of the participant in Group II, with respect to the technical and clinical quality of the online session was assessed by the respective participant on a three-point scale: '3'- highly satisfied, '2'- somewhat satisfied, '1'- not at all satisfied. This assessment was based on the following aspects: 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

Each recorded session was rated by three SLPs for its quality on a six - item, five - point scale (from '1'- highly dissatisfied to '5'- highly satisfied). Quality rating is divided into two sections: technical and clinical quality. The technical quality was assessed in terms of three items: the quality of sound, delay in signal reception and image quality. Clinical quality was also assessed in terms of three items: degree of control over the participant during the session, attainment of clinical goals and degree of compliance of the participant with the instructions given by the researcher (Sicotte et al., 2003).

3.3.6 Assessment of stuttering through automatic speech processing

Figure 1 shows the process used for identifying the dysfluencies in the stuttered speech from the recorded speech using automatic speech processing. A trained Deep Neural Network (DNN) was used. This model detects whether a particular type of dysfluency is present in a 10ms frame of speech or not. Using this, the time stamps for dysfluencies can be found in an audio file. Mel-filter bank features were extracted in a 25 ms window with 10 ms shift. The Mel-filter bank features was then normalized (mean variance) using the Voice Activity Detection (VAD) information. For intonation, fundamental frequency was extracted in a window of 50 ms and sampled every 1.8 ms.

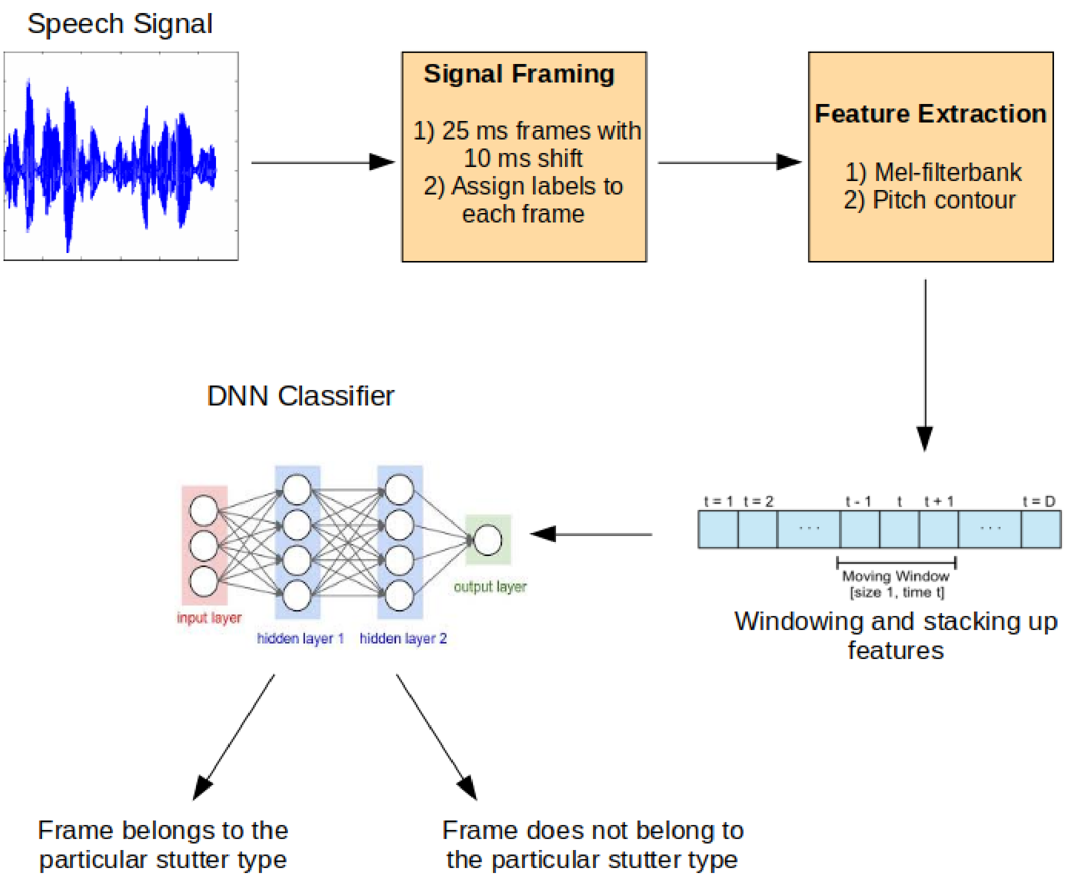


Figure 3.1

Block schematic showing the process for automatic assessment of stuttering parameters

The final feature vector for every frame was then obtained by stacking up features from 3 frames before the central frame and 3 frames after the central frame. A binary classifier was then trained to detect whether a particular type of dysfluency is present in the 10 ms audio frame or not. The DNN had 2 hidden layers with 100 and 50 units each. The binary cross entropy loss function was used. Since, the number of frames having normal speech was of much higher number as compared to the frames having dysfluency; the random under-sampling approach was used to do balanced training. An 80:20 train-test ratio was used. Once the type of dysfluency and the dysfluent word is identified, the algorithm estimates the frequency and duration of the stuttered moments.

3.3.7 Parameters assessed

a. Frequency: Frequency of stuttering is assessed as the percentage of words stuttered which is equal to number of stuttered moments per 100 words.

b. Duration: Duration is assessed by measuring the time period of three longest stuttered moments and by calculating their mean duration.

3.4 Analysis

- Recordings of speech from participants in Group II were analyzed through a matlab based code (developed by a person who is experienced in matlab) to automatically identify the frequency and duration of prolongations, filled pause and repetitions. Recordings of speech from participants in Group I served as the normative data for this analysis.
- The recorded speech of participants in Group II was analyzed by the researcher to find out the frequency and duration of prolongations, filled pauses and repetitions. The researcher's findings were validated by three experienced SLPs.
- The audio data record was taken from Zoom; it's converted into .wav file. Then wave file was extracted using PRAAT software. Starting point and ending point of each word was noted. To find out whether the word is fluent or dysfluent, researcher used following codes for perceptual evaluation. '0' - Clean speech,

‘1’- Interjection, ‘2’- Sound repetition, ‘3’ - Part word repetition, ‘4’ - Whole word repetition, ‘5’ - Phrase repetition, ‘6’ - Revision &‘7’ - Prolongation.

- The frequency and duration of prolongations, filled pauses and repetitions obtained through Matlab based code were compared with the results of perceptual evaluation of the researcher and the difference was analyzed.
- The obtained results of perceptual assessment and automatic speech recognition were subjected to statistical analysis in Statistical Package for the Social Sciences (SPSS) software (Version 21.0.). Descriptive statistics was carried out to calculate the mean, median, and standard deviation for both the groups. Shapiro Wilk test and Kolmogorov-Smirnov test was done to test the normality. As the obtained data was non-normalized, a non-parametric analysis was performed. That is, the Wilcoxon Signed Ranks Test was done to compare between perceptual assessment and automatic speech recognition process.
- The accuracy of fluency parameters derived through automatic speech processing across different levels of severity was analyzed.

