Change evaluation of Bayesian detector for dysfluent speech assessment

Tomáš Lustyk[1], Petr Bergl[1], Roman Čmejla[1],
Jan Vokřál[2]

[1]Faculty of Electrical Engineering Czech Technical University in Prague
lustytom@fel.cvut.cz, berglpet@fel.cvut.cz, cmejla@fel.cvut.cz
[2]Department of Phoniatrics, 1st Faculty of Medicine at Charles University and General Faculty Hospital in Prague

Abstract – **This paper is focused on a parameter which could be used for automatic and objective assessment of dysfluent speech. The measurement is based on the analysis of *read* audio recordings of stutterers. The designed parameter analyzes the signal in the spectral domain, which correlates with the subjective assessment of two phoniatrics experts with a coefficient of -0.789. A Bayesian change-point detector (BACD) has been used for spectral analysis. This parameter could become a part of an automatic and objective assessment system.**

I. INTRODUCTION

Stuttering is one of the major fluency disorders in speech, and appears in many forms and causes. Signs of stuttering are repetition (repetition of sounds, syllables, words), prolongation (unnatural extension of sounds), frequent pauses and broken words [1], [2].

Correct assessment of the disorder's severity and its follow-up treatment are very difficult tasks. Determination of severity is based on subjective speech assessment, currently performed by clinicians. Hence a method that would be able objectively and automatically to determine the degree of speech fluency disorder would be useful. Application of such a method would be: 1) Determination of disorder severity. 2) Assessment of treatment results. 3) Comparison of treatment approaches.

The symptoms of stuttered speech may be found in audio recordings to determine the degree of speech disorder. The article [3] concentrates on finding repetitions and prolongations in the speech signal. A simple VAD (Voice Activity Detector) and time thresholds were used to detect repetitions. The detection of the formant frequencies was applied for finding of prolongations. A more complex method involving HMM (Hidden Markov Model) is used for recognizing blockades with repetition and prolongations on fricative phonemes in [4].

The parameters do not have to look for the symptoms of 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 [5]. In the time domain, for example, they are: the average length of silence, the ratio of the total length of silence and speech, the number of segments of silence/speech and the parameters exploring speech signal energy. In the frequency domain for example: the number of maxima BACD (Bayesian Change-point Detector) output and the standard deviation of distance maxima BACD.

In this paper, a brief view is provided of one parameter in the frequency domain using BACD, and its results are discussed.

II. DATABASE

The speech signal database has been 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 differences in age and degree of speech fluency disorder. Utterances are *read* and *spontaneous*, recorded with and without DAF (Delayed Auditory Feedback). Thirty utterances are control recordings of healthy speakers. The range of DAF varies from 10 to 110 ms. The sampling frequency was 44 kHz when recording. The signals were down-sampled to 16 kHz for following analysis.

The *read* text consists of about 70 words. Each utterance takes approximately 60 s. Spontaneous utterances are formed as an image description. The *read* part of the database was used in all experiments described in this paper.

Subjective assessment of stuttering and dysfluency was conducted in 2008 by two clinicians at the Phoniatrics Clinic in Prague. The degree of speech fluency disorder has been described on a five-level scale, from 0 – no occurrence of dysfluent speech to 4 – very severe degree of speech fluency disorder. This data serves as control data for all of our analysis.

III. BAYESIAN CHANGE-POINT DETECTOR

Detection using the Bayesian approach could be used in many various roles, in our case as a detector

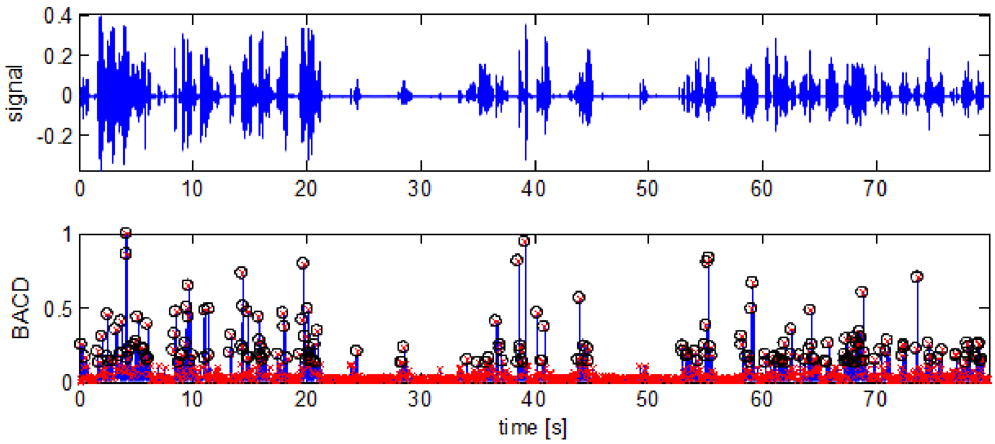

Figure 1: An example of speech signal (up) and BACD output (down).

of spectral changes. Spectral changes corresponding to the phoneme boundaries are found in audio signals. See Ref. [6] for more details.

An example of speech signal and BACD output is shown in Figure 1. All positions of spectral changes are indicated by the red x-marks as local maxima. Many local maxima do not correspond to significant changes as phoneme boundaries. These maxima are excluded from further analysis by applying a threshold.

The analysis, carried out on the detector outputs from different participants, showed that the threshold should not be the same for all signals in the database. In thesis [5] an original method was presented for threshold extraction. The threshold is determined as a fragment of one selected maxima. The adaptive approach was studied in thesis [7], but experiments did not have as good results as the fixed threshold.

The significant spectral changes (maxima) are marked by black circles in Figure 1.

Other methods, such as GLR (General Likelihood Ratio) and cepstral distance, or analysis by HTK (the Hidden Markov Model Toolkit), could be used for finding spectral changes. The following analysis will be identical.

IV. THE AVERAGE NUMBER OF BACD CHANGES IN A SHORT SEGMENT

In Section I, it has been mentioned that dysfluent speech consists of many prolongations, frequent pauses and broken words. Audio recordings were analyzed using a parameter that takes into account the presence of these symptoms.

Prolongation is the unnatural extension of sounds, in which the signal spectrum does not change in the course of its duration. Long segments of silence behave similarly. Speech activity is also interrupted by intervals of silence with various lengths, a phenomenon visible in Figure 1. All these signs of dysfluent speech could be detected by the detector of spectral changes.

The difference in the number of spectral changes could be significant for segments with speech activity opposite segments with silence (segments without spectral changes) if the output of the detector was processed in short segments.

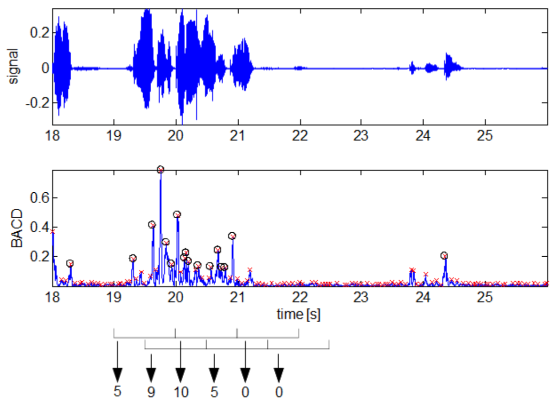

Figure 2: Procedure of calculating of *the average number of BACD changes in a short interval*. *Above:* Speech signal. *Below:* Output of Bayesian detector.

The procedure of analysis using the parameter *the average number of BACD changes in short interval* as follows: 1) all maxima are found in the detector output, 2) maxima less than the threshold are filtered, 3) the output of detector is processed in short segments.

Significant maxima are separated from others by threshold. The value of the threshold is derived as a fraction of the selected maximum (for example one third of the second highest).

Evaluation of *the average number of BACD changes a short interval* is simple: the output is processed in short segments with half-overlap. The segment length was 1, 2 and 4 s. The number of spectral changes is found in each segment and the average is computed from these data, in a process shown in Figure 2.

It is expected for participants with speech fluency disorder that *the average number of changes in a short segment* is smaller. More silence appears in

Table 1: Correlation coefficient between *the average number of BACD changes in a short interval* and the subjective assessment of second clinician, length of segment is 1 s. Value of *k* defines which maxima is used for threshold determining, M_k is the value of maxima.

k	$0.1 \cdot M_k$	$0.15 \cdot M_k$	$0.2 \cdot M_k$	$0.25 \cdot M_k$	$0.3 \cdot M_k$
1	-0.766	-0.751	-0.701	-0.646	-0.576
2	**-0.771**	-0.768	-0.731	-0.681	-0.624
3	-0.767	**-0.774**	-0.743	-0.697	-0.647
4	**-0.774**	**-0.783**	-0.752	-0.71	-0.664
5	**-0.771**	**-0.781**	-0.759	-0.721	-0.688
6	**-0.77**	**-0.786**	-0.768	-0.736	-0.703
7	-0.765	**-0.789**	**-0.775**	-0.746	-0.711
8	-0.764	**-0.787**	**-0.779**	-0.756	-0.72
9	-0.762	**-0.786**	**-0.783**	-0.763	-0.731
10	-0.759	**-0.786**	**-0.782**	-0.765	-0.736

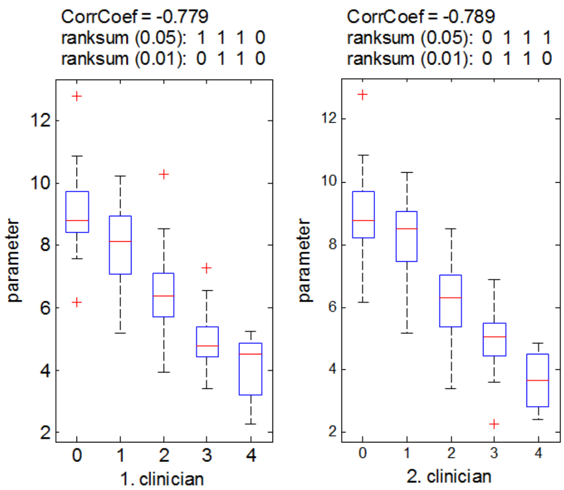

Figure 3: *The average number of BACD changes in a short interval* for threshold determined by seventh maxima with strength 0.15. *Boxplot*, correlation coefficient and results of Wilcoxon rank sum test.

stuttered than in fluent speech. The number will vary and in many cases is zero. The number of changes in healthy speakers will be more stable in contrast to stutterers and the average should be higher.

V. RESULTS

The subjective assessment of phoniatrics experts has been noted in section concentrating on the database. A score from doctors serves as the control data for all experiments.

The results of the algorithm are compared in several aspects with the subjective assessment. The correlation coefficient is the first and shows the level of agreement between the results of the evaluation parameter and the assessment of clinicians. 121 signals from the *read* part of the database were used to verify the suitability of the parameter for evaluation of the speech disorder severity. The value of *the average number of spectral changes in a short interval* is calculated for each signal. These results are compared with those of clinicians.

Correlation coefficients with the second clinician assessment and various settings of the threshold are given in Table 1; the segment length during processing was 1 s. Rows specify from which maximum the threshold is derived, columns define what is considered as the maximum value. Correlation coefficients higher than -0.77 are highlighted in bold.

High values of coefficients can be seen for the multiple 0.15. The best achieved score is -0.789 for threshold specified by the seventh highest maximum and multiple 0.15. Detailed results of this setting are shown in Figure 3.

The result is displayed using the Matlab function *boxplot*. Five boxes are plotted in Figure 3, one for each level. If the parameter is indicative of the speech disorder severity, the boxes should be of different height and their size should be minimized. The correlation coefficients with the assessment of clinicians and results of the Wilcoxon rank sum test for two significance levels ($\alpha = 0.05$ and $\alpha = 0.01$) are written in the header.

The Wilcoxon test performs a rank sum test of the hypothesis that two independent samples come from distributions with equal medians. Note that “0 0 1 0” means that the hypothesis about equal medians is rejected between groups 3 and 4. Then, the parameter could be useful for differentiation between these two groups. The ideal parameter would achieve eight *ones*. The Wilcoxon rank sum test is another method for comparison of the results of the parameter with clinicians.

The parameter presented in Figure 3 achieved five of the eight *ones*. It turns out that the decision between peripheral groups (0 – 1 and 3 – 4) is a difficult task.

For clarity, the results are presented in Table 1 for only one phoniatrics expert and length of window 1 s. The comparison of the parameter with another clinician achieved similar results. Approximately the same number of settings exceeds the coefficient -0.77.

The lengths of windows 2 and 4 s were also used in addition to the above- mentioned length of segment 1 s. The results were very similar, with many settings of the threshold exceeding the coefficient -0.77. The value of -0.789 has not been improved.

VI. CONCLUSION

This paper deals with a method that could be used in automatic evaluation of dysfluent speech: parameter searches in the frequency domain of the speech signal. It also takes into account phenomena associated with stuttered speech such as prolongation and frequent pauses.

A Bayesian change-point detector is used for spectral analysis. The suitability of the parameter *the average number of BACD changes in a short segment* for use in assessing severity of the dysfluency disorder was tested on 121 *read* utterances.

Parameter results were compared with the evaluation of two clinicians. The highest correlation coefficient was achieved for one of the doctors -0.789. Several different algorithms settings were also tested, some of which exceeded the correlation coefficient of -0.77.

Results of comparing with clinicians suggest that the examined parameter could be useful for evaluating the severity of speech fluency disorders. Future work may focus on the use of other spectral change detectors and application of the designed parameter to spontaneous recordings.

ACKNOWLEDGMENT

We would like to thanks MUDr. M. Hrbková and MUDr. J. Černý for evaluation speech signals. This work was supported by the research plan "Transdisciplinary research in biomedical engineering" (No. MSM6840770012) and grant "Analysis and modelling of biological signals" GACR No. 102/08/H008, and "Speech recognition under Real-World Conditions", GACR No.102/08/0707 and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS10/180/OHK3/2T/13.

REFERENCES

[1] Škodová E., Jedlička I. and collective: *Klinická logopedie*, Prague, Portál, 2003. (in Czech)[2] Lechta V., *Poruchy plynulosti řeči*, Prague, Scriptorium, 1999. (in Czech)

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

[4] Wiśniewski M., Kuniszyk-Jóźkowiak W., Smołka E., and Suszyński W., *Automatic detection of disorders in a continuous speech with the hidden Markov models approach*, in *Computer Recognition Systems 2*, vol. 45 of *Advances in Soft Computing*, pp. 445–453, Springer, Berlin, Germany, 2007.

[5] Bergl P., *Objektivizace poruch plynulosti řeči*. Dissertation thesis, Faculty of Electrical Engineering, CTU in Prague, 2010. (in Czech)

[6] Čmejla R., Sovka P., *Audio Signal Segmentation Using Recursive Bayesian Change-point Detectors*. In Proceedings of the WSEAS International Conferences [CD-ROM]. New York : WSEAS Press, 2004, vol. 1, pp. 1087–1091.

[7] Lustyk T., *Analýza neplynulé řeči*. Diploma thesis, Faculty of Electrical Engineering, CTU in Prague, 2010. (in Czech)

[8] Bergl P., Čmejla R., Hrbková M., Černý L., *Systém pro automatické hodnocení koktavosti*. Automatizace 2010, vol. 53, no. 1-2, pp. 49–52. ISSN 0005-125X.