Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 178–189

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Bayesian changepoint detection for the automatic assessment of fluency and articulatory disorders

Roman Cmejla [a,*], Jan Rusz [a,b], Petr Bergl [a], Jan Vokral [c]

[a] Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Circuit Theory, Czech Republic
[b] Charles University in Prague, First Faculty of Medicine, Department of Neurology and Centre of Clinical Neuroscience, Czech Republic
[c] Charles University in Prague and General University Hospital in Prague, First Faculty of Medicine, Department of Phoniatrics, Czech Republic

Received 7 November 2011; received in revised form 30 April 2012; accepted 3 August 2012
Available online 16 August 2012

Abstract

The accurate changepoint detection of different signal segments is a frequent challenge in a wide range of applications. With regard to speech utterances, the changepoints are related to significant spectral changes, mostly represented by the borders between two phonemes. The main aim of this study is to design a novel Bayesian autoregressive changepoint detector (BACD) and test its feasibility in the evaluation of fluency and articulatory disorders. The originality of the proposed method consists in its normalizing of a posteriori probability using Bayesian evidence and designing a recursive algorithm for reliable practice. For further evaluation of the BACD, we used data from (a) 118 people with various severity of stuttering to assess the extent of speech disfluency using a short reading passage, and (b) 24 patients with early Parkinson's disease and 22 healthy speakers for evaluation of articulation accuracy using fast syllable repetition. Subsequently, we designed two measures for each type of disorder. While speech disfluency has been related to greater distances between spectral changes, inaccurate dysarthric articulation has instead been associated with lower spectral changes. These findings have been confirmed by statistically significant differences, which were achieved in separating several degrees of disfluency and distinguishing healthy from parkinsonian speakers. In addition, a significant correlation was found between the automatic assessment of speech fluency and the judgment of human experts. In conclusion, the method proposed provides a cost-effective, easily applicable and freely available evaluation of speech disorders, as well as other areas requiring reliable techniques for changepoint detection. In a more modest scope, BACD may be used in diagnosis of disease severity, monitoring treatment, and support for therapist evaluation.

Keywords: Changepoint detection; Speech pathology; Speech signal processing; Disfluency; Articulation disorder

1. Introduction

Communication disorders have a strongly negative effect on our daily employment opportunities and everyday social life. The cost of care, as well as the drastic limitation of employment for people with speech disabilities, has a major impact on national economies. These circumstances indicate that communication disorders are one of the main medical challenges in the 21th century (Ruben, 2000). Acoustic analysis has the potential to provide a quantitative, objective, and precise tool to help depict the presence, severity, and characteristics of speech disorders, and to help the monitoring of deterioration or improvement in speech according to the disease's progression, recovery, or treatment effects (Kent et al., 1999). Moreover, the integration of automated speech signal processing techniques may significantly contribute to better rehabilitation for such disorders, forming a useful biomarker for diagnosis and remote monitoring of the disease (Harel et al., 2004), as well as reducing the cost of care for affected persons (Ruggiero et al., 1999).

* Corresponding author. Address: Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic. Tel.: +420 224 352 236; fax: +420 233 339 805.
E-mail address: cmejla@fel.cvut.cz (R. Cmejla).

0167-6393/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2012.08.003

Among the most common manifestations of communication disorders are those related to stuttering and dysarthria. The first, stuttering, is a chronic speech disorder characterized by repeated speech movements and fixed articulatory postures, affecting the natural fluency of speech production (Conture, 2001). Developmental stuttering is a poorly understood communication disorder with a prevalence of 5–15% in preschool children, typically starting between 2 and 7 years of age (Yairi and Ambrose, 1999). Furthermore, stuttering is estimated to persist in 1% of the adult population. Second, Parkinson's disease (PD) is a chronic neurodegenerative disorder characterized by progressive loss of dopaminergic neurons, affecting 1–2% of persons over the age of 60 years (Hornykiewicz, 1998; de Rijk et al., 1997). Patients with PD typically develop alterations of speech, characterized as hypokinetic dysarthria, that have been associated with a reduced range of articulatory movements (Ackermann et al., 1997; Duffy, 2005). Although approximately 30% of affected patients themselves regard the reduced ability to communicate as one of their most difficult symptoms, only 3–4% patients receive speech therapy (Hartelius and Svensson, 1994).