3.5 Ethical consideration

The research’s aim and objectives were informed to all the participants along with the study's procedures and purpose. Their safety and confidentiality were assured, and an oral consent was taken before the study.

Chapter IV

RESULTS

Aim of the present study was to develop an online technique for assessment of stuttering using automatic speech processing in Kannada speaking adults. Frequency and duration of stuttering events and the severity of stuttering for each of the 15 adults with stuttering were evaluated online through Zoom video conferencing application, using SSI – 4. The frequency and duration of filled pauses, repetitions, and prolongations were also evaluated using automatic speech processing of the recorded passage. The standard Kannada passage spoken by 30 normal participants recorded using the Zoom application was used as the normative speech for automatic speech processing. The goal was to see the similarity between the values of fluency parameters acquired from automatic speech processing and the values obtained through online perceptual assessment.

4.1 Characteristics of participants

4.1.1 Characteristics of normal participants

Thirty adult, literate, native Kannada speakers with normal speech and language skills, as well as normal hearing, vision, communication and cognition participated in the study. The characteristics of these participants are shown in Table 4.1.

Table 4.1

Characteristics of normal participants

Sex	No. of participants	Mean age	Standard Deviation
Male	15	25.87	4.03
Female	15	25.73	2.37
Over all	30	25.80	3.25

4.1.2 Characteristics of participants with stuttering

Fifteen literate, native Kannada speaking adults, who were diagnosed to have mild to severe stuttering participated in the present study, they had normal language skills and had no history of hearing, vision and any other communication problems. Also, all the participants knew how to use the Zoom for video call. Participants with symptoms of neurological, social, emotional, cognitive, or psychiatric disorders were excluded from the study. The characteristics of these participants are shown in Table 4.2.

Table 4.2

Characteristics of participants with stuttering

Severity	No. of participants		Mean age	Standard Deviation
	Male	Female		
Mild	2	1	28.33	3.512
Moderate	6	3	24.89	4.485
Severe	3	0	24.00	5.292
Over all	11	4	25.40	4.437

4.2 Online assessment of the severity of stuttering for each of the 15 adults with stuttering using SSI – 4

4.2.1 Severity assessment

The researcher assessed the severity of stuttering for fifteen adults with stuttering through virtual mode with Zoom platform using SSI-4. Nine participants were diagnosed to have moderate stuttering; three each were diagnosed to have mild and severe stuttering respectively. The diagnosis of the researcher was validated by three SLPs and the validation results were in 100% agreement with the researcher's diagnosis.

4.2.2 Assessment of frequency and duration of filled pauses, repetitions and prolongations

Through online interaction, the researcher determined the severity of stuttering using SSI-4 for each participant with spontaneous speech and passage. The researcher also assessed the frequency and duration of prolongations, filled pauses and repetitions from the recorded passage through perceptual evaluation.

The following code was used to indicate the stuttering events of each participant:

0 - Clean Speech

1 – Filled Pause

2 - Sound Repetition

3 - Part- Word Repetition

4 - Word Repetition

5 - Phrase Repetition

6 - Revision

7 – Prolongation

Table 4.3

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

Partici-pants	Stuttering Severity	Filled pauses		Prolongations		Part word repetitions		Word repetitions		Duration in seconds
		No.	Freq (%)	No.	Freq (%)	No.	Freq (%)	No.	Freq (%)	
1	Moderate	0	0	8	11	13	17	0	0	8
2	Moderate	0	0	1	1	14	19	1	1	6
3	Moderate	0	0	1	1	2	3	0	0	4
4	Moderate	0	0	6	8	12	16	0	0	4
5	Severe	0	0	14	19	9	12	3	4	8
6	Moderate	0	0	1	1	5	7	0	0	2
7	Mild	4	5	1	1	4	5	7	9	2
8	Severe	0	0	5	7	20	27	0	0	4
9	Moderate	0	0	4	5	11	15	3	4	6
10	Moderate	0	0	1	1	11	15	1	1	6
11	Mild	0	0	0	0	0	0	1	1	4
12	Mild	0	0	1	1	1	1	1	1	6
13	Moderate	17	23	0	0	2	3	1	1	4
14	Moderate	0	0	10	13	3	4	1	1	2
15	Severe	0	0	2	3	1	1	0	0	6

Table 4.3 shows the results of assessment of frequency and duration of stuttering events through perceptual evaluation using SSI-4 for all the 15 participants. The number and frequency of the following stuttering parameters were assessed: - prolongations, word repetitions, part-word repetitions and filled pauses. Frequency of stuttering is assessed as the percentage of words stuttered which is equal to number of stuttered moments per 100 words. Highest frequency (9) of word repetitions was

observed in participant no. 7 with mild stuttering. Highest frequency (27) of part-word repetitions was observed in participant no. 8 with severe stuttering. Highest frequency (19) of prolongations was observed in participant no. 5 with severe stuttering. Highest frequency (23) of filled pauses was observed in participant no. 13 with moderate stuttering. Frequency was calculated using the method mentioned in Section 3.3.6. Durations were measured in seconds. Average length of the 3 longest stuttering events was taken for the duration measurement. Duration scale defined in SSI-4 was used.

The stuttering events identified by the researcher were compared with the assessment of two independent judges who was not aware of the purpose of the experiment. There was 90 percent correlation between the two sets.

4.3 Assessment through automatic speech processing

Table 4.4 represents the fluency parameters assessed through automatic speech recognition from the recorded standard passage for all the fifteen participants with stuttering. Similar to perceptual evaluation, highest frequency (5) of word repetitions was observed in participant no. 7 with mild stuttering. Highest frequency (16) of part-word repetitions was observed in participant no. 8 with severe stuttering, which is also similar to the results of perceptual evaluation.

Differing from perceptual evaluation, highest frequency (11) of prolongations was observed in participant no. 14 with moderate stuttering. Highest frequency (15) of filled pauses was observed in participant no. 13 with moderate stuttering, again similar to perceptual evaluation.

Table 4.4

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

Participants	Stuttering Severity	Filled pauses	Prolongations	Part word repetitions	Word repetitions	Duration in seconds
		Freq.	Freq.	Freq.	Freq.	
1	Moderate	0	8	11	0	8
2	Moderate	0	1	11	1	6
3	Moderate	0	1	3	0	4
4	Moderate	0	5	12	0	4
5	Severe	0	7	13	3	8
6	Moderate	0	1	5	0	2
7	Mild	4	1	3	5	2
8	Severe	0	4	16	0	4
9	Moderate	0	3	9	3	6
10	Moderate	0	1	9	1	6
11	Mild	0	0	0	1	4
12	Mild	0	1	1	1	6
13	Moderate	15	0	1	1	4
14	Moderate	0	11	3	1	2
15	Severe	0	3	1	0	6

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

Comparison between results of perceptual evaluation and automatic speech processing was done using Wilcoxon Signed Ranks Test. Results (Table 4.5) show a significant difference in prolongation (Z= 0.020) and Part-word repetition (Z= 0.003).

No significant differences were observed in filled pauses (Z = 0.180) and word repetition (Z = 0.102), (P > 0.05).

Table 4.5

Comparison between assessment of frequency of stuttering events through automatic speech processing and perceptual evaluation

	Median	Inter-quartile Range
Perceptual Evaluation		
Filled pauses	0.00	0
Prolongation	1.00	5
Part- word repetition	5.00	10
Word repetition	1.00	1
Automatic speech processing		
Filled pauses	0.00	0
Prolongation	1.00	3
Part- word repetition	4.00	7
Word repetition	1.00	1

4.5 Accuracy of fluency parameters derived through automatic speech processing across different levels of severity

Error in the number of the fluency parameters derived through automatic speech processing for each participant was calculated using the below formula:

Percentage error 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 followed for other fluency parameters also.

Table 4.6 shows the error in fluency parameters derived through automatic speech processing across different levels of severity such as mild, moderate and severe stuttering. Highest error (50%) in part-word repetitions was observed in participant 13 (moderate). Highest error (35%) in filled pauses was observed in participant 13 (moderate). Highest error (64%) in prolongations was observed in participant 5 (severe). Highest error (43%) in word repetitions was observed in participant 7 (mild).

Table 4.6

Error in assessment of number of stuttering events through automatic speech processing across mild, moderate and severe stuttering

Participants	**Stuttering Severity**	**No. of filled pauses**	**No. of prolongations**	**No. of Part word repetitions**	**No. of Word repetitions**
1	Moderate	0	25%	38%	0
2	Moderate	0	0	43%	0
3	Moderate	0	0	0	0
4	Moderate	0	33%	25%	0
5	Severe	0	64%	11%	33%
6	Moderate	0	0	20%	0
7	Mild	25%	0	50%	43%
8	Severe	0	40%	40%	0
9	Moderate	0	50%	36%	33%
10	Moderate	0	0	36%	0
11	Mild	0	0	0	0
12	Mild	0	0	0	0
13	Moderate	35%	0	50%	0
14	Moderate	0	20%	33%	0
15	Severe stuttering	0	0	0	0

Table 4.7

Overall error 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 error**
Filled pauses	21	14	33%
Prolongations	55	36	35%
Part-word repetitions	108	74	31%
Word repetitions	19	14	26%

Overall error in each of the fluency parameters derived through automatic speech process was calculated using the below formula.

Overall percentage error in part – word repetition = {(Total number of part – word repetition in perceptual assessment – Total number of part – word repetition in automatic speech processing) / Total number of part – word repetition in perceptual assessment} x 100. The same formula was followed for each fluency parameter.

Table 4.7 shows the overall errors in fluency parameters derived through automatic speech processing. Maximum error (35%) was seen in prolongations, whereas the minimum error (26%) was observed in word repetitions. Occurrences of filled pauses and word repetition were less compared to prolongation and part word repetition.

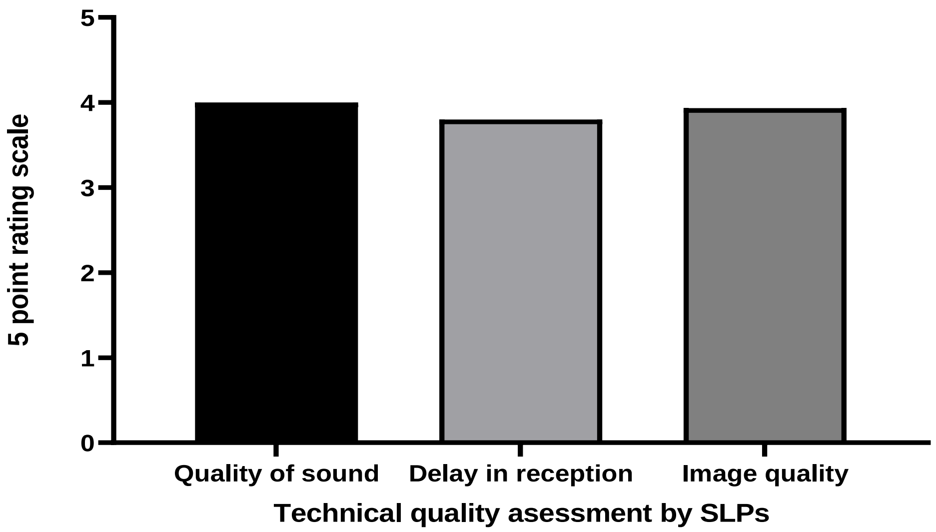
4.6 Technical and clinical quality of the online assessment of persons with stuttering

4.6.1 Technical quality based on assessment by the SLPs

Figure 4.1 represents the average scores of technical quality assessment which was done by 3 SLPs, by analyzing the recorded video of the sessions for each participant. The assessment was based on a five point rating scale, where '1' indicates highly dissatisfied and '5' indicates highly satisfied. Technical quality was evaluated on three aspects:-quality of sound, delay in reception and image quality. Quality of sound was found to be highly satisfied (5) for two participants, while the rating was satisfactory (3 and above) for all other participants. Delay in reception was reported to be highly satisfied (5) for two participants, while the rating was satisfactory (3 and above) for all other participants. Image quality was assessed to be highly satisfied (5) for the three participants, while the rating was satisfactory (3 and above) for all other participants.

Figure 4.1

Technical quality based on evaluation by SLPs on a 5 point scale (1 indicates highly dissatisfied and 5 indicate highly satisfied)

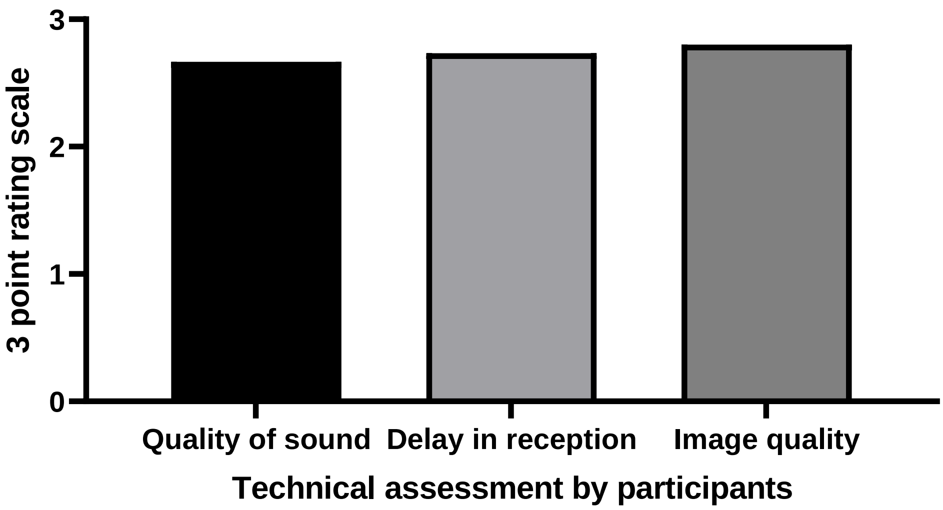


4.6.2 Technical quality based on assessment by the participant

Figure 4.2 represents the technical quality assessment which was done by the participants with stuttering, at the end of the online session. It was based on three point rating scale, where '3' was considered as highly satisfied, '2' - somewhat satisfied and '1' - not at all satisfied, with respect to quality of sound, delay in reception and image quality for all the fifteen participants with stuttering. Among fifteen participants, delay in reception and image quality was rated to be 'highly satisfied' (3) by eleven participants and as 'somewhat satisfied' (2) by four participants. Quality of sound was rated to be 'highly satisfied' (3) by ten participants, while others rated as 'somewhat satisfied' (2).

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).

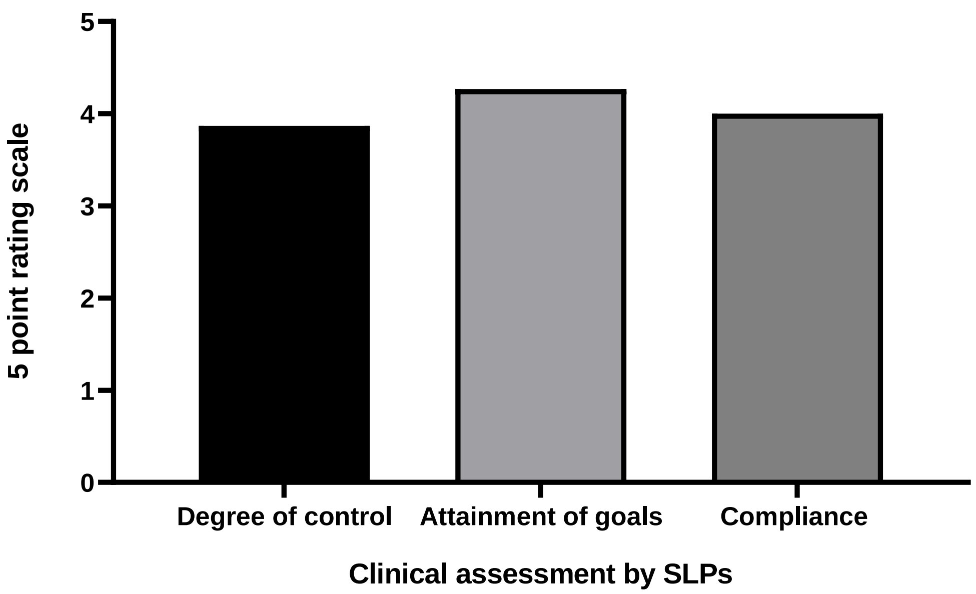


4.6.3 Clinical quality based on assessment by the SLPs

Figure 4.3 represents the average scores of clinical quality assessment which was done by 3 SLPs, by analyzing the recorded video of the sessions for each participant. The assessment was based on five point rating scale, where ‘1’ indicates ‘highly dissatisfied’ and ‘5’ indicates ‘highly satisfied’ with respect to degree of control, attainment of goals and compliance, for all the fifteen participants with stuttering. Among fifteen participants; five participants indicated ‘highly satisfied’ (5) rating for attainment of goals. The remaining ten participants gave a rating of ‘3’ and above. Degree of control was rated to be highly satisfied (5) by the two participants; thirteen participants gave a rating of ‘3’ and above. Compliance was rated as ‘highly satisfied’ (5) by one participant, whereas all others gave a rating of ‘3’ and above.

Figure 4.3

Clinical quality based on evaluation by SLPs on a 5 point scale (1 -highly dissatisfied to 5 -highly satisfied)

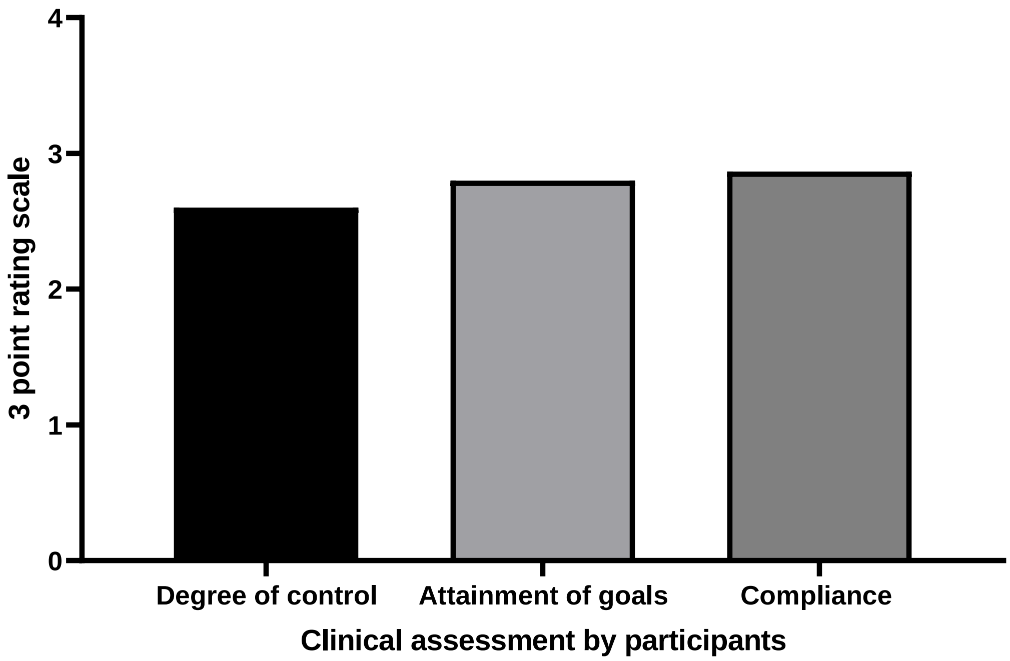


4.6.4 Clinical quality based on assessment by the participant

Figure 4.4 represents the clinical quality assessment which was done by the participants, at the end of the online session. It was done on three point rating scale, where '3' indicates 'highly satisfied', '2' indicates 'somewhat satisfied', and '1' indicates 'not at all satisfied', with respect to degree of control, attainment of goals and compliance for all the fifteen participants with stuttering. Out of fifteen participants, eleven participants rated 'highly satisfied' (3) for attainment of goals and compliance, whereas four participants rated 'somewhat satisfied' (2). Nine participants rated 'highly satisfied' (3) for degree of control and six participants rated 'somewhat satisfied' (2)

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)



Chapter V

DISCUSSION

The present study attempted to answer the following questions:

1. Is it feasible to assess through automatic speech processing, the frequency and duration of filled pauses, repetitions and prolongations?
2. Is there a significant similarity between the value of fluency parameters derived through automatic speech processing and the values obtained through online perceptual assessment?
3. Is the accuracy of fluency parameters derived through automatic speech processing different across various levels of severity?
4. Is the technical and clinical quality of the online assessment satisfactory?

