Indian Journal of Science and Technology, *Vol 10(21), DOI: 10.17485/ijst/2017/v10i21/106092, June 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Comparative Study of Recognition Technique Used for Development of Automatic Stuttered Speech Dysfluency Recognition System

Swapnil D. Waghmare[1]*, Ratnadeep R. Deshmukh[1], Pukhraj P. Shrishrimal[1], Vishal B. Waghmare[2], Ganesh B. Janvale[3] and Babasaheb Sonawane[4]

[1]Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad – 431004, Maharashtra, India; waghmare.swapnil21@gmail.com, ratnadeep_deshmukh@yahoo.co.in, pukhraj.shrishrimal@gmail.com,
[2]Department of Computer Science, Vivekanand College, Tarabai Park, Kolhapur – 416003, Maharashtra, India; vishal.pri12@gmail.com
[3]MGM's Institute of Biosciences and Technology, Aurangabad – 413003, Maharashtra, India; ganesh.janvale@gmail.com
[4]Department of Computer Science and Engineering, Maharashtra Institute of Technology, Aurangabad – 411038, Maharashtra, India; bssonawane@gmail.com

Abstract

Objectives: This paper is an attempt to compare the work done around the world for development of stuttered speech database and approaches for analysis of stuttered speech and recognition system. **Methods/Statistical Analysis:** In particular we have compared the different methods adopted by the researchers around the world for development of speech database and the techniques implemented on these developed databases. We have compared the databases on the basis of utterances, gender, age group, speech dysfluencies and type of samples. The recognition systems are compared on the basis of feature used, classification techniques and the accuracy. **Findings:** Speech recognition based application is getting more popularized and now being implemented at various places. However, the developed speech recognition systems cannot handle the speech dysfluencies. Very less work had been carried out till date for stuttered speech recognition system. The work for Indian languages is very negligible. The only work carried out is for Kannada. There is no major contribution for other Indian Languages. This paper shows the current status and the notable work carried in other languages. **Application/Improvements:** There is a need to develop more such systems for other Indian languages which will be very helpful for multilingual society like India.

Keywords: Stuttered Speech, Stuttered Speech Dysfluency Recognition System, Speech Dysfluency, Stuttered Speech Database

1. Introduction

Speech is the most effective and developed biological means to express feelings, ideas and thoughts via interpersonal verbal communication. Speech contains precise pathological acoustic and linguistic background description which consist a set of information, which is not limited to verbal but also consists of emotional state and intension. Production of a fluent speech requires the combination of cognitive, linguistic and motor processes[1–5].

The main aim of speech is to convey messages in a linguistic form and it consists of articulation, voice and fluency pattern. Sometimes speech becomes unintelligible in some people due to some disorder. These dysfluency in speech is badly affect the performance of Automatic Speech Recognition (ASR) system and makes such system

**Author for correspondence*

unusable to the user suffering from speech disorders. The study of Dysfluencies for speech based system and analysis has gained more attention in the field of healthcare, military, security and machine learning scenarios.

This study report focuses on the speech disorders observed in the human being. The paper is divided in following six sections. Section 1 gives introduction and Section 2 explains about speech disorder and it gives standard definitions and different categories of speech disorder observed in the human being. Section 3 gives idea about various stuttered speech database developed by institutions in all over the world. Section 4 elaborate the various acoustic models are established for recognition. Section 5 discusses the various speech and no speech characteristics Section 6 gives conclusion[4–9].

1.1 Language Disorder Include Following Disorder

1.1.1 Cluttering

Cluttering disorder concern with lack of language knowledge. The rapid rate with breakdown fluency but not included repetition or hesitation. Speech is erratic and dysrhythmic, the person who clutter they are thinking and disorganize while actually speaking, because of disorganization their speech in bursts or rapid jerky sound.

1.2 Speech Disorder Include Following Disorders

1.2.1 Aprexia (Dyspraxia)

Aprexia identified as oral-motor speech disorder. It is the serious impairment of development of motor coordination that is not solely explicable in terms of general intellectual retardation, in this disorder patient has problem for muscle movement and difficult to formulate speech sound into words.

1.2.2 Articulation Disorder

Articulation is known as *artic* disorder, person having difficulties in pronouncing speech sound is below the appropriate level of its mental age. Functional causes having difficulty to producing speech sound of phonemes correctly.

1.2.3 Stuttering

By enter into stuttering, it encompasses threefold type of symptoms: **linguistic** (disturbance of speech flow, as suddenly fluctuating standard rhythm, phonation break, involuntary repetition, prolongation, extraneous delay, producing additional sound.), **psychological** (fear of speech surrounded by particular fear of speaking, lack of enthusiasm, unwillingness to speak, also known as logophobia) and **neurophysiological** (accompanied by involuntary muscle contradiction, dis-coordination of various articulation, respirator and phonetic muscle, motor control)[10,–15]

2. Stuttered Speech

Speech is not sound in a smooth manner; it is disrupted by disorder with unspecified etiology. It is commonly assumed that stuttering consequences from cohesion of biological[16], psychological and even social reactions. Voice carries various acoustic and linguistic characteristic like basic pitch subsequent formant frequency etc. Every individual involves different noticeable flow of a speech, rhythms[17] determined by structure and arrangement of larynx, pharynx, oral and nasal activity, paranasal sinuses, and thorax. The description of stuttering is still lacks of consensus to consistently constitutes sphere of stuttering in despite of carries large number of research. Spoken languages influence to communication among the human being. Speech has the ability of being used as mode of interaction with computer[18–20]. Human beings have been motivated to develop a system that can understand and recognize voice for normal speech. Based on previous research we found types of stuttering.

2.1 Types of Stuttering

2.1.1 Developmental Stuttering

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

2.1.2 Neurogenic Stuttering

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

2.1.3 Psychogenic Stuttering

Psychogenic stuttering is directly connected with patients' mental stress and speaking behaviors[21,22]. In the United Kingdom [UK] stuttering is identified as stammering, it can found to be a very serious and complex disorder in speech pathology. As per the global scenario it occurs in approximately 1% of the entire population and has found that it affects 1:3 or 4 times in female to male ratio. This dysfluency may cause difficulties in interpersonal language communication as well as reluctance to speak, sense of guilty, and low self-esteem[23].

