THE RELATION BETWEEN SPECTRAL CHANGES DISTANCE AND PROLONGATION IN SPEECH OF STUTTERERS

T. Lustyk[1)], *P. Bergl*[1)], *R. Cmejla*[1)]

[1)]Faculty of Electrical Engineering, Czech Technical University in Prague

Abstract

The paper deals with an algorithm that could be used to identify prolongations in audio recordings of disfluent speech. The measurement utilizes the detector of abrupt spectral changes and the voice activity detector. The experiment is based on the analysis of read audio recordings of stutterers. The highest correlation coefficient with control data is 0.72. This algorithm could become a base of a parameter in an automatic and objective assessment system.

1 Introduction

Stuttering is one of a speech fluency disorders and appears in many forms and causes. Symptoms mainly occur in speech, these are sounds, syllables, words and phrases repetition, prolongation, pauses, revision, incomplete phrases and broken words [1, 2, 3]. Correct assessment of disorder severity and follow-up treatment are very difficult tasks.

Determination of the disorder severity is based on a subjective speech assessment, one currently performed by clinicians. Hence a method, that would be able objectively and automatically to determine the degree of the speech fluency disorder, would be useful. Application of such a method would be: 1) The determination of disorder severity; 2) The evaluation of treatment results; 3) The comparison of treatment approaches.

The symptoms of a stuttered speech may be found in audio recordings to determine the degree of speech disorder. Several researches have been carried out to find method for detection of prolongations. The article [4] concentrates on finding repetitions and prolongations in the speech signal. The detection of formant frequencies was applied for finding of prolongations. More complex method involving the HMM (Hidden Markov Model) is used to recognize blockades with repetition and prolongations on fricative phonemes in [5, 6] or to find repetition and prolongation in [7].

The parameters do not have to look for the symptoms of the speech disorder but they could process the signal as a whole. The group of parameters in the time and frequency domain has been described in thesis [8]. They are: the average length of silence, the ratio of the total length of silence and the parameters exploring speech signal energy and the standard deviation of distance maxima BACD (Bayesian change-point detector).

A brief view is provided of one algorithm looking for prolongation in speech signals and its results are discussed in this paper.

2 Database and evaluation

The speech signal database was created in the past few years at the Department of Phoniatrics of the 1st Faculty of Medicine at Charles University and the General Faculty Hospital in Prague. The database contains recordings of approximately 160 Czech native speakers with different age and different degree of speech fluency disorder. The read part of database consists of 121 recordings is used in the described experiment.

The read text includes about 70 words. Each utterance takes approximately 60 s. The sampling frequency was 44 kHz when recording. The signals were down-sampled to 16 kHz for following analysis.

To verify the suitability of the measurement the control data are necessarily. The control data were produced by means of the LBDL (the Lidcombe Behavioral Data Language of stuttering), see [9] for more details. The LBDL considers seven descriptors of stuttering symptoms: syllable repetition (SR), incomplete syllable repetition (ISR), multisyllable unit repetition (MSUR), fixed posture with audible airflow (FPWAA), fixed posture without audible airflow (FPWOAA), superfluous verbal behaviors (SVB), superfluous nonverbal behavior (SNB). All descriptors are detectable in speech signal except the descriptor SNB. The descriptor SNB is not used in the experiment. The *overall*, *repeated* (SR + ISR + MSUR) and *fixed* (FPWAA + FPWOAA) descriptors are also considered in this paper. All read recordings were evaluated by one evaluator.

3 Methods

The algorithm for the disfluent speech evaluation could be imagine as a black box. At the beginning there is a speech signal, which is processed by an algorithm (black box) and at the end there is a number indicating the level of speech fluency disorder.

The algorithm for finding prolongations uses two instruments for processing. The first is the voice activity detector (VAD) and the second is the detector of abrupt spectral changes.

The VAD is based on the Mel-spectrum filter bank. A brief view on its procedure: 1) Estimation of power spectra (computed by the Welch's method). 2) Application of the triangular Mel-frequency filters. 3) Decision about speech activity in each frequency band by means of adaptive thresholds. 4) Final decision about the speech activity (speech/silence).

Detection using Bayesian approach (BACD) is used to detect spectral changes in audio signals. The spectral changes should correspond to phoneme boundaries. A detail description and applications of the BACD can be seen in [10] or [11]. An example of speech signal and BACD output is shown in the Figure 1. All positions of spectral changes are indicated by the red x-mark as local maxima. Many local maxima do not correspond to significant changes at phoneme boundaries. These maxima are excluded from further analysis by applying a threshold. The analysis carried out on detector outputs from different participants showed that the threshold should not be the same for all signals in the database. In thesis [8] an original method was presented for the threshold extraction. The threshold is determined as a fragment of one selected maximum. In this case the fourth highest maximum and the multiple 0.25 was used. The significant spectral changes (maxima) are marked by black circles in Figure 1.

The algorithm of finding prolongations is then very simply. We combine the output of the VAD and the BACD to find significant spectral changes that are included in speech segments. The distances between all peaks are calculated, those that are longer than a threshold are considered to be prolongations. The number of long distances is then divided by the number of all intervals between spectral changes. We gain the relative number of long distances (prolongations).

The examined values of the threshold (the shortest distance which can be considered as a prolongation) were: 100, 200, 300, 400, 500, 600, 700, 800, 1000, 1200, 1500 and 2000 ms. Some of threshold values as shorter (100, 200, 300 ms) and longer (1500, 2000 ms) are supposed not to achieve good results. The reasons are as follows: 1) duration of phonemes is comparable to the threshold (short distances 100, 200, 300 ms) and there will be a large number of candidates; 2) prolongations of 1500 ms and longer are not frequent.

Duration of typical prolongation is not determined precisely. Each listener perceives the stuttered-like phonemes in different way [12] but the boundary is approximately at 300 and

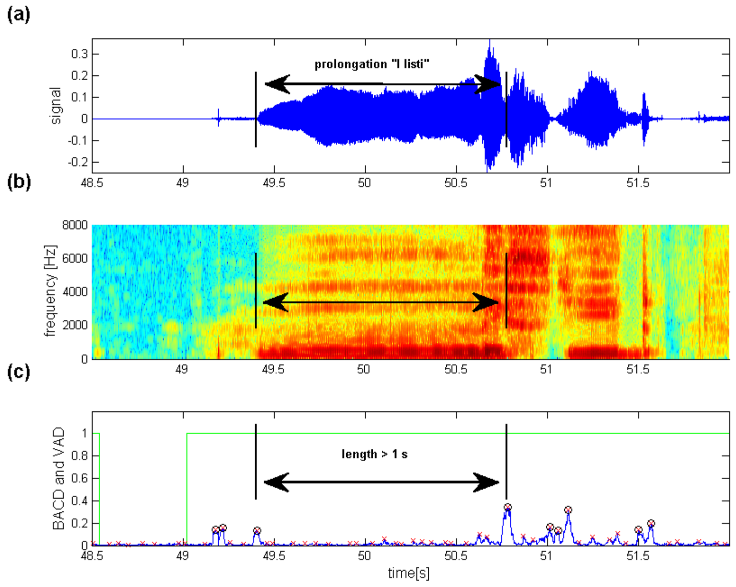

Figure 1: The prolongation in the speech signal "l listi" (in Czech), highlighted by the arrow. a) Speech signal. b) Spectrogram. c) Output of the BACD (*blue*) and the VAD (*green*), candidates of abrupt changes marked by red x-marks, selected spectral changes marked by black circles.

Table 1: The correlation coefficient (r) and the level of significance (p) for the algorithm depending on the prolongation threshold [ms] in comparison with the control data.

threshold	100	200	300	400	500	600	700	800	1000	1200	1500	2000
r	0.35	0.39	0.41	0.44	0.51	0.55	0.60	0.65	0.72	0.65	0.60	0.65
p	< 0.001											

500 ms [13, 12, 14]. There exists the possibility of false detections because there is no additional procedure in algorithm to prove that a candidate is speech or not.

The additional procedure is applied to emphasize the difference between fluent and disfluent speech. The procedure (used for VAD outputs) is based on the removal of short speech segments. The removal is successive, at first segments shorter than 125 ms are removed then segments shorter than 150 ms, the final threshold is 1700 ms. It is considered that the procedure removes short parts of speech as repetitions or broken words which are present in the stuttered speech and they will not influence further processing. The threshold 1700 ms was utilized in this experiment.

4 Results

All recordings were processed by the algorithm described above. Each signal obtained the number which should correspond to the number of prolongation. This number is compared to the control data based on the LBDL. The correlation coefficient r and the level of significance p are used to show how the algorithm is able to indicate the speech disfluency. The comparison of the algorithm of different threshold settings and control data is given in the Table 1.

The results show that the correlations rise depends on the threshold to the value of 1000 ms then they fall. The highest correlation coefficient of 0.72 was obtained for threshold 1000 ms. The thresholds of 100 and 200 ms earned the lowest correlation correlation 0.35 and 0.39, respectively.

There is also an interesting issue how the algorithm behaves in comparison with the

Table 2: The correlation coefficient (r) and the level of significance (p) for the algorithm in comparison to the control data.

descriptor	r	p
SR	0.336	$<$ 0.001
ISR	0.09	0.27
MSUR	0.40	$<$ 0.001
FPWAA	0.72	$<$ 0.001
FPWOAA	0.26	0.003
SVB	0.23	0.008
repeated	0.21	0.018
fixed	0.49	$<$ 0.001
overall	0.36	$<$ 0.001

control data. These results are shown in the Table 2. The comparison is made for the setting with the best results according to the threshold it is the threshold 1000 ms. One can see a good correlation coefficient especially for the descriptor FPWAA (prolongation). Other particular descriptors obtained lower correlation $<$ 0.40. Very important conclusion is that all repeated descriptors (SR, ISR and MSUR) have a little agreement with the algorithm. A small correlation is not necessary a negative result.

5 Conclusion

The paper deals with the measurement which can be used to assess the speech fluency disorder. The algorithm aims to finding prolongations in a speech signal. The algorithm utilizes the voice activity detector and the Bayesian change-point detector. The analyses were compared to the control data created by means of the Lidcombe bahavioral data language of stuttering. The results of comparison with control data suggest that the algorithm can be useful for evaluation of disfluent speech. The highest correlation coefficient is 0.72. The important conclusion is also that the algorithm does not have strong agreement with *repeated* symptoms.

Future work can be focused on the improvement of the algorithm ans its application for evaluation of spontaneous speech recordings or using different types of detectors.

Acknowledgement

We would like to thank Dr. Ing. Jan Vokral for providing recordings and Ing. Tereza Tykalova for evaluation of speech signals. The work has been supported by grant GACR 102/12/2230 and by the Grant Agency of the Czech Technical University in Prague, grant SGS12/185/OHK4/3T/13.

References

[1] O. Bloodstein and N. Bernstein Ratner. *A handbook on Stuttering.* Delmar, Cengage Learning, sixth edition, 2008.

[2] E. Skodova, I. Jedlicka, and colective. *Klinicka logopedie.* Portal, Prague, 2003. (in Czech).

[3] V. Lechta. *Poruchy plynulosti reci.* Scriptorium, Prague, 1999. (in Czech).

[4] P. Howell, A. Hamilton, and A. Kyriacopoulos. Automatic detection of repetitions and prolongations in stutterred speech. *Speech Input/Output: Techniques and Applications*, pages 252–256, 1986.

[5] M. Wisniewski, W. Kuniszyk-Jozkowiak, E. Smolka, and W. Suszynski. Automatic detection of disorders in a continuous speech with the hidden markov models approach. *Computer Recognition Systems 2*, vol. 45 of Advaces in Soft Computing:445–453, 2007. Springer, Berlin, Germany.

[6] M. Wisniewski, M. Niewski, W. Kuniszyk-Jozkowiak, E. Smolka, and W. Suszynski. Automatic detection of prolonged fricative phonemes with the hidden markov models approach. *Journal of Medical Informatics & Technologies*, 11:293–297, 2007.

[7] E. Noth, H. Niemann, T. Haderlein, M. Decher, U. Eysholdt, F Rosanowski, and T. Wittenberg. Automatic stuttering recognition using hidden markov models. *ICSLP-2000*, 4:65–68, 2000.

[8] P. Bergl. *Objektivizace poruch plynulosti reci.* PhD thesis, Faculty of Eletrical Engineering, CTU in Prague, 2010. (in Czech).

[9] K. Teesson, A. Packman, and M. Onslow. The lidcombe bahavioral data language of stuttering. *Journal of Speech, Language, and Hearing Research*, 46:1009–1015, 2003.

[10] R. Cmejla, J. Rusz, P. Bergl, and J. Vokral. Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders. *Speech Communication*, 2012. in press, available online 16 August 2012.

[11] R. Cmejla and P. Sovka. Audio signal segmentation using recursive bayesian change-point detectors. In *WSEAS International Conferences [CD-ROM]*, volume 1, pages 1087–1091. New York : WSEAS Press, 2004.

[12] N. Schaeffer and N. Eichorn. The effects of differential vowel prolongations on perceptions of speech naturalness. *Journal of Fluency Disorders*, 26(4):335–348, 2001.

[13] P. M. Zebrowski. Duration of sound prolongation and sound/syllable repetition in children who stutter: Preliminary observations. *Journal of Speech and Hearing Research*, 37:254–263, 1994.

[14] N. Kawai, E. C. Healey, and T. D. Carrell. The effects of duration and frequency of occurrence of voiceless fricatives on listeners' perceptions of sound prolongations. *Journal of Communication Disorders*, 45:161–172, 2012.

Tomas Lustyk
lustytom@fel.cvut.cz

Petr Bergl
berglpet@fel.cvut.cz

Roman Cmejla
cmejla@fel.cvut.cz