5.1 Selection of participants and their characteristics

In Group I, a total of 30 adult (Mean age = 25.80 and SD = 3.25), literate, native Kannada speakers participated in the study. 15 male and 15 female participants were selected through an interview and screening/testing procedures. The participants had no difficulty in understanding speech and had no hearing-related, or otological, or neurological problems and also had no problems related to vision, communication, and cognition. In Group II, 15 adult literate (11 Male and 4 Female), native Kannada speakers (Mean age = 25.40 and SD = 4.437), who were diagnosed by qualified SLPs to have mild to severe stuttering by certified SLPs, participated in the study. It was ascertained through a structured interview that, all the participants who were included knew how to use Zoom application for video call. It was also ascertained by the researcher that the participants had no symptoms of neurological, social, emotional, or psychiatric disorders.

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

Frequency and duration of the fluency parameters were derived from the online recorded passage through a code written in Matlab. The algorithm was effective in identifying the stuttering events, as shown by the 100% agreement while assessing duration (Table 4.3 and 4.4). However, the error in detecting the frequency was found to vary between 26% to 35% (Table 4.7). While running the algorithm on recorded speech, it was observed that there is difficulty in identifying speech variability such as intra and inter-speaker variability, recognition units, language complexity, ambiguity, and environmental conditions. In the algorithm used for the present study, MFCC was used as the method of feature extraction. MFCC works well with noiseless speech data (Khara et al., 2018).The speech data used in this study was recorded online through Zoom app. The speaker was at his home while recording and hence the recording environment was not noiseless always. RASTA performs very well with noisy speech signals, but it needs to be used with PLP for better accuracy (Khara et al., 2018). The error in the number of stuttering events identified through automatic speech processing is between 26% and 35% (in comparison with online perceptual assessment) in the present study. If the error can be reduced by using RASTA with PLP for feature extraction, it is feasible to assess the fluency parameters through automatic speech processing. We were unable to implement RASTA with PLP in our study because of the time constraint. RASTA and PLP require complex algorithms for implementation, whereas MFCC based algorithms are comparatively simpler.

Another probable reason for lower accuracy in the present study could be the lower sample size. The automatic speech processing systems become well-trained with a larger data size. As the sample size in the present study was only 15, the system could not be trained properly. Thus it is expected that the accuracy of the system can be further improved by training the system on a larger sample size.

ASR poses several challenges. Those are variability in speech, recognition units, language complexity, and ambiguity. Also speaker, gender variation, speaking rates, vocal efforts, regional accents, speaking styles creates a wide range of variations. Also, ASR does not account for physical concomitants of the participants for assessment of stuttering concerning the severity, where SSI4 includes physical concomitants for the severity. Considering all these challenges, an accuracy of 74% in identifying the word repetitions achieved in the present study is promising.

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

No significant differences were observed in the duration; a 100% agreement was found between perceptual evaluation and automatic speech processing. A significant difference is seen in the frequency of prolongation and part-word repetition. No significant difference in the frequency of filled pauses and word repetition was observed between the online perceptual evaluation and automatic speech processing. The prolongation was observed more on vowels, and part-word repetition types of dysfluencies were observed on bilabial sounds and retroflex. Total number of prolongations (55) and part-word repetitions (108) were higher compared to filled pauses (21) and word repetitions (Table 4.7). This could be the reason for showing a significant difference between the frequency of prolongations and part word repetitions derived through automatic speech processing and the values obtained through perceptual assessment. Comparison between perceptual assessment and automatic speech processing was also done by calculating the error percentage, considering perceptual assessment as the reference. The percentage error was found to be 33%, 35%, 31% and 26% respectively for filled pauses, prolongations, part-word repetitions and word repetitions. The automatic speech processing based system developed by Ghonem et al. (2017) for classifying speech samples as normal, repetition or prolongation showed an accuracy of 69.56%, 40%, 50% for repetition, prolongation and pauses respectively. The accuracy obtained in the present study is 67%, 65%, 69% and 74% respectively for filled pauses, prolongations, part-word repetitions and word repetitions. This is comparable with the system developed by Ghonem et al. (2017).

In a study done by Esmaili et al. (2016), dysfluency classification was done using MFCC, PLP and Filter Bank Energies (FBE). Sensitivity of 96.16% in detection of prolongation was obtained when PLP features were used. Best specificity of 99.95% was observed in phrase repetition when MFCC features were used. For prolongation detection the classification accuracy of MFCC features with correlation similarity measure was found to be 99.82%. Syllable or word repetition detection rates for FBE features are 99.85%. When PLP and MFCC features were combined, accuracy, and sensitivity and specificity were found to be 97.80%, 85.45%, and 98.09% respectively. The higher accuracy reported in their study may be due to the reason that they used PLP and FBE features in addition to MFCC features for identifying the dysfluencies. PLP and FBE works better in noisy situations. We could not use PLP and FBE in our study because of the complexity in developing these algorithms. Moreover, for these algorithms to work effectively a larger sample size is required. Our sample size was limited to 15.

PLP based feature extraction showed better performance in prolongation detection and MFCC based extraction showed better performance in repetition detection (Esmaili et al., 2016). Both PLP and MFCC features resulted in the recognition accuracy of 97.80%, 85.45%, and 98.09%.

5.4 Accuracy of fluency parameters derived through automatic speech processing across various levels of severity

Accuracy was calculated for each of the stuttering events such as pauses, repetitions (part word and word) and prolongations in terms of error percentage. The percentage error was found to vary between 25% and 36%. Independent-Samples Kruskal-Wallis Test was done for each fluency parameter to find out whether the distribution of percentage error values was same across the mild, moderate and severe categories. The test revealed that percentage error values of fluency parameters did not show any relation to the mild, moderate or severe category.

Esmaili et al. (2016), extracted features like MFCC, PLP or FBE from speech samples of 39 adults with stuttering to detect sound prolongation and repetition in stuttered speech. The average accuracy achieved for detection of prolongation, word and phrase repetition were 99.84%, 98.07% and 99.87%, respectively across different severe category. Results of their study are similar to the present study considering the observation that the accuracy of automatic speech recognition did not show any relation to the severity of stuttering. The factors influencing the accuracy of the automatic speech recognition system include difficulty in identifying speech variability such as intra and inter-speaker variability, recognition units, language complexity, ambiguity, and environmental conditions (John, 2013). As these factors are independent of the severity, it may be concluded that the severity of the disorder does not influence the error in detection.

5.5 Technical and clinical quality of the online assessment

Technical and clinical quality of the online sessions was evaluated by the participant as well as three experienced SLPs. The SLPs used the videos of the sessions recorded through Zoom for quality evaluation. Technical quality evaluation by the SLPs focused on three aspects: - quality of sound, delay in reception and image quality. The rating was done on a five point scale, where '5' indicates highly satisfied and '1' indicates highly dissatisfied. All the three aspects were rated as '3' and above (average of the rating by three SLPs) for all the participants (Table 4.8). This indicates that the quality of sound and image was satisfactory and also there was no unsatisfactory delay reported in signal reception. Clinical quality was evaluated by the SLPs on 3 aspects such as degree of control, attainment of goals and compliance. Average scores of 3 SLPs were '3' and above for all the participants (Table 4.10), for all the three aspects. This shows that the researcher had good degree of control, was successful in attaining the set goals and also was compliant. Similar results were obtained by Sicotte et al. (2003) where 6 adult participants were rated by the SLPs with respect to technical and clinical quality on 6 items with the same rating scale used in present study. Technical quality was judged by the SLP as being moderately good. Overall, 50% of the session ratings were of 3 and above on the five-point scale and 43% of the session ratings were of 4 and above for the technical quality. In the present study also, technical quality judged by the SLPs as being good and overall 56% of the technical parameters ratings were of 4, 18% of the technical parameters ratings were of 5 and 27% of the technical parameters rating were of 3 on five-point rating scale.

In Sicotte et al. (2003), among the three indices of technical quality, image quality was judged the least successful, with 63% of the ratings in the middle of the scale. Clinical quality was judged more positively by the SLPs where 81% of the ratings were on the positive side of the scale that is the SLP was satisfied 53% of the time and highly satisfied 28% of the time.

In the present study, clinical quality was judged by the SLPs as good and overall 67% of the clinical parameters ratings were of 4, 18% of the clinical parameters ratings were 5 and 16% of the clinical parameters ratings were 3 on the five-point rating scale. Overall, SLPs judged clinical quality as good and almost 67%-81% of the ratings were on the positive side of the scale.

Technical and clinical quality was also assessed by the 15 stuttering participants at the end of each session. Technical quality evaluation by the participants was based on the same three aspects (quality of sound, delay in reception and image quality) as used by the SLPs. A three-point rating scale was used, where ‘3’ was considered as highly satisfied, ‘2’-somewhat satisfied and ‘1’-not at all satisfied. All the participants rated ‘2’ and above for all the three aspects (Table 4.9). This shows that the participants were either ‘somewhat satisfied’ or ‘highly satisfied’ with the quality of sound and image quality. Not even a single participant gave ‘not satisfied’ rating for delay in reception. Clinical quality evaluation by the participants was based on the same three aspects (degree of control, attainment of goals and compliance) as used by the SLPs. Instead of a five-point scale, a three-point rating scale was used. All the participants gave a rating of ‘2’ and above for all the three aspects (Table 4.11). This indicates that the participants were satisfied with the online session conducted by the researcher. Similar results were obtained by Sicotte et al.

(2003) where they used the similar patient satisfaction with measures for assessment of technical quality and clinical quality of the intervention, on a rating scale similar to the one used in the present study. Sicotte et al., 2003 observed that technical quality was scored at the highest level by all but one of the participants.

In the present study the participant's satisfaction with respect to the technical quality was assessed on rating scale which was similar to the Sicotte et al. (2013) study. Perception of all the fifteen participants was positive at both the technical and clinical levels. In technical quality among fifteen participants, delay in reception and image quality was rated to be 'highly satisfied' by eleven participants and as 'somewhat satisfied' by four participants. Quality of sound was rated to be 'highly satisfied' by ten participants, while others rated as 'somewhat satisfied'. With respect to clinical quality, out of fifteen participants, eleven participants rated 'highly satisfied' for attainment of goals and compliance, whereas four participants rated 'somewhat satisfied'. Nine participants rated 'highly satisfied' for degree of control and six participants rated 'somewhat satisfied'.

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

When the results of automatic speech processing based assessment were compared with the results of perceptual evaluation significant difference was observed in prolongation and part word repetition. Whereas, filled pause and word repletion did not show any significant difference. These results show that speech processing can be used for online assessment for evaluation of stuttering with improvements in recognizing prolongations and part word repetition. One of the limitations observed in the present study was the noise in the online recorded samples, which is also the reason for the low accuracy. The error in detection was found to vary between 26% and 35%. It is possible to reduce the error if RASTA is used with PLP (Khan et al., 2019), instead of MFCC. This will enhance the recognition of speech in noisy conditions and therefore will provide better accuracy. The technical and clinical quality evaluation of the online sessions by the SLPs and the participant provide satisfactory feedback. Considering the above, it can be concluded that it is feasible to use automatic speech processing based assessment as an online technique for assessment of stuttering.

Chapter VI

SUMMARY AND CONCLUSIONS

Many researchers have attempted to assess fluency disorders using automatic speech processing. However, no research was conducted on online assessment of stuttering in Kannada speaking adults deriving fluency parameters through automatic speech processing. In the present pandemic situation, it is difficult to have the conventional face to face assessment. The possibility of conducting online assessment of fluency parameters in Kannada speaking adults was explored in the present study.

A total of 45 literates, Kannada speaking participants (age range: 18-35 years) divided into two groups, (Group I - 30 normal participants, and Group II -15 participants with stuttering) participated in the study. The researcher used the most widely used video conference app - Zoom for assessment through virtual mode. The standardized passage in Kannada (Annexure-A) was used as the reading material to assess stuttering through automatic speech processing. The researcher assessed the severity of stuttering for fifteen adults with stuttering where nine participants were diagnosed to have moderate stuttering; three each were diagnosed to have mild and severe stuttering respectively. The researcher also assessed the frequency and duration of prolongations, filled pauses and repetitions from the recorded passage through perceptual evaluation. The fluency parameters were also assessed through automatic speech recognition from the recorded standard passage for all the fifteen participants with stuttering. At the end of each session, satisfaction of the participant in Group II, with respect to the technical and clinical quality of the online session was assessed by the respective participant. Each recorded session was also rated by three Speech Language Pathologists for its technical and clinical quality.

Results showed no significant difference between perceptual evaluation and automatic speech recognition for filled pause and word repetition. Overall technical quality and clinical quality were satisfactory as per the judgment of 3 SLPs and the participants.

6.1 Important results of the study

The important findings of the study are summarized below:-

- The present study showed concurrence between perceptual evaluation and automatic speech recognition for frequency of filled pauses and word repetitions.
- The study also showed 100% agreement between perceptual evaluation and automatic speech recognition for duration of all fluency parameters.
- The accuracy of detection of fluency parameters through automatic speech processing did not vary across mild, moderate and severe category of stuttering.

6.2 Implications of the study

- The study has established that assessment of fluency parameters through automatic speech processing can be effectively used for online assessment of fluency disorders.
- The study has also established that the technical and clinical quality of online assessment is acceptable to the professionals as well as the patients. Hence, the study has shown the feasibility of employing telepractice for assessment of fluency disorders.

6.3 Limitations of the present study

- Only the words in the standard passage were considered for assessment in the study.
- Assessment of fluency parameters using automatic speech processing was done only for reading task.
- The study considered only adult population and there was no gender balance in the participants.
- Sample size considered for the study is relatively small.
- ASR does not account for physical concomitants of the participants for assessment of stuttering.

6.4 Future recommendations

- The study may be extended on a larger sample size representing each **c**lass of severity.
- The study may be repeated with phrases and sentences for spontaneous speech.
- It can be extended over children.
- The study may be repeated on a group of participants with equal gender distribution.

6.5 Significance of the results of the study

Very few studies have been reported on tele-assessment of fluency disorders. Moreover, no research is conducted on online assessment of stuttering in Kannada speaking adults using automatic speech recognition techniques. In the present pandemic situation, it is difficult to have the conventional face to face assessment.

The possibility of conducting online assessment of fluency disorders has been established by the study. The present study also showed that the technical and clinical quality of the online mechanism used in the present study is acceptable to the clinician as well as the patient.

REFERENCES

Arjun, K. N., Karthik, S., Kamalnath, D., Chanda, P., &Tripathi, S. (2020). Automatic Correction of Stutter in Disfluent Speech. *Procedia Computer Science*, *171*, 1363-1370. https://doi.org/10.1016/j.procs.2020.04.146.

Armfield, N.R., Gray, L.C., & Smith, A.C. (2012). Clinical use of Skype: A review of the evidence. *Journal of Telemedicine and Telecare*, *18*, 125-127. https://doi.org/10.1258%2Fjtt.2012.SFT101

Bayerl, S. P., Hönig, F., Reister, J., &Riedhammer, K. (2020, September). Towards Automated Assessment of Stuttering and Stuttering Therapy. In *International Conference on Text, Speech, and Dialogue*, 386-396.Springer, Cham. DOI: 10.1007/978-3-030-58323-1_42

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., &Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech communication*, *49*(10-11), 763-786.https://doi.org/10.1016/j.specom.2007.02.006

Blaiser, K.M., Behl, D., & Callow-Heusser, C. (2013). Measuring costs and outcomes of tele-intervention when serving families of children who are deaf/hard-ofhearing. *International Journal of Tele-rehabilitation*, *5*(2), 3-10.doi: 10.5195/ijt.2013.6129

Caplan, L. (1983). Some comments on aspects of the methodology for the measurement of stuttering. *South African Journal of Communication Disorders*, *30*(1), 30-33.

Carey, B., O'Brian, S., Onslow, M., Packman, A., &Menzies, R. (2012). Webcam delivery of the Camper Down Program for adolescents who stutter: A Phase I trial. *Language, Speech, and Hearing Services in Schools*. https://doi.org/10.1044/0161-1461(2011/11-0010)

Carey, B., O'Brian, S., Onslow, M., Block, S., Jones, M., &Packman, A. (2010). Randomized controlled non inferiority trial of a tele-health treatment for chronic stuttering: The Camper Down Program. *International Journal of Language and Communication Disorders,45,* 108–120. https://doi.org/10.3109/13682820902763944

Chee, L. S., Ai, O. C., Hariharan, M., &Yaacob, S. (2009). Automatic detection of prolongations and repetitions using LPCC. In *2009 international conference for technical postgraduates(TECHPOS),* 1-4.IEEE.https://doi.org/10.1109/TECHPOS.2009.5412080

Cohn, E.R. (2012). Tele-ethics in telepractice for communication disorders.*Perspectives on Telepractice, 2*(1), 3-15 https://doi.org/10.1044/tele2.1.3

Constantino, C. D., Leslie, P., Quesal, R. W., &Yaruss, J. S. (2016).A preliminary investigation of daily variability of stuttering in adults. *Journal of Communication Disorders*, *60,* 39-50.https://doi.org/10.1016/j.jcomdis.2016.02.001

Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research, 37,* 279–294.https://doi.org/10.1044/jshr.3702.279

Cordes, A. K., & Ingham, R. J. (1995). Judgments of stuttered and non-stuttered intervals by recognized authorities in stuttering research. *Journal of Speech and Hearing Research, 38,* 33–41, https://doi.org/10.1044/jshr.3801.33

Craig, A., Hancock, K., Tran, Y., Craig, M., & Peters, K. (2002*). Epidemiology of stuttering in the community across the entire life span.* Davidow, J. H., & Scott, K. A. (2017). Intrajudge and Interjudge Reliability of the Stuttering Severity Instrument–Fourth Edition. *American journal of speech-language pathology*, *26*(4), 1105-1119.https://doi.org/10.1044/2017_AJSLP-16-0079

Denton, D.R. (2003). Ethical and legal issues related to telepractice. *Seminars in Speech and Language, 24,* 313-322.DOI: 10.1055/s-2005-863794

Esmaili, I., Dabanloo N. J., &Vali, M. (2016).“Automatic classification of speech dysfluencies in continuous speech based on similarity measures and morphological image processing tools”, *Biomedical Signal Processing and Control*, *23,* 104-114.https://doi.org/10.1016/j.bspc.2015.08.006

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

Ghonem, S. A., Abdou, S., Shoman, M., &Ghamry, N. (2017). “Classification of Stuttering Events Using I-Vector, *Egyptian Journal of Language Engineering, 4*(1), 11-18.https://dx.doi.org/10.21608/ejle.2017.59395

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

Hayhow, R. (2018). The assessment of stuttering and the evaluation of treatment. In *Approaches to the Treatment of Stuttering*,15-46. Routledge.

Heeman, P. A., Lunsford, R., McMillin, A., &Yaruss, J. S. (2016).Using Clinician Annotations to Improve Automatic Speech Recognition of Stuttered Speech.In *INTERSPEECH*, 2651-2655.http://dx.doi.org/10.21437/Interspeech.2016-1388

Irani, F., & Gabel, R. (2011). Intensive stuttering therapy with telepractice follow-up: A case study. *Perspectives on Fluency and Fluency Disorders, 21,* 11-21.https://doi.org/10.1044/ffd21.1.11

Khan, I., Ullah, A., &Emad, S. M. (2019). Robust Feature Extraction Techniques in Speech Recognition: A Comparative Analysis. *KIET Journal of Computing and Information Sciences*, *2*(2), 11-11.

Khara, S., Singh, S., &Vir, D. (2018). A comparative study of the techniques for feature extraction and classification in stuttering. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 887-893. IEEE.DOI**:** 10.1109/ICICCT.2018.8473099

Kully, D. (2000). Tele-health in speech pathology: Applications to the treatment of stuttering. *Journal of Telemedicine and Telecare, 6,* 39–41. https://doi.org/10.1258%2F1357633001935509

Kumar, N. V., Sai, Y. P., &Prakash, C. O. (2015). “Design and Implementation of Silent Pause Stuttered Speech Recognition System”, *International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, 4(3),* 1253-1260.

Lewis, D., &Doaothy, S. (1951). Measuring the severity of stuttering. *I. Speech Hearing D/s., 16,* 320-326.

Mackay, D. G., & Macdonald, M. C. (1984). Stuttering as sequencing and timing disorder. In *In*.

Mahesha, P., &Vinod, D. S. (2013). "Classification of Speech Dysfluencies Using Speech Parameterization Techniques", *Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, QSHINE* 2013, LNICST 115, 298-308.DOI: 10.1007/978-3-642-37949-9_26.

Manning, W. H., &DiLollo, A. (2017). *Clinical decision making in fluency disorders*. Plural Publishing.

McGill, M., Noureal, N., & Siegel, J. (2019). Telepractice Treatment of Stuttering: A Systematic Review. Telemedicine and e-Health, 25(5), 359-368 https://doi.org/10.1089/tmj.2017.0319.

McGill, M., Cullen, L., & Webb, H. (2019). Clients' Experiences of Telepractice for Stuttering. *Perspectives of the ASHA Special Interest Groups*, 1-10.https://doi.org/10.1044/2019_PERS-SIG18-2018-0025

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.

Mendhakar, A. M., Mahesh, S., Rakshith, S. (2016). Application of Automatic Speech Recognition in Communication Sciences and Disorders, *Journal of Acoustical Society of India, 43*(2), 108-115.

Minifie, F. D., & Cooker, H. S. (1964). A disfluency index. *Journal of Speech and Hearing Disorders, 29*(2), 189-192.

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.

Nöth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., & Wittenberg, T. (2000). Automatic stuttering recognition using hidden Markov models. In *Sixth International Conference on Spoken Language Processing*.

Ravikumar, K. M., Reddy, B., Rajagopal, R., &Nagaraj, H. (2008). Automatic detection of syllable repetition in read speech for objective assessment of stuttered dysfluencies. *Proceedings of world academy science, engineering and technology*, *36*, 270-273.

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., & Bakker, K. (2009). *SSI-4: Stuttering severity instrument*. PRO-ED, an International Publisher.

Rousseau, I., Packman, A., Onslow, M., Dredge, R., & Harrison, E. (2002).Australian speech pathologists' use of the Lidcombe Program of early stuttering intervention. *Acquiring Knowledge in Speech, Language and Hearing*, *4*(2), 67-71

Sicotte, C., Lehoux, P., Fortier-Blanc, J., & Leblanc, Y. (2003). Feasibility and outcome evaluation of a telemedicine application in speech–language pathology. *Journal of Tele-medicine & Tele-care*, *9*(5), 253-258.https://doi.org/10.1258%2F135763303769211 256

Świetlicka, I., Kuniszyk-Jóźkowiak, W., & Smołka, E. (2009). “Artificial Neural Networks in the Disabled Speech Analysis”, In: Kurzynski M., Wozniak M. (eds) Computer Recognition Systems 3, *Advances in Intelligent and Soft Computing, Springer, Berlin, Heidelberg, 57,* 347-354.https://doi.org/10.1007/978-3-540-93905-4_41

Tahmasebi, N., Shafie, B., Karimi, H., &Mazaheri, M. (2018). A Persian-version of the stuttering severity instrument-version four (SSI-4): how the new additions to SSI-4 complement its stuttering severity score? *Journal of communication disorders*, *74*, 1-9. https://doi.org/10.1016/j.jcomdis.2018.04.005

Todd, H., Mirawdeli, A., Costelloe, S., Cavenagh, P., Davis, S., & Howell, P. (2014). Scores on Riley’s stuttering severity instrument versions three and four for samples of different length and for different types of speech material. *Clinical linguistics & phonetics*, *28*(12), 912-926.https://doi.org/10.3109/02699206.2014.926991

Van Riper, C. (1982). The nature of stuttering. Prentice Hall. DOI: https://doi.org/10.1017/S0142716400004458

Waghmare, S. D., Deshmukh, R. R., Shrishrimal, P. P., Waghmare, V. B., Janvale, G. B., &Sonawane, B. (2017). A Comparative Study of Recognition Technique Used for Development of Automatic Stuttered Speech Dysfluency Recognition System. *Indian Journal of Science and Technology*, *10*(21), 1-14.DOI: 10.17485/ijst/2017/v10i21/106092

Yairi, E., Ambrose, N., & Cox, N. (1996). Genetics of stuttering: a critical review. *J Speech Lang Hear Res 3*9, 771–784.

Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., Danielson, M. L., Bitsko, R. H., & Boyle, C. A. (2019). Prevalence and trends of developmental disabilities among children in the United States*: 2009–2017*. *Pediatrics*, *144*(4).

Annexure- A

KANNADA PASSAGE

ಬೆಂಗಳೂರು ನಮ್ಮ ರಾಜ್ಯದ ದೊಡ್ಡ ಊರು. ಈ ಊರನ್ನು ನಮ್ಮ ರಾಜ್ಯದ "ಬೊಂಬಾಯಿ" ಎನ್ನುವರು. ಇಂಡಿಯಾದ ದೊಡ್ಡ ನಗರಗಳಲ್ಲಿ ಇದೂ ಒಂದು. ಈ ಊರನ್ನು ನೋಡಲು ಜನರು ಬೇರೆ ಬೇರೆ ಊರುಗಳಿಂದ ಬರುವರು. ಇದಲ್ಲದೆ ನಮ್ಮ ರಾಜ್ಯದಲ್ಲಿರುವ ಬೇಲೂರು, ಜೋಗ್, ನಂದಿ, ಇವುಗಳನ್ನು ನೋಡಲು ಜನರು ಬರುವರು. ಈ ನಾಡಿನಲ್ಲಿ ರೇಷ್ಮೆಯನ್ನು ಬೆಳೆಯುವರು.

ಕೃಷ್ಣಾ ನದಿಯು ಸಹ್ಯಾದ್ರಿ ಪರ್ವತಗಳಲ್ಲಿ ಮಹಾಬಲೇಶ್ವರದ ಹತ್ತಿರ ಹುಟ್ಟುತ್ತದೆ. ಈ ಪ್ರದೇಶವು ರಮಣೀಯವಾದ ಸ್ಥಾನ. ಇದು ಮಹಾರಾಷ್ಟ್ರ, ಕರ್ನಾಟಕ ಮತ್ತು ಆಂಧ್ರಪ್ರದೇಶಗಳಲ್ಲಿ ಹರಿದು ಬಂಗಾಳ ಕೊಲ್ಲಿಯನ್ನು ಸೇರುತ್ತದೆ. ಇದಕ್ಕೆ ಉಪನದಿಗಳು ಹಲವು. ಕೊಯಿನಾ, ತುಂಗಭದ್ರಾ, ಘಟಪ್ರಭಾ, ಭೀಮಾ, ಮಲಪ್ರಭಾ – ಅವುಗಳಲ್ಲಿ ಕೆಲವು. ಕೊಯಿನಾ ನದಿಗೆ ಅಣೆಕಟ್ಟನ್ನು ಕಟ್ಟಿ ವಿದ್ಯುತ್ತನ್ನು ಉತ್ಪಾದನೆ ಮಾಡುತ್ತಾರೆ.