World Health Organization (WHO) recognize stuttering with the code *F98.5*, and finalize stable definition as a speech having *frequent repetition or prolongation of sounds or syllables or words, or by frequent hesitations or pauses that disrupt the rhythmic flow of speech*[14]. It is evident from the past and concurrent literatures survey that, stuttering can be appraised a genetic disordered since 1930s[10,14,24] and employed several different dysfluencies are: Interjection, revision, repetition, prolongations, and blocks.

- **Interjection**

These are negligible extraneous sound, word, syllable or phrase that doesn't alter the meaning of the original sentence. It depends on language, in English "*um, uh, well, like*" are frequently occurs. For example: The baby *um-um- uh* was *um um* hungry. Interjections are also known as *filled pause or fillers*.

- **Revisions**

It occurs when speaker corrects the contents or grammatical structure of phrases. It can change the meaning of original message. Example of revision is *broken words,* sentence like "*I'd like to chang... I'll modify....*".

- **Repetition**

It takes place if any part of sentence is involuntary said more than one. There are two repetition phenomenons, first is *Syllable repetitions:* A syllable or sound is repeated at the beginning of the words. For example: "I had a *c-c-c-coffee*". Another phenomenon is *Word repetitions:* Here example: "The *baby-baby* had the soup".

- **Prolongation**

Unduly prolong the sounds or syllable for example: *muuuuummmy* has gone there.

- **Broken Words**

When speaker trying to pronounce syllable with too much force, and broke the whole word with pause. Example: "It was *won[pause]derful*"[23–27].

Larynx and brain activity of stutter patient had studied by Dr. Freeman, speech pathologist from University of Texas Medical Centre and she found that normal speech affected by malfunctioning of larynx muscle that rapidly control the transition between opening and closing configuration of vocal cord. Muscle impairment is linked with less blood flow, there is either increase or decrease electrical activity in region of the brain which involved in speech production[28]. Figure 1 illustrates the structures involved in speech and voice production.

Research carried out by[1], it was observed that stuttering cannot be thoroughly cured, but it may go into remission of time. People Who Stutter (PWS) can overcome from speech dysfluency by shaping the tempo, loudness, or duration of their utterance and learn to control the speech fluency under the supervision of appropriate speech pathology treatment[1]. In some research motor measures are achieved good result, but on other hand audiovisual indication are used to better classify occurrence of stuttering. The study carried out by *Archibald and de Nil* used kinematic measures that are observed the movement of jaw to measure the relationship between speech motor deficiencies and stuttering severity, helpful for differentiate between fluent and stuttered speaker[29]. From an acoustic point of view, disordered speech can be analysis through the electric signal processing. The result of this analysis extracts the information about the articulation process and can form the basis for diagnosis of the patient[29–31].

In earlier stuttering, repetition and prolongation are the ubiquitous dysfluency, as opposed to the other type of dysfluency. Hence, they are commonly used in stuttering

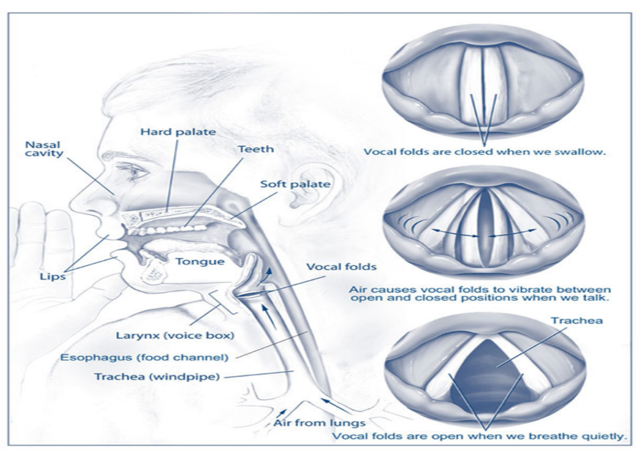

Figure 1. Structures involved in speech and voice production.

assessment process for analysis the performance of stutter before and after therapy. Conventionally, a Speech Language Pathologists (SLP) are employed to counts the dysfluencies, measurement of severity and classify the episode of stuttering manually, to keep track the enhancement of treatment. These types of stuttering assessment is subjective, inconsistent, time consuming and prone to error[32,33].

Therefore, it might be better if stuttering assessment can be done automatically and get the maximum time for the treatment session. One of the important aspects of dysfluencies detection in speech technology is to accumulate the ASR system to decrease the recognition error. In past two decades owing to cutting edge modern electronic multimedia system can be useful to many researchers to developed objective methods, procedures and norms for disfluencies recognition, identify characteristics speech parameters and voice synthesis and also develop different stuttering devices based on Altered Auditory Feedback namely are: Delayed Auditory Feedback (DAF), Frequency Shifted Auditory Feedback (FAF), and Masked Auditory Feedback (MFA) and also Digital Speech Aid (DSA) is widely used to rehabilitation the stuttering and facilitate the SLP during therapy[34–37]. Figure 2 shows the general block diagram of stuttering recognition. *Kuniszyk-Jozkowiak* compared the speech envelops of utterance of fluent and stuttered speakers[38]. The most often used techniques of facilitating speech fluency are called as: *fluency shaping* and *stuttering modification*[39].

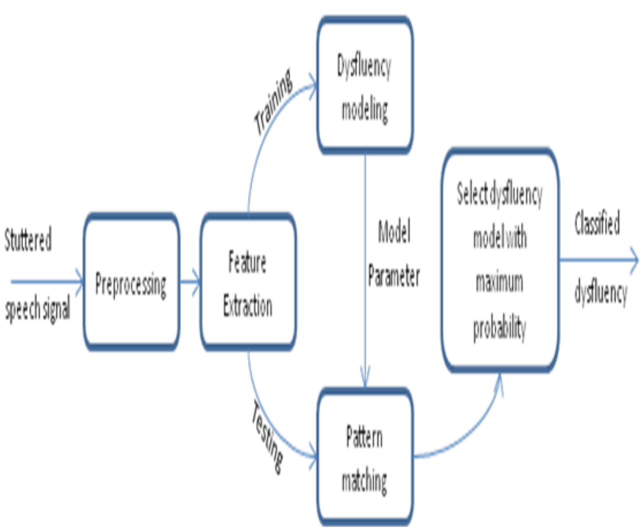

Figure 2. General block diagram of stuttering recognition system.

2.2 Fluency Shaping

Fluency Shaping focuses on the speaker's speech motor control capabilities and apply various approaches to facilitate new speech production patterns. A drawback of this method is that it does not incorporate the individual's feelings and reactions to the disorder[40]. Furthermore, Guitar and Peters stated that by using this method, speech becomes monotonous and artificial as the client strives for fluent speech[41,42].

2.3 Stuttering Modification

The goal of stuttering modification is to reduce speech-related avoidance behaviors, negative attitudes, and fears. This may be skilfully done by reducing struggle behaviours, tension, and the rate of stuttering. The World Health Organization (WHO) would therefore not consider the treatment effective because it does not reduce the impairment level[42].

3. Stuttered Speech Database

There are several database has been developed for stuttered speech for diagnosis and recognition purpose, but most of the work carried out on the UCLASS database. Following are the details of Speech database developed for the stuttered recognition and analysis purpose.

3.1 UCLASS Stuttered Speech Database

A spontaneous stuttered speech database of English language has been developed to establish acoustic properties with various research application areas and for training clinicians about the characteristics as well as investigate language and speech behaviour of stuttered speaker at Psychology Department of University College London (UCL) over the last 20 years[43], known as UCLASS i.e., UCL Archive of Stuttered Speech. UCLASS releases in version 1 and version 2 which is online freely accessible since 2004 to compare results of work around the world. UCLASS version 1 contains monologues recording, it includes 18 females and 120 male speakers, whose ages ranges from 5 years 4 months to 47 years. UCLASS version 2 consists of three types of recording: monologs, reading and spontaneous conversation. The samples are chosen to cover broad range of age from 5 years 4 months to 20 years 7 months. The speech sample with the content of 'One more week to Easter' and 'Arthur the rat', each of these two passage contains more than 300 words (90% it's content is monosyllable words). Most of the records in UCLASS contains from school-age children who were referred to clinics in London. The database is available in wav, mp3, SFS, CHILDES and PRAAT format. Table 1 show the details description about the UCLASS database[44,45].

3.2 Polish Language

A Stuttered Speech Database of Polish language has been developed for recognition and categorization of various speech disorders like blocks, Syllable repetitions, Prolongation at Department of Physics, University of Life Science, Lublin, Poland. This Polish Language Speech Database consist of 33 fluent and non-fluent (dysfluency) speakers, out of which 153 utterances with 4 second fragments of 19 dysfluency patients recordings were differentiated according to gender and age groups which were selected for analysis. These syllables were recorded under various stages of therapy and two situations such as reading story and illustrations. The recording was done in acoustic refuse with 22050Hz frequency. The vocabulary size of polish database is shown in Table 2[15].

3.3 Northern Sotho Language

University of Limpopo, Sovenga, South Africa has been developed a Dysfluency Speech Database of Northern Sotho Language which one of the official language out of eleven official language of South Africa for enhancing the existing automatic speech recognition system for speech disordered people of South Africa. This database contains the 1642 speech syllable of various age groups available on Center of Excellence for Speech Technology. Perl scripts and Bash is used for language implementation. The main aim behind the speech collection of these languages is to develop a speech interaction system and Speech Synthesis for eleven official language of South Africa[46].

Table 1. UCLASS database

Version	Sample Type	Age	Gender	
			Male	Female
Version 1	Monolog	5years 4 month – 47 years 0 month	120	18
Version 2	Monolog	7years 10 month – 20years 1 month	76	6
	Reading	7 years 10 month – 20 years 7 month	93	15
	Conversion	5years 4 month – 20 years 7 month	110	18

Table 2. Polish language speech database

Stuttering Types	Non-Fluent (Patient)		Fluent Speaker	
	Gender	Age	Gender	Age
Blocks	9 Male, 2 Female	10y – 23y	4 Male, 4 Female	22y – 50y
Syllable Repetitions	4 Male, 2 Female	11y – 23y	4 Male, 2 Female	22y – 53y
Prolongation	6 Male, 0 Female	10y – 25y	2 Male, 2 Female	24y – 51y

3.4 Indian Regional Language (Kannada)

Over a last 20 years, All India Institute of Speech and Hearing (AIISH), Mysore, India has been developed a Disfluent Speech database of Kannada Language is one of the south Indian languages of Dravidian origin to develop a Disfluency Assessment Procedure for Children (DAPC) to differentiate between distinguish Normal Non Fluency (NNF) and stuttering in children and classification of speech and non-speech characteristics. This database collects speech sample in two groups. Group I consist 25 disfluent children (18 males and 7 females) and in group II having 26 disfluent children (19 males and 7 females) I the range of 2:3 to 6 years. All the children were allied audiologic problems and mental retardation. The speech sample collect with a detailed 25 questionnaires was prepared based on stuttering chronicity prediction checklist. These questions were designed to extract historical, attitudinal, and behavioural indicators of disfluency, motor, speech and language development, and scholastic history, and sequences of six Panchatantra stories were used to obtain narrative samples from the children. These syllables samples were recorded with the National Panasonic Portable tape recorder[11].

3.5 Peninsular Spanish Language

Monolingual spontaneous conversational Peninsular Spanish developmental stuttering speech database for inspect the developmental change in disfluency from mainly function words to mainly content words of

Spanish at University College London, funded by a grant from the Wellcome Trust. 46 speakers (37 male and 9 Female) 31043 Speech samples are divided into five age groups (G1-G5) from 3 years to 68 years. Database information is shown in Table 3. Speakers are participated from the various parts of Spain are not associated with neurological, psychological, or other medical problems. Speech material contain function word included pronouns, articles, prepositions, conjunctions, and auxiliary verbs, and content words included nouns, main verbs, adverbs, and adjectives[47].

3.6 German Language

Fifty monolingual German stuttered speakers speech database has been developed by the Department of Psychology, University College London for examine and compares overall phonetic complexity between German and English across stuttered and fluent words to examine whether or not they affect stuttering rate. These German speakers are classified into four different age group(1 Adult and 3 children) with 15463 numbers of words. Samples were obtained by spontaneous question and answer dialogs prompt to speakers, duration of sample from 2 minutes to 25 min [48]. German database is shown in Table 4.

4. Existing Established Acoustic Models

Stuttered dysfluencies recognition is a multidisciplinary field of research such as speech pathology, physiology, psychology, signal and acoustic analysis. Based on previous and recent years work, many researchers concentrate on feature extraction algorithms and classification methods to develop recognition system for automatic

Table 3. Monolingual spontaneous conversational peninsular Spanish speech database

Group Name	Age	Speakers	Male / Female	Number of words in samples
G1	3y – 5y	7	3M/4F	2798
G2	6y – 9y	11	9M/2F	5472
G3	10y – 11y	10	9M/1F	6751
G4	12y – 16y	9	9M	4316
G5	20y – 68y	9	7M/2F	11706
Total			**37M/9F**	**31043**

Table 4. Conversational German speech database

Group Name	Age	Speakers	Male / Female	Number of words in samples
G1	2y10m – 6 y5m	11	9M/2F	4135
G2	6y6m – 8y11m	10	8M/2F	5246
G3	9y – 11y11m	14	10M/4F	3228
G4	16y3m – 47y3m	15	10M/5F	2854
Total			**37M/13F**	**15463**

detection of stuttered events and development of speech database for both fluent, non-fluent sample and non-speech characteristics for statistics approach and acoustic analysis. Analysis of stuttered speech includes 1. Mean duration of sound/syllable repetition and sound prolongation, 2. Mean number of repeated units per instance of sound/syllable and whole-word repetition, and 3. Various related measures of the frequency of all between- and within-word speech dysfluencies[49]. Now a day's variety of research focuses on feature extraction, acoustic analysis, design the experiment and evaluate the result of stuttered speech. In this section represent literature survey of previous work which focuses on how the automatic stuttering recognition is being performed. How they develop database, design the experiment, evaluate and assessment of the results, it gives rough idea of different approaches in existing literature. Table 5 is in chronological order of publication, year publication in second column, third column for feature properties, and last column for recognition percentage.

4.1 ANNs

Artificial Neural Networks (ANNs) are newly emerging mathematical tools or computational modelling tools inspired by biological counterparts, that have found that it is used to solve many real world complex problems from various different scientific disciplines such as develop intelligent systems, pattern recognition, prediction, optimization, associative memory and control. ANNs have the remarkable capability of information processing characteristics such as high parallelism, fault and noise tolerance, and learning and generalization capabilities and distinguishing between similar signals[50–52] hence ANNs play remarkable vital role in speech signal processing especially in speech and speaker recognition[11,32,53]. In recent decade history, ANNs are extensively employed

Table 5. Summary of several previous collection of research work on stuttered speech recognition

Year	Database	Features	Classifiers	Results(%)
1995	-	Autocorrelation function and envelope parameter	ANNs	80%
1995-1997	12 speakers	Duration, energy peaks, spectral of word based and part word based.	ANNs	78.01%
2000	51 speakers	Age, sex, type of dysfluency, frequency of dysfluency, duration, physical concomitant, Rate of speech, historical, attitudinal, and behavioral scores, family history.	ANNs	92%
2000	37 speakers	Duration and frequency of dysfluent portions, speaking rate	HMMs	-
2003	6 Non Stuttered normal + stop gaps sample	Frequency, 1st to 3rd formant's frequencies and its amplitude	ANNs & Rough Set	73.25 % & ≥ 90.0%
2003	-		Hopfield network	-
2003	10 normal+10 stuttering children	Formant pattern, speecd of transitions, F2 transition duration and F2 transition	-	-
2006	8 speakers	Spectral measure (FFT 512)	MLP, Kohonen	76.67%
2006	-	Perceptral linear prediction (PLP)	HMM-SVM	-
2007	38 samples for prolongation of fricatives + 30 samples for stop blockade + free–of–silence samples	MFCC	HMMs	70%
2007	-	MFCC	HMMs	80%
2007	15 normal + 15 artificial stuttered speech	MFCC	HMMs	96%
2008	10 speakers	MFCC	Perceptron	83%
2009	15speakers	MFCC	SVM	94.35%
2009	8 stutter speakers + 4 normal speakers (59 fluent speech samples + 59 Non fluent speech samples)	Spectral measures(FFT 512)	MLP,RBF	88.1% 94.90%
2009	8 speakers	22.05kHz, Applying Filters	Kohonen networks	76%
2009	10 samples from UCLASS	MFCC	K-NN, LDA	90.91%
2009	10 samples from UCLASS	LPCC	K-NN, LDA	89.77%
2010	2 speakers	MFCC	HMMs	80%
2010	121 speakers	Time Domain, Spectral domain	Batesian detector, HMM,LDA	63%
2012	39 samples from UCLASS	MFCC	k-NN,LDA	92.55%
2012	39 samples from UCLASS	LPC,LPCC,WLPCC	k-NN	92.16% 96.47% 97.45%
		LPC,LPCC,WLPCC	LDA	94.90% 97.06% 98.04%
2012	53 speakers	Spectral measures(FFT 512)	ANNs	84%
2013	39 samples from UCLASS	LPCC,WLPCC,MFCC	SVM	95%
2014	16 samples from UCLASS	MFCC	SVM	98.00%
2014	10 speakers	volume, zero crossing rate, spectral entropy, high-order derivatives, VH curve, and VE curve and end-point detection according (EPD)	DTW	83%
2015	50 samples from UCLASS	MFCC	GMM	96.43%

by different way on stuttered speech, for classification of fluent and dysfluent in stuttered speech and recognition of prolongation and repetition in stuttered speech.

In[45]first to research and develop a new area of knowledge to find the particular stuttered events are repetition and prolongations by using ANNs as classification technique. The proposed recognition technique was achieved by each word is judged as fluent, repetitions, prolongations or other dysfluency categories[44,45]. They provide a combination of the autocorrelation function and spectral information and Envelope parameters as input vector to ANNs. Around approximately 80% best accuracy were obtained by them.

In[25,46] they carried out their research up to next advanced level. The 12 students who were stuttered employed for this study and speech samples were obtained by reading 376-word passage "Arthur the rat". Speech samples was recorded on DAT tape and down sampled to 20 kHz and transferred to computer for further processing. They differentiate fluent and dysfluent speech by using part work duration, fragmentation measure, energy and spectral measures. The ANNs gave 78.07% correct identified accuracy of both prolongation and repetition dysfluent word. In[11] focuses their research on distinction between normal non-fluency (NNF) and stuttering in children by using ANN. 25 dysfluent children were used to train the ANN and 26 dysfluent children were used for ANN to predict the classification. They used 10 different variables to make the differentiation between NNF and stuttering. These variables are age, sex, type of dysfluency, frequency of dysfluency, duration, physical concomitant, rate of speech, historical, attitudinal, and behavioural scores, family history. ANNs predict the classifications between normal non-fluency and stuttering with 92% accuracy by using these variables. In[54] used both ANNs and rough set to automatic recognition of stuttering episode based on the stop-gaps, discerning vowel prolongations, detection of syllable repetitions. They employed 6 fluent speech sample and 6 with stop-gaps speech sample. Results implied that the better accuracy score was obtain from rough set-based system was more than 90% than the ANNs average accuracy equal to 73.25%. Other researcher[51] depict the neural network tests on ability of recognition and categorizing the non-fluent and fluent speech record. They have used Sound Blaster for recording speech samples of 8 stuttering people. Speech Samples were analyzed by FFT 512 with the use of a 21 digital 1/3-octave filters of centre frequencies between 100 Hz and 10 kHz. The authors obtained 76.67% best classification result with the best network, built of 171 input neurons, 53 neurons in hidden layer and 1 output neuron.

In[23] introduced novel automatic detection method for syllable repetition in read speech for objective assessment of different types of stuttered dysfluencies. This detection method has four stages comprising of segmentation, feature extraction, score matching and decision logic. They prepared 150 words Standard English language passage database. Sample collected in database were manually segmented. They employed 12MFCC for feature extraction and for recognition system they used neural networks (Perception), this Perception was the first iterative algorithm for learning linear classification to make decision whether a syllable is repeated or not. Authors were collect 10 speech samples and achieved 83% accuracy, out of those 8 samples are for Perceptions classifier training while remaining 2 samples are for testing.

In[55] carried their research work on automatic detection of dysfluency in stuttered speech. They were collected 8 stuttered people age range from 10 to 23 years. 59 non- fluent and 59 fluent speech samples selected from recording for this concerned analysis. They used 21 digital 1/3-octave filters of centre frequencies between 100Hz and 10 kHz to analyze the speech samples and these parameters work as input to the networks. Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks applied to recognize and classify fluent and non-fluent in speech samples. These networks give classification correctness for all networks ranged between 88.1% and 94.9%. In[15] they proposed a system to identification of the block, syllable repetitions, syllable-initial prolongation types of dysfluency from the continuous speech on the basic of hierarchical ANN structures. They employed 19 stuttered and 14 fluent people. 153 non-fluent and 153 fluent speech samples were selected from recording. They applied FFT512 and 21 digital 1/3-octave filters of centre frequencies between 100 Hz and 10 kHz for speech sample analysis. They yield correct classification ranged between 84% to 100% depending on the dysfluency type.

4.2 HMMs

HMM stand for Hidden Markov Model is a stochastic model that calculates the statistical real world data. HMMs are specially widely used in stuttered speech recognition and dysfluency recognition such as prolongation and repetition because speech signals are treated as piecewise or a short-time stationary signals[56–61].

Researcher[32] present an innovative combined work of speech recognition system and SLP to evaluate the degree of stuttering during therapy session. This System can use statistical analysis approach for counting and classification of typical repetitions, pauses and phoneme duration. The measurable factors used to classify the degree of stuttering are frequency of dysfluent portions in the speech, duration of the dysfluencies and speaking rate. The database consists of 16 non stuttered and 37 patients with stuttering symptoms in that either reading all or the beginning of a passage. Results of word and phoneme accuracies of the stuttered text in relation to the number of detected dysfluencies showed a correlation coefficient of up to 0.99. In[61] needs more experiments in the future, especially with stutterers belonging with repetition or blocking type.

In 2007[62] have presented 2 papers about an automatic detection system utilizes HMM as the classification technique. They collect 30 samples for stops blockade of recognition model, 30 samples for summary model and 38 samples for prolongation of fricatives recognition model. The samples were down-sampled to 22050 Hz and the samples were parameterized using MFCCs. In their first paper the best recognition accuracy was achieved for free silence equal to 70%. In their next paper 2010, the sampling frequency of the sound samples were 22050 Hz and all the records were normalized to the same dynamic range – 50dB. They got a best result of approximately 80%. In[10]they utilizes HMM for Malay Speech Therapy Assistance Tools (MSTAT) is a system which helps SLP to not only diagnose children and to train with stutter but also keep track their clients. The normal and disordered speech signals are used to train HMM model. This project database consists of 20 normal speech samples and 15 samples of artificial stutter speech data. 10 samples of each normal and artificial stutter speech are used to train speech models and remaining 5 sample of each speech data used to test on HMM models. They set the threshold value, if the score is less than the threshold then it is diagnosed as stutter. The average percentage of recognition rate for normal speech is 96% and for the artificial stutter speech is 90%.

In[63] find method to automatically and objectively determine the degree of speech fluency disorders on the basis of analysis of audio recordings stutterers. This database consists of 121 speakers speech signals. Dysfluency was analysed on number of parameters like the total length of silence and speech, the number of segments of silence/speech, the periodicity of speech signal energy, parameter voice/unvoiced speech and spectral changes. This system correctly identify the degree of dysfluency is 63% in individual.

4.3 SVM

Support Vector Machine (SVM) is widely used powerful machine learning tool in the field of pattern recognition to attempt good separating hyper-plane between two classes in the higher dimensional space[17]. In 2009[26] same as their previous work, they proposed automatic detection method for syllable repetitions in read speech. This assessment done in four stages segmentation comparison, feature extraction, score matching and decision logic. As compare to their previous work SVM was used to differentiate between normal and fluent speech. The 15 dysfluent speech samples were recorded from 15 patients out of that 12 samples are used for training and remaining 2 samples used to testing. The system secures the 94.35% accuracy which is better than previous work.

In[64] present comparison of effectiveness of classifier for classifying speech dysfluencies. In this work speech database is carried from UCLASS database. There are 43 different speakers 38 male and 5 female having 107 recordings were selected for experiments. Speech samples were down-sampled to 16000Hz. They used 10-fold cross-validation scheme to show the reliability of classification result. These three k-NN, LDA and SVM were yield 95% accuracy. In 2014[65]Palfy, present algorithms of speech transformation to discrete symbolic sequence for searching complex repeated pattern in speech using SVM classifier. The lack of uniformity of speech sample selection from UCLASS database and speech segmentation for experiments, due to this their results are not easily comparable. The system yields 98% accuracy.

4.4 GMM and DTW

GMM (Gaussian Mixture Model) as statistical model to evaluate and differentiate the category of dysfluencies. GMM is parameterized by the mean vectors, covariance matrices and mixture weights. Researcher in[66], proposes a method for identify repetitions in stuttered speech. This proposed methodology was verified on Mandarin Chinese speech sample. They Collect 10 male stuttered people of Mandarin Language. DTW (dynamic time warping) used as classifier. The experiment results of DTW had accuracy of 83%.In recent[67], UCLASS speech samples are employed to carry the task. The short-time window analysis was used for feature extraction. The classification accuracy is 94.43%.

4.5 k-NN and LDA

In[60,61] publish two articles to analyze the effectiveness of k-NN and LDA for classification of repetition and prolongation by using two feature extraction methods MFCC and LPCC. They used 43 different speakers 10 recording samples obtained from UCLASS and each sample were down sampled to 16k HZ. In LDA, a 10^{th} fold cross-validation was applied on MFCC with 90.91% accurate classification. In[61] autocorrelation analysis used to convert LPC parameter into LPC cepstral coefficient. Feature set of were divided into the basic of 60:40 ratio for training and testing respectively. The experiment was 10 times repeated for each k-values. They yield 89.77% accuracy for k-NN and 87.50 % for LDA.

Researcher in[68] compared two speech parameterized techniques MFCC and LPCC to recognition of stuttered events namely repetition and prolongation. In this work 39 samples were employed from UCLASS archives and down sampled to 16 kHz. According to window overlapping percentage, frame length selection and α value in the first order high pass filter LPCC is slightly better than MFCC. 25 MFCC features gave 92.55% best accuracy whereas 21 LPCC features present the best accuracy of 94.51%.

5 Discussion

It is observed in the several researches, pathology treatment, case studies that there are several speech and no speech characteristics are carry by the person who stuttered (PWS) which affects the fluency[70,71,73].

Gender and Age

As mention earlier age and gender ratio are the important factor to differentiate and diagnosis between fluency and stuttering. A review of existing disfluent speech databases literature serves that male affection ratio is more than the female subjects.

Duration of Disfluency

Several researches comparing the speech disfluencies of young early stutterers, adults and their non-stuttering peers has typically focused on the frequency and type of disfluencies. A complexity and length on stuttered speech has been observed in a several studies. Literature survey reveals the overall duration of the sound/syllable repetitions and sound prolongations from minimum 2-4 seconds to 20 minutes. Whereas length and complexity have been found to affect fluency, the level of influence is unclear and disfluency is appearing to vary individually.

Rate of Speech

In past decades the rate of speech is subjectively observed in slow, average and fast utterance. In general, as compare the rate of speech of children is faster than adult speech rate. The variation is observed in different communication context to its relationship with rhythm and speech rate.

Severity

Measuring stuttering is very difficult. It was observed that in the large extent the stuttering severity (disfluency percentage) and symptomatology variably changed because there are several factor affects the severity like children stutter in different ways than adults, some stutterers can substitute words and appear to never stutter. The Syllables Per Minute (SPM) and Percent Disfluency (PD) were calculated using following formulae:

$$\mathrm{SPM} = \frac{\text{Total number of syllables read}}{\text{Total time in seconds}} \times 60$$

$$\mathrm{PD} = \frac{\text{Total number of disfluent syllable}}{\text{Total number of syllable}} \times 60$$

Physical Concomitants

Other than speech characteristics there are several non-speech characteristics are present in the person who stuttered (PWS), such as eye blinking, nose flaring, grimaces, or frowning stamping, abnormal lip, hand, and head movements. Some stuttering children also demonstrated certain avoidance behaviours such as poor eye contact, looking away, and low volume of voice associated with their stuttering, but these secondary non speech characteristics does not associate with the degree of severity[69–73].

6. Conclusion

In this paper we have discussed some of the speech database developed in different languages for stuttered speech.

We observed that the developed databases are used for recognition and analysis purpose. This paper also puts on some light on the different approaches followed by the researchers around the world for the analysis and recognition of stuttering. The work carried out for stuttered speech recognition or analysis is very less.

This review paper will help to identify the attempts taken by the researchers in the field of stuttered speech recognition and analysis systems. Now speech recognition systems are getting more importance and the use of those application is increasing however; there is a need for development of robust speech recognition which can also analysis and recognize the stuttered speech. The researcher's needs to develop a robust speech based system which can also handle the speech dysfluencies like stuttering and general speech. This paper is an attempt to highlight the work carried for stuttered speech databases and the approaches used for analysis and development of recognition system for stuttering.

This work can further be extended for multilingual societies like where there are many different languages. During the study it was observed that there is only one speech database for stuttered speech in only one Indian language (i.e., Kannada). There is need of some similar work in other Indian Languages also. This paper can help other researchers to identify the work carried out till date for stuttered speech recognition and analysis.

7. Acknowledgement

This work is supported by University Grants Commission under the scheme Major Research Project entitled as "Development of Database and Automatic Recognition System for Continuous Marathi Spoken Language for agriculture purpose in Marathwada Region". The authors would also like to thank the Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing the infrastructure to carry out the research.

8. References

1. Awad S. The application of digital speech processing to stuttering therapy. Proceedings of Instrumentation and Measurement Technology Conference Sensing Processing Networking; 1997. p. 1361–7. Crossref
2. Borsel JV, Achten E, Santens P, Lahorte P, VoetT. FMRI of developmental stuttering a pilot study. Journal of Brain and language. 2003; 85(3):369–76.
3. Bloodstein O. A handbook on stuttering. Chicago the National Easter Seal Society; 1987.
4. Waghmare VB, Deshmukh RR, Shrishrimal PP, Janvale GB. Emotion recognition system from artificial marathi speech using mfcc and ldatechniques. Proceeding of 5th International Conference on Advances in Communication Network and Computing–CNC; 2014. p. 1–9. PMCid:PMC4090972
5. Landge MB, Deshmukh RR, Shrishrimal PP. Analysis of variations in speech in different age groups using prosody technique. International Journal of Computer Applications. 2015; 126(1):1–4.
6. Guntupalli VK, Kalinowski J, Saltuklaroglu T. The need for self-report data in the assessment of stuttering therapy efficacy repetitions and prolongations of speech. The stuttering syndrome. International Journal of Language and Communication Disorders. 2006; 41(1):1–18. PMid:16272000. Crossref
7. Fisher SE, Khadem FV, WatkinsKE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder. Nature Genetics. 1998; 18:168–70. PMid:9462748. Crossref
8. American speech-language-hearing association. Scope of Practice in Speech-Language Pathology; 2007.
9. Skljarov O, Bortnik T. Chaos and speech rhythm. Proceedings of IEEE International Joint Conference on Neural Networks, IJCNN'05; 2007. p. 2070–5.
10. Tan TS, Ariff AK, Ting CM, Salleh SH. Application of malay speech technology in malay speech therapy assistance tools. Proceeding of International Conference on Intelligent and Advanced Systems IEEE-ICIAS; 2007. p. 330–4. Crossref
11. Geetha YV, Pratibha K, Ashok R, Ravindra SK. Classification of childhood disfluencies using neural networks. Journal of Fluency Disorders. 2000; 25(2):99–117. Crossref
12. World Health Organization. Available from: Crossref
13. Dewey D. What is developmental dyspraxia. Journal of Brain and Cognition. 1995; 29(3):254–74. PMid:8838385. Crossref
14. World Health Organization. International Statistical Classification of Diseases and Related Health Problems; 2004.
15. Świetlicka I, Jóźkowiak WK, SmołkaE. Hierarchical ANN system for stuttering identification. Journal of Computer Speech and Language. 2013; 27(1):228–42. Crossref
16. Szabelska E, Kruczyńska A. Computer-based speech analysis in stutter. Applied Computer Science. 2013; 9(2):34–42.
17. Czyżewski A, Kostek B, Skarżyński H. Technika komputerowa w audiologiifoniatriiilogopedii. Akademicka Oficyna Wydawnicza Exit; 2002.
18. Chen WY, Chen SH, Lin CJ. A speech recognition method based on the sequential multi-layer perceptrons. Journal of Neural Networks. 1996; 9(4):655–69. Crossref

19. Shriberg EE. Phonetic consequences of speech disfluency. Sri International Menlo Park CA; 1999.
20. Shrishrimal PP, Deshmukh RR, Waghmare VB. Development of isolated words speech database of marathi words for agriculture purpose. Asian Journal of Computer Science and Information Technology. 2013; 2(7):217–8.
21. Ward D. Sudden onset stuttering in an adult: Neurogenic and psychogenic perspectives. Journal of Neurolinguistics. 2010; 23(5):511–7. Crossref
22. Krishnan G, Tiwari S. Revisiting the acquired neurogenic stuttering in the light of developmental stuttering. Journal of Neurolinguistics. 2011; 24(3):383–96. Crossref
23. Ravikumar K, Reddy B, Rajagopal R, Nagaraj H. Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies. Proceedings of World Academy Science Engineering and Technology. 2008; 36:270–73.
24. Subramanian A, Yairi E. Identification of traits associated with stuttering. Journal of Communication Disorders. 2006; 39(3):200–16. PMid:16455103. Crossref
25. Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter. Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers. Journal of Speech, Language, and Hearing Research. 1997; 40(5):1073–84. PMid:9328878 PMCid:PMC2000472. Crossref
26. Ravikumar KM, Rajagopal R, Nagaraj HC. An approach for objective assessment of stuttered speech using MFCC features. ICGST International Journal on Digital Signal Processing DSP. 2009; 9:19–24.
27. Howell P, Sackin S, Yeung JA. Assessment procedures for locating stuttered events. Proceedings of the 2nd World Congress on Fluency Disorders; 1998.
28. Jacqueline S. Speaking easy for kids who stutter early treatment can make the words flow smoothly. Journal of Health. 1991.
29. Archibald L, De Nil LF. The relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter. Journal of Fluency Disorders. 1999; 24(1):25–42. Crossref
30. Howell P, Yeung JA, Sackin S. Exchange of stuttering from function words to content words with age. Journal of Speech Language and Hearing Research. 1999; 42(2):345–54. PMid:10229451 PMCid:PMC2013932. Crossref
31. Awad SS, Coreless MW, Merson R. Computer assisted treated for motor speech disorders.. Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference; 1999. p. 595–600. Crossref
32. Nöth E, Niemann H, Haderlein T, Decher M, Eysholdt U, Rosanowski F, Wittenberg T. Automatic stuttering recognition using hidden Markov models. Proceeding of 6th International Conference on Spoken Language Processing. 2000; 4:65–8.
33. Hollingshead K, Heeman P. Using a uniform-weight grammar to model disfluencies in stuttered read speech a pilot study. Center for Spoken Language Understanding; 2004. p. 1–22. PMid:15196814.
34. Lincoln M, Packman A, Onslow M. Altered auditory feedback and the treatment of stuttering- A review. Journal of Fluency Disorders. 2006; 31(2):71–89. PMid:16750562. Crossref
35. Voigt T, Hewage K, Alm P. Smartphone support for persons who stutter. Proceedings of the 13th International Symposium on Information Processing in Sensor Networks IEEE Press; 2014. p. 293–94.
36. Miyamoto C, Komai Y, Takiguchi T, Ariki Y, Li I. Multimodal speech recognition of a person with articulation disorders using AAM and MAF. IEEE International Workshop on Multimedia Signal Processing (MMSP); 2010. p. 517–20. Crossref
37. Ai OC, Yunus J. Overview of a computer-based stuttering therapy. Regional Postgraduate Conference on Engineering and Science; 2006. p. 207–11.
38. Jóźkowiak WK, Smołka E, Adamczyk B. Effect of acoustical visual and tactile echo on speech fluency of stutterers. Folia phoniatrica et logopaedica. 1996; 48(4):193–200. Crossref
39. Awad SS, Piechocki C. Speech therapy software on an open web platform. 10th IEEE Computer Engineering Conference (ICENCO); 2014. p. 53–6. Crossref
40. Awa SS, Corless MW, Przebienda L, Merson R. Development of a computer based speech fluency treatment aid. The Proceedings of the 4th IEEE International Conference on Electronics Circuits and Systems ICECS97; 1997. p. 245–8.
41. Blomgren M, Roy N, Callister T, Merrill RM. Intensive stuttering modification therapy a multidimensional assessment of treatment outcomes. Journal of Speech, Language, and Hearing Research. 2005; 48(3):509–23. Crossref
42. Guitar B, Peters TJ. Stuttering- An Integration of Contemporary Therapies; 1980.
43. Sidavi A, Fabus R. A review of stuttering intervention approaches for preschool-age and elementary school-age children. Contemporary Issues in Communication Science Disorders. 2010; 37:14–26.
44. Howell P. Assessment of some contemporary theories of stuttering that apply to spontaneous speech. Contemporary Issues in Communication Science and Disorders CICSD. 2004; 31:122–39. PMid:18259590 PMCid:PMC2231590.
45. Howell P, Sackin S. Automatic recognition of repetitions and prolongations in stuttered speech. Proceedings of the 1st World Congress on Fluency Disorders. 1995; 2:372–4.
46. Howell P, Sackin S, Glenn K. Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutterI. Psychometric procedures appropriate for selection of training material for

lexical dysfluency classifiers. Journal of Speech, Language, and Hearing Research. 1997; 40(5):1073–84. PMid:9328878 PMCid:PMC2000472. Crossref
47. Ramaboka M, Manamela J, Gasela N. Automatic Speech Recognition for People with Speech Disorders; 2012. p. 1–2.
48. Yeung JA, Gomez IV, Howell P. Exchange of disfluency with age from function words to content words in Spanish speakers who stutter. Journal of Speech Language and Hearing Research. 2003; 46(3):754–65. Crossref
49. Dworzynski K, Howell P. Predicting stuttering from phonetic complexity in German. Journal of Fluency Disorders. 2004; 29(2):149–73. PMid:15178130. Crossref
50. Zebrowski PM. Duration of the speech disfluencies of beginning stutterers. Journal of Speech Language and Hearing Research. 1991; 34(3):483–91. Crossref
51. Szczurowska I, Jozkowiak WK, Smolka E. The application of Kohonen and Multilayer Perceptron Networks in the speech nonfluency analysis. Archives of Acoustics. 2014; 31(4):205–10.
52. Wisniewski M, Jozkowiak WK, Smolka E, Suszynski W. Automatic detection of disorders in a continuous speech with the hidden Markov models approach. Proceeding of Computer Recognition Systems Springer Berlin Heidelberg. 2007; 2:445–53. Crossref
53. Wisniewski M, Jozkowiak WK, Smolka E, Suszynski W. Automatic detection of prolonged fricative phonemes with the hidden Markov models approach. Journal of Medical Informatics and Technologies. 2007.
54. Czyzewski A, Kaczmarek A, Kostek B. Intelligent processing of stuttered speech. Journal of Intelligent Information Systems. 2003; 21(2):143–71. Crossref
55. Szczurowska I, Jozkowiak WK, Smolka E. Speech nonfluency detection using Kohonen networks. Neural Computing and Applications. 2009; 18(7):677–87. Crossref
56. Prakash B. Acoustic measures in the speech of children with stuttering and normal non-fluency-a key to differential diagnosis. Workshop on Spoken Language Processing; 2003.
57. Johnson MH, Gunderson J, Penman A, Huang T. HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria. IEEE International Conference on Acoustics Speech and Signal Processing ICASSP 2006 Proceedings. 2006; 3:1–4. Crossref
58. Swietlicka I, Jozkowiak WK, SmolkaE. Artificial neural networks in the disabled speech analysis. Springer Berlin Heidelberg. Proceeding of Computer Recognition Systems. 2009; 3:347–54. Crossref
59. Poroshin AN, Jarov OP. An internet system of partner-learning special type. In International Conference on Physics and Control. 2003; 2:703–6. Crossref
60. Chee LS, Ai OC, Hariharan M, Yaacob S. MFCC based recognition of repetitions and prolongations in stuttered speech using k-NN and LDA. IEEE Student Conference on Research and Development (SCOReD); 2009. p. 146–9. Crossref
61. Chee LS, Ai OC, Hariharan M, Yaacob S. Automatic detection of prolongations and repetitions using LPCC. International Conference for Technical Postgraduates (TECHPOS); 2009. p. 1–4. Crossref
62. Wisniewski M, Jozkowiak WK, Smolka E, Suszynski W. Automatic detection of prolonged fricative phonemes with the hidden Markov models approach. Journal of Medical Informatics and Technologies. 2007.
63. Bergl P, Lustyk T, Cmejla R, Cerny L, Hrbkova M. Assessment of dysfluency in stuttered speech. Proceeding of Technical Computing of Bratislava; 2010. p. 1–3.
64. Fook CY, Muthusamy H, Chee LS, Yaacob SB, Adom AHB. Comparison of speech parameterization techniques for the classification of speech disfluencies. Turkish Journal of Electrical Engineering and Computer Sciences. 2013; 21(1):1983–94. Crossref
65. Pálfy J. Analysis of dysfluencies by computational intelligence. Information Sciences and Technologies. 2014; 6(2):1–14.
66. Yeh PH, Yang SL, Yang CC, Shieh MD. Automatic Recognition of repetitions in stuttered speech using end-point detection and dynamic time warping. Proceeding of Social and Behavioral Sciences. 2015; 193:356. Crossref
67. Mahesha P, Vinod DS. Gaussian mixture model based classification of stuttering dysfluencies. Journal of Intelligent Systems. 2016; 25(3):387–99. Crossref
68. Ai OC, Hariharan M, Yaacob S, Chee LS. Classification of speech dysfluencies with MFCC and LPCC features. Journal of Expert Systems with Applications. 2012; 39(2):2157–65. Crossref
69. Hariharan M, Chee LS, Ai OC, Yaacob S. Classification of speech dysfluencies using LPC based parameterization techniques. Journal of Medical Systems. 2012; 36(3):1821–30. PMid:21249515. Crossref
70. Wingate EM, Howell P. Foundations of stuttering. The Journal of the Acoustical Society of America. 2002; 112(4):1229–31. Crossref
71. Cordes AK, Ingham RJ, Frank P, Ingham JC. Time-interval analysis of interjudge and intrajudge agreement for stuttering event judgments. Journal of Speech Language and Hearing Research. 1992; 35(3):483–94. Crossref
72. Kully D, Boberg E. An investigation of interclinic agreement in the identification of fluent and stuttered syllables. Journal of Fluency Disorders. 1988; 13(5):309–18. Crossref
73. Palfy J, Pospichal J. Pattern search in dysfluent speech. Proceeding of International Workshop on Machine Learning for Signal Processing (MLSP); 2012. p. 1–6. Crossref