Diagnostic testing of pathological utterances and severity of speech intelligibility is traditionally based on perceptual evaluations by clinical experts (Hirano, 1981; Kent et al., 1989; Robertson and Thompson, 1986; Dejonckere et al., 2001). For this purpose, several evaluation scales such as the Voice Handicap Index (Jacobson et al., 1997), Stuttering Severity Instrument (Riley, 1972), and Dysarthria Rating Scale (Darley et al., 1969a,b) have been introduced. Nevertheless, a certain extent of intra-rater variability commonly occurs among experts in perceptible speech ranking, and therefore overall assessment of the vocal impairment is traditionally based on several independent evaluators. Hence there is an urgent need for reliable, automatic, and cost-effective objective methods to track the severity of speech performances and provide feedback in voice treatment. Currently, objective analyses of pathological speech are commonly performed using computer programs. A well-known commercially available computer system introduced by Kay Elemetrics Corp. provides several measures for speech assessment in its Multi-Dimensional Voice Program (Kay Elemetrics, 2003). As an example of freely available software, PRAAT also offers various voice and speech reports (Boersma and Weenink, 2001). However, these software packages mostly require user control of the analysis procedure. Recently, a number of studies have appeared making use of innovative methods for voice and speech disorders detection on the basis of signal processing techniques (Sapir et al., 2010; Rusz et al., 2011a), speech recognition (Middag et al., 2008; Su et al., 2008), machine learning techniques (Godino-Llorente and Gomez-Vilda, 2004; Henriquez et al., 2009), acoustic modeling (Bocklet et al., 2012), as well as new automated software (Maier et al., 2009).

When considering the precise evaluation of subjects' speech performances, we have to keep in mind that the speech signal consists of a sequence of segments that represent the individual phonemes of the utterances. Hence, we can expect abnormalities in spectral changes between phonemes in pathological utterances when compared to healthy speech. From the standpoint of signal processing, the borders between these phonemes are characterized by significant changepoints. The process of changepoint detection was first introduced by (Basseville and Nikoforov, (1993)), and it includes cumulative sum control charts and likelihood tests. Changepoint analysis covers a broad range of practical application domains such as meteorology (Chu and Zhao, 2004), climatology (Reeves et al., 2007), hydrology (Wong et al., 2006), history (Western and Kleykamp, 2004), biomedicine (Prochazka et al., 2008), astronomy (Dobigeon et al., 2005), process control (Hawkins and Zamba, 2005), and telecommunications (Ureten and Serinken, 1999). Generally, the main challenge of changepoint detection lies in the accurate discovery of points between segments with different statistical properties (Cmejla and Sovka, 2001).

In this article, we introduce a novel recursive Bayesian Autoregressive Changepoint Detector (BACD) for the automated identification of signal changepoints. Subsequently, we test its reliability in the automatic evaluation of speech disorders using data from two population groups of people afflicted with stuttering and PD. Since the symptoms of both speech pathologies are quite different, we demonstrate the potential of our detector regarding (a) the separation of several degrees of speech disfluency with motivation of support for therapeutic evaluation, and (b) the articulation-based differentiation of parkinsonian patients from healthy speakers with the possibility of making an early diagnosis of the disease. In addition to these applications, the recursive BACD is a suitable technique for use in a number of other areas dealing with signal processing applications. Furthermore, its code is freely available on http://sami.fel.cvut.cz/bacd/.

The outline of this paper is as follows. In the section 'Methods', we describe the speech data and participants, detail the mathematical description of recursive BACD, design the methods of fluency and articulation disorders assessment, and explain the statistics used in this study. The section 'Results' includes the results obtained. In the section 'Discussion', we provide a discussion of our general findings. Finally, the paper is concluded with a short summary in the section 'Conclusion'.

2. Methods

2.1. Fluency disorder: people with stuttering

During the last ten years, 118 Czech native speakers (90 men and 28 women) with different severity of disfluency and stuttering were recruited. Their age ranged from 8 to 50 years (Mean = 18.1; SD = 9.9). All participants were instructed to read a standardized short text of 75 words (a paragraph from the classic Czech novel 'Grandmother'

by Bozena Nemcova). The recorded data were reported as a part of a previous thesis (Bergl, 2010). The speech samples were recorded with a close-talking microphone at 44.1 kHz sampling frequency and 16 bit resolution.

In addition to the recording, the severity of speech disfluency for each participant had been perceptually evaluated by two independent professional speech-language pathologists. A five-point rating scale comprised of severity levels 0 through 4, including 0 = normal healthy speech (without frequent signs of disfluency), 1 = mild disfluency (approximately up to 5% of disfluent words), 2 = moderate disfluency (approximately 5–20% of disfluent words), 3 = severe disfluency (approximately 20–60% of disfluent words), and 4 = very severe disfluency (more than 60% of disfluent words) was applied to rate the relative frequency of disfluent words in each recording. This scale has been adopted by speech therapists in the Czech and Slovak Republic to assess speech fluency (Lastovka et al., 1998; Lechta, 2004). To ensure the correctness of clinical evaluation, the count of disfluencies for each recording was also assessed using the developmental stuttering taxonomy (Teesson et al., 2003). The overall disfluency percentage was calculated as the total number of disfluencies divided by the total number of words and multiplied by 100. The Pearson correlation indicated very high relationships between the overall disfluency percentage and the evaluation of the first ($r = 0.92$, $p < 0.001$) and the second ($r = 0.93$, $p < 0.001$) speech therapists, as well as very high inter-analyzer reliability ($r = 0.91$, $p < 0.001$) between the evaluations of both therapists. Therefore, both of the expert judgements were averaged for further assessment. In case an ambiguous situation occurred during evaluation, for example if the first expert ranked subject performance as 2 and the second one as 3, the higher degree of speech disfluency was retained. According to this evaluation scale, 15, 24, 41, 31, and 7 participants were separated into the individual groups of 0, 1, 2, 3, and 4 respectively.

2.2. Articulatory disorder: people with PD

Data for assessment of articulation accuracy was used from the original study (Rusz et al., 2011b), in which a total of 46 Czech native participants were observed. Twenty-four of them (20 men and 4 women) were newly diagnosed with an early stage of idiopathic PD, before the start of treatment with dopaminergic medication. Their ages ranged from 34 to 83 years (Mean = 60.9; SD = 11.2), with the duration of PD symptoms ranging from 6 to 84 months (Mean = 31.3; SD = 22.3), the Hoehn & Yahr stage (disability scale comprised of stages 1–5, where 5 is most severe) ranging from 1 to 3 (Mean = 2.2; SD = 0.5), and the Unified PD Rating Scale III (motor rating scaled from 108, where 108 represents severe motor impairment) ranging from 5 to 32 (Mean = 17.4; SD = 7.1). In addition, a healthy control (HC) group consisting of 15 men and 7 women of comparable age ranging from 40 to 91 years (Mean = 58.7; SD = 14.6) was included. Each participant was instructed to perform at least two times a diadochokinetic (DDK) task, which requires rapid steady /pa/-/ta/-/ka/ syllable repetition as long and constantly as possible, at least 5 times on one breath. A total of 116 vocal samples were obtained (56 for PD and 60 for HC). The speech data was recorded using an external condenser microphone placed at approximately 15 cm from the mouth and coupled to a Panasonic NV-GS 180 video camera. The voice signals were sampled at 48 kHz, with 16-bit resolution.

2.3. The normalized recursive Bayesian autoregressive changepoint detector

In this study, changepoint detection is based on autoregressive (AR) modeling, where prediction of the current signal sample is generally expected from a linear combination of previous samples and additive white noise, which is non-correlated with the normal distribution and zero mean value. In other words, speech can be modeled by the piecewise autoregressive model which is characterized by abrupt changes of its order and coefficients. Therefore, we further assume that the signal is composed of two sections which can be described by two different autoregressive models evaluating the magnitude of the change, the left AR model with M_1 parameters a_k and the right AR model with M_2 parameters b_k, i.e., Eq. (1)

$$x[n] = \begin{cases} \sum_{k=1}^{M_1} a_k \cdot x[n-k] + e[n], & \text{pro } n \leqslant m \\ \text{“left” signal of AR order } M_1 & \\ \sum_{k=1}^{M_2} b_k \cdot x[n-k] + e[n] & \text{pro } n > m \\ \text{“right” signal of AR order } M_2 & \end{cases} \quad n = 1, \cdots, N, \tag{1}$$

where $e[n]$ is zero mean noise, a and b are vectors of the model parameters, and m is the changepoint position.

Furthermore, we assume *a priori* that all parameters are equally likely. Using Bayesian marginalization, the parameters are integrated out, and therefore it is possible to find the relationship for calculation of a posteriori probability changes in the signal without the knowledge of any information regarding the values of the noise level or the linear coefficients (O Ruanaidh and Fitzgerald, 1996), i.e., Eq. (2)

$$p(\{m\}|\mathbf{x}, M) \approx \frac{(\mathbf{x}^{\mathbf{T}}\mathbf{x} - \mathbf{x}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}}(\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}})^{-1}\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{x})^{\frac{-(N-M)}{2}}}{\sqrt{\det(\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}})}}, \tag{2}$$

where m is the changepoint position, $\mathbf{x}$ is the column data vector of length N with elements $\mathbf{x}[n]$, $\mathbf{G_A}$ represents the matrix of the basis function, and number of matrix columns M depends on the AR model orders M_1, M_2 ($M = M_1 + M_2$). Hence, the original derived relationship (2) by O Ruanaidh and Fitzgerald (1996) introduced for two piecewise linear models cannot be used if the signal contains increased multiple changepoints. This limitation is very restrictive in practice, since certain multiple changepoints have always been presented in speech utterances. This disadvantage may be treated by calculating the posterior probability in the sliding window.

As pointed out in (O Ruanaidh and Fitzgerald (1996)), the posterior probability (2) is influenced only by the nominator of the Bayesian formula because it is derived under the condition that a given data segment is constant. Thus, the Bayesian evidence in the denominator of Bayesian formula is also constant. On the other hand, in the event that the posterior probability (2) is repeatedly used for a segmented signal, the data are not constant, and therefore Bayesian evidence has to be used to normalize posterior probability (Cmejla and Sovka, 2004), i.e., Eq. (3)

$$p(\{m\}|\mathbf{x},M) \approx \frac{(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{x}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}}(\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}})^{-1}\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{x})^{\frac{-(N-M_1-M_2)}{2}}}{\sqrt{\det(\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}})}} \times \frac{\sqrt{\det(\mathbf{G}_{\mathbf{E}}^{\mathbf{T}}\mathbf{G}_{\mathbf{E}})}}{(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{x}^{\mathbf{T}}\mathbf{G}_{\mathbf{E}}(\mathbf{G}_{\mathbf{E}}^{\mathbf{T}}\mathbf{G}_{\mathbf{E}})^{-1}\mathbf{G}_{\mathbf{E}}^{\mathbf{T}}\mathbf{x})^{\frac{-(N-M_1)}{2}}}, \quad (3)$$

where matrix $\mathbf{G}_{\mathbf{E}}$ consists of all segment samples without the separation into the "left" and "right" parts. Use of Bayesian evidence allows us to compare results between different signal segments, which subsequently allow the formation of the sliding window algorithm. This approach overcomes the disadvantage of using the classical BACD (2), which lies in its inability to compare results between different signal segments.

Moreover, the usage of logarithms can suppress the other negative aspects of numerical instability in the original relationship (2). First, we neglect both determinants, since their influence on the resulting probability density is very small in comparison with other terms in the resulting equation. Second, we omit multiplicative constants, which originate from the application of logarithms on the basic relationship (3). After using these two adjustments, we obtain an expression for calculating the posterior probability of the change position in the sliding window, i.e., