International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 11976-11980
© Research India Publications. http://www.ripublication.com

Identification and Validation of Repetitions/Prolongations in Stuttering Speech using Epoch Features

G. Manjula
Research Scholar, Department of Electronics and Communication Engineering,
Associate Professor, Department of Telecommunication Engineering,
GSSS Institute of Engineering and Technology for Women,
KRS Road, Metagalli Industrial Area, Mysuru, Karnataka 570016, India.
Orcid Id: 0000-0003- 1160-6393

M. Shiva Kumar
Professor and Head, Department of Electronics and Instrumentation Engineering,
GSSS Institute of Engineering and Technology for Women,
KRS Road, Metagalli Industrial Area, Mysuru, Karnataka 570016, India.
Orcid Id: 0000-0003- 1247-6009

Y. V. Geetha
Professor Emeritus, Speech Language Pathology.
All India Institute of Speech and Hearing, Mysuru, Karnataka 570016, India.

T. Kasar
Department of Electronics and Communication Engineering,
GSSS Institute of Engineering and Technology for Women,
KRS Road, Metagalli Industrial Area, Mysuru, Karnataka 570016, India.

[1,2]Orcid: 0000-0003-1160-6393, 0000-0003-1247-6009

Abstract

Stuttering is an involuntary disturbance in the fluent flow of speech characterized by disfluencies such as sound/syllable repetition, prolongation, or blocks. There are high proportion of prolongations and repetitions in stuttering. Automatic recognition of features in stuttering is difficult and has been a diagnostic challenge to the speech/language pathologists (SLP) for long. This current work is intended for automatic recognition of instances such as prolongations and repetitions in stuttering speech using epoch features by monitoring the speech at glottal closure instants. In this paper, we propose a method for detecting glottal closure instants in the stuttering speech based on the harmonics of the phase of the source signal. The source signal is extracted by eliminating the effect of vocal tract resonances from the stuttering speech signal by zero frequency filter (ZFF). The proposed zero frequency filter brings out the region of glottal activity during excitation. Similarly, the normalized error helps to distinguish between regions of stuttering events and normal speech. The current work was carried out at All India Institute of Speech and Hearing (AIISH), Mysuru on the speech data of 20 adults with stuttering. Experimental investigation helped us in the identification of stuttering events which were validated for the subjective identifications of stuttering by the speech language pathologists.

Keywords: Stuttering, Epochs, Zero Frequency Filter

INTRODUCTION

Speech is a complex neuromotor activity controlled by the nervous system. Speech attributes such as nasal voicing, dental, describe the way speech is articulated. The peripheral systems involved in the production of speech are respiration, phonation and articulation. Speech is the most frequently used form of communication among the human beings. Not all human beings are blessed with normal speech. Stuttering is one of the most common speech problems associated with speech pathology. Stuttering problem affects more male members compared to females a ratio of 1:3[1].

Stuttering is a speech disorder characterized by certain types of speech disfluencies, such as Sound/syllable repetitions, prolonged sounds, and Dysrhythmic phonations or articulatory fixations [2]. With proper speech therapy techniques, the persons with stuttering (PWS) can be trained to shape their stuttering speech into fluent speech.

The common method followed by the Speech language pathologist (SLP) for the assessment of stuttering is by counting and classifying the occurrences of disfluencies such as repetitions, prolongations, and articulatory fixations. The traditional method is subjective, inconsistent, time consuming and prone to error [4].Hence, automatic stuttering recognition methods are used to automate the measurement of Disfluency count and identify the type of Disfluency which helps in providing an objective and consistent assessment of stuttering.

This research work focuses on automatic detection of prolongation and repetition events in stuttering. Objective assessment of stuttering can aid the SLPs in their diagnostic procedures. In the recent years, many efforts have been made in the objective assessment of speech disorders. Epoch is the instant of significant excitation of the vocal tract within a pitch period. Epochs are due to the glottal closure instants (GCIs) of the glottal cycles. Most epoch extraction methods rely on the error signal derived from the speech waveform after removing the predictable portion (second-order correlations). Epochs yield important information for the analysis of speech. This paper focuses on extraction of epoch features of stop consonants from zero-frequency filtered signal. The proposed zero frequency filtered signal elevates the region of glottal activity during excitation. The classification results obtained from the epoch were successful in identifying the instances of repetitions and prolongations in stuttering.

This research paper is organized as follows: In Section III, the methodology of the system is presented. Furthermore, this section covers database that was used in the experiment, feature extraction algorithm and classification technique. Experimental results and discussion are presented in Section IV. Finally conclusions and future works are discussed in Section V.

REVIEW OF LITERATURE

The literature survey on the analysis of stuttering speech indicates that there are variations in the production of stuttering speech with respect to normal speech [9]. This analysis of stuttering speech helps in identifying the best analysis suitable for the automatic identification of instances of stuttering. The analysis of speech can be based on spectral features or non spectral features. The most commonly used classifiers and feature extraction techniques in spectral analysis to identify the types of disfluencies and also to distinguish between stutterers and nonstutters are Artificial Neural Networks (ANN), Hidden Markov Model (HMM), Support Vector Machine (SVM), Mel Frequency Cepstral Co efficient (MFCC), Perceptual linear prediction, Dynamic Time Warpping (DTW).

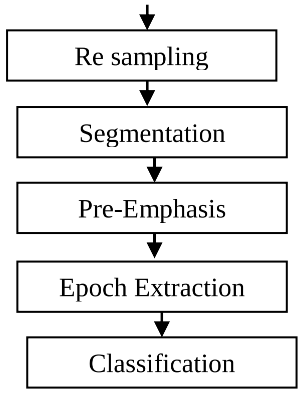
The voice onset time (VOT) is defined as the difference between the time of burst release and the time of phonation. Normal speech is a result of efficient coordination of oral-facial muscles and the vibration of vocal folds, people who stutter (PWS) often lack this coordination. This can be observed in acoustic analysis, and through further examination of VOT shed further details on articulatory instability in stutters. A study showed that non stutters produce a longer VOT for voiceless stops, while stutters present a longer VOT for both voiced and voiceless stops when compared to non stutters. The non spectral analysis of speech signal provides information about short time energy, magnitude, zero crossing count (ZCC), and VOT. The speech of adults with stuttering is characterized by longer vowel and phrase durations and increased proportion of voicing durations during the closure for stop consonants [12]. The increased proportions of voicing segments in stuttering, decreased durations of silence during intervocalic durations for voiced stops and longer VOT [13].

Most epoch extraction methods rely on the error signal derived from the speech waveform after removing the predictable portion (second-order correlations). The error signal is usually derived by performing linear prediction (LP) analysis of the speech signal. The first contribution to the detection of epochs using the determinant of the auto variance matrix was examined by Sobakin et al [14], It was shown that the determinant of the autocovariance matrix of the speech signal was maximum at the beginning of the interval. Strube [15] proposed a slightly modified version for the determination of epochs. In his work, predictor methods were based on LP analysis for the determination of the epochs. These methods do not always yield reliable results because the LP residual contains peaks of random polarity around epochs. Cheng et al [16] proposed maximum-likelihood epoch detection (MLED) method in which the strongest positive pulse in the MLED signal indicates the epoch location within a pitch period. However, the MLED signal creates not only a strong and sharp epoch pulse, but also a set of weaker pulses which represent the suboptimal epoch candidates within a pitch period. The minimum phase signals and group delay function was proposed for the extraction of epochs [17, 18, 19]. Brookes et al [20] proposed average group-delay, zero frequency weighted phase. All the four measures of group delays were used for the detection of epochs. It is a challenging task to detect epochs for voiced consonants group-delay, energy-weighted group-delay, and energy.

From the literature survey of stuttering speech analysis and recognition, it is observed that there are certain deviations in stuttering speech when compared to normal speech. Stuttering speech analysis and recognition can be done by analyzing spectral features or temporal features. In case of spectral analysis of stuttering speech Mel Frequency Cepstral Coefficient (MFCC) feature extraction technique with Hidden Markov Model (HMM) as classifier is proven to give highest accuracy 96% [32]. Stuttering speech is due to variation in speech processing ability with temporal stability. Voice on set time is a significant temporal feature for analysis and classification of stuttering speech [11].

METHODOLOGY

In the current study non spectral method was used for the identification of the stuttering events. The non spectral methods will enhance the information of the excitation source of the vocal tract. The primary and most significant mode of excitation is due to the activity of glottis. Figure 1 shows block diagram to extract the low frequency information and the breathy noise part of the glottal excitation in stuttering speech.



Stuttering Speech Data Base

Figure 1: Block diagram of stuttering signal processing

Speech Data

Twenty participants with stuttering (PWS) in the age group of 15 to 35 years were considered for the study which comprised the clinical group. The clinical group included 16 males and 4 females with stuttering, they were diagnosed as having stuttering by qualified speech-language pathologists. All the participants were instructed to read a rainbow passage and reading was recorded in a recording room at AIISH using a PRAAT tool and higher end camera with 44KHz sampling.

Speech signal pre-processing

The stuttering speech signals were sampled at 44.1 kHz since most of the salient features for speech processing are within 8 kHz bandwidth. By using decimation method, the signals were down sampled to 8 kHz by a factor of n where n = 44.1 KHz/8KHz.

Segmentation

In this work, two types of disfluencies in stuttering speech such as repetitions and prolongations were considered. Data was transcribed and segmented using Praat tool for acoustic analysis. The consonant stops /pa/, /ta/, /ka/, and /da/ were identified and extracted. The segmented samples were subjected to further processing of speech signal.

Pre-Emphasis

Stuttering speech signal had smaller magnitude of energy in the low frequency range, and hence processing of the signal to emphasize higher frequency energy is required. Pre-emphasis was performed on the digitized stuttered speech signal to flatten the magnitude spectrum and to balance the high and low frequency components. The output of the pre-emphasizer network was related to the input network by the equation.

$$Y[n] = X[n]-0.95\ X[n-1]\ldots.. \qquad (1)$$

The purpose of pre-emphasis stage is to boost the amount of energy in the high frequencies.

Epoch Extraction

Events occur at several levels of speech production and it is important to identify events before further processing of the signal. In this work, the instances at the production level are considered, that is, the instances occurring due to major source of glottal vibrations (epochs). Fig 2 indicates the speech signal, filtered output, and the normalized error for stuttered BA and Normal BA.

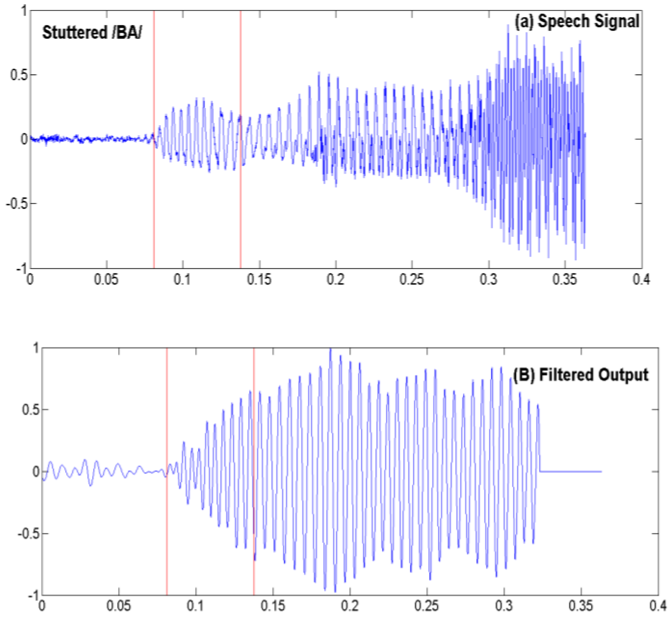
The following steps are involved in processing the stuttering speech signal to derive the glottal excitation instances.

A) The difference of the stuttering speech signal S[n] is taken to eliminate any slowly varying component introduced by the recording device i.e.,

B) Pass the signal x[n] twice through an ideal resonator at zero frequency. That is,

$$y_1\,[n] = -\sum_{k=1}^{2} a_k\, y1[n-k] + x[n] \qquad (2)$$

$$y_2\,[n] = -\sum_{k=1}^{2} a_k\, y2[n-k] + x[n] \qquad (3)$$



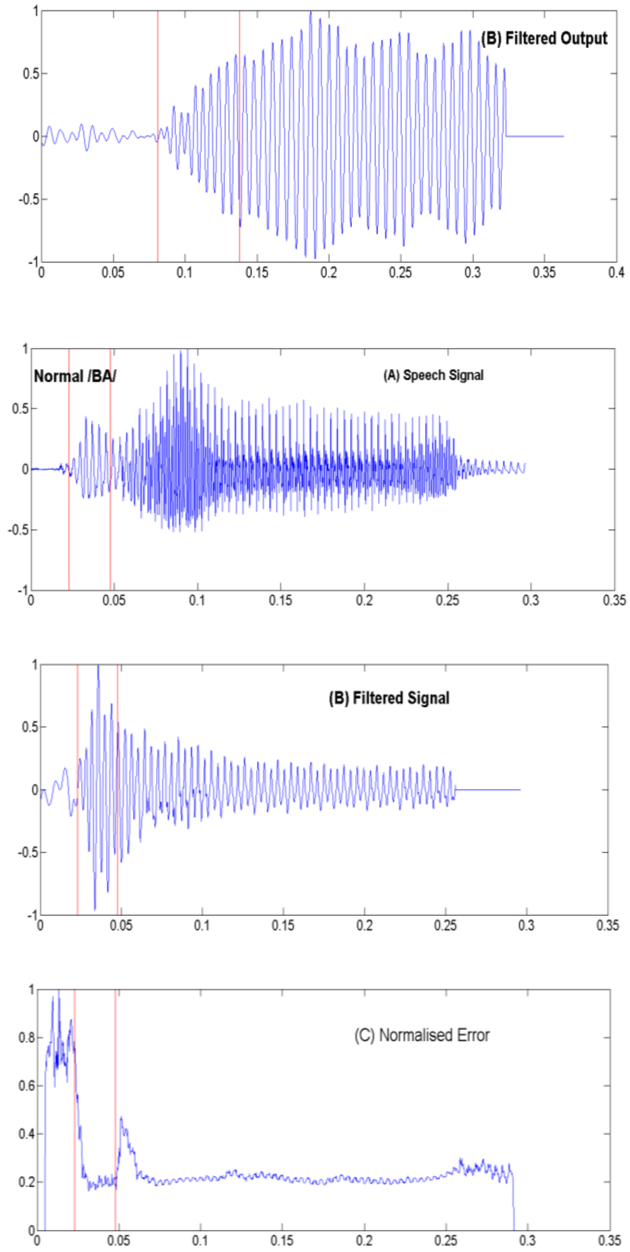


Figure 2: The speech signal, filtered output, and the normalized error for stuttered BA and Normal BA.

C) Remove the trend in $y_2[n]$ by subtracting the average of $y_2[n]$ over 20 ms at each sample. The resulting signal $y[n]$ is called the zero-frequency filtered stuttering speech signal.

D) As the LP spectrum provides the vocal tract characteristics, the vocal tract resonances (formants) can be estimated from the LP spectrum. The LP residual is computed as follows:

$$e[n]=x[n]-\sum_{k=1}^{p} a_k x[n-k] \qquad (4)$$

where a_k's are the LPCs obtained by solving the autocorrelation equations.

RESULTS

The VOT obtained manually from the zero frequency filter and normalized error are represented by the notations for the burst release (B) and onset of glottal activity as (V) in Fig 2. The interval between these two activities is VOT. The Table 1 shows the comparative results of VOT detected by using zero frequency filters of six standard syllables /BA/, /DA/, /GA/, /PA/, and /TA/ both for normal and stuttering speech for categories of CV units ending with the vowel /A/. The proposed method consistently performs better even under degradation of speech due to stuttering.

Table 1: VOT analysis of Stop Consonants in Stuttering

Stop Consonants	VOT for Stuttering Speech in ms	VOT for Normal Speech in ms
/BA/	58	26.2
/DA/	47.2	13
/GA/	20.5	15.2
/PA/	20	12
/TA/	25	23

The improved performance of the proposed method may be attributed to the following reasons. 1) The entire speech signal is processed at once to obtain the filtered signal. 2) The proposed method is not dependent on the energy/magnitude of the signal. 3) There is only one parameter involved in the proposed method; the length of the window for removing the trend from the output of 0-Hz resonator. 4) There are no critical thresholds involved in identifying the epoch locations.

DISCUSSION AND CONCLUSION

The Principal finding of this study, is that there is a durational increase of temporal measures in production of stop consonants (BA/, /DA/, /GA/, /PA/, and /TA/) in stuttering speech. This paper proposed a source based voice/non-voice detection method by extracting the harmonics of the amplitude spectrum phase of Zero Frequency Filter (ZFF) speech. The Zero Frequency Filter is utilized to detect excitations of stuttering events created by the glottal activity. The VOT of Stuttering speech was successfully extracted for the stop consonants and was compared with the VOT of normal stop consonants. The proposed is employed only for stop consonants of stuttering speech and normal speech. The system can be further improved for extracting VOT of complete sentences and also for multi modal stuttering.

REFERENCES

[1] Andrzej Czyzewski, Andrzej Kaczmarek1 and Bozena Kostek, “Intelligent processing of stuttered speech”, Journal of Intelligent Information Systems, vol.21, pp.143-171, Sep. 2003.

[2] Hariharan.M, Vijean.V, Fook.C.Y, and Yaacob.S, “Speech stuttering assessment using sample entropy and Least Square Support Vector Machine (published conference proceedings style)”, in Proc. 8th IEEE International Colloquium Signal Processing and its Applications (CSPA), Malaysia, 2012, pp. 240-245.

[3] Michael D. McClean, Stephen M. Tasko, and Charles M. Runyan, “Orofacial Movements Associated With Fluent Speech in Persons Who Stutter” Journal of Speech, Language, and Hearing Research, Vol. 47, pp 294–303, April 2004.

[4] Jiang J, Lu C, Peng D, Zhu C, and Howell . (2012, June). “Classification of Types of Stuttering Symptoms Based on Brain Activity”, PLoSONE, Vol 6 (Issue7), Available:http://doi.org/10.1371/journal.pone.0039747.

[5] Arghavan Bahadorinejad, and Farshad Almasganj, “Delayed Auditory Feedback for Speech Disorders”, in Proc International Conference on Biomedical Engineering(ICoBE), Penag, 2011,pp. 585-588.

[6] Marius Cristian, and Adrian Graur, “Developing a Logopaedic Mobile Device Using FPGA”,in Proc. in 4th International Symposium on Applied Computational Intelligence and Informatics, Romania, 2007, pp. 89-92.

[7] Walter Dosch, and Nontasak Jacchum, “Stuttering Removal- Developing Mealy and Moore Style Implementations of an Interactive Component”, in IEEEXplore International Conference on Computer Systems and Applications, 2009, pp. 302-308.

[8] Juray Palfy, and Jirt Pospichal, “ Pattern Search in Disfluent Speech”, in Proc. IEEE International Workshop on Machine Learning for Signal Processing, Spain, 2012, pp.1-6.

[9] V. Naveen Kumar , Y. Padma Sai , and C. Om Prakash, “Design and Implementation of Silent Pause Stuttered Speech Recognition System”, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering(IJAREEIE), Vol. 4, pp. 1254 -1260, Mar. 2015.

[10] Lim Sin Chee, Ooi Chia Ai, and Sazali Yaacob, “ Overview of Automatic Stuttering Recognition System” in Proc. International Conference on Man-Machine Systems (ICoMMS), MALAYSIA, Oct 2009, pp. 5B7-1-5B7-6.

[11] Arcuri CF, Osborn E, Schiefer AM, and Chiari BM, “Relationship between the stuttering severity index and speech rate”, Sao Paulo Medical Journal, Vol.121 no.2, pp. 45-50, Jan-Mar 2009.

[12] R. Wayland and A. Jongman, “ Acoustic correlates of breathy and clear vowels: The case of Khmer,” Journal of Phonetics, Vol. 31, no. 2, 2003, pp. 181-201.

[13] M. Gardon and P. Ladefoged, “Phonation types: a cross-linguistic overview,” Journal of Phonetics, Vol.29, no. 4, 2003, pp.383-406.

[14] A. N. Sobakin, “Digital computer determination of formant parameters of the vocal tract from a speech signal,” Soviet Phys.-Acoust., vol. 18, pp. 84–90, 1972.

[15] H. W. Strube, “Determination of the instant of glottal closures from the speech wave” J. Acoust. Soc. Amer., vol. 56, pp. 1625–1629, 1974.

[16] Y. M. Cheng and O’Shaughnessy, “Automatic and reliable estimation of glottal closure instant and period, IEEE Trans. Acoustic Speech Signal Process, vol. 27, no. 12, pp. 1805–1815, Dec. 1989.

[17] R. Smits and B. Yegnanarayana, “Determination of instants of significant excitation in speech using group delay function,” IEEE Trans. Speech Audio Process, vol. 3, pp. 325–333, Sep. 1995.

[18] B. Yegnanarayana and R. L. H. M. Smits, “A robust method for determining instants of major excitations in voiced speech,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process, Detroit, May 1995, pp. 776–779.

[19] P. S. Murty and B. Yegnanarayana, “Robustness of group-delay-based method for extraction of significant excitation from speech signals,” IEEE Trans. Speech Audio Process, vol. 7, no. 6, pp. 609–619, Nov. 1999.

[20] M. Brookes, P. A. Naylor, and J. Gundnason, “A quantitative assessment of group delay methods for identifying glottal closures in voiced speech,” IEEE Trans. Audio, Speech Lang. Process, vol. 14, no. 2, pp. 456–466, Mar. 2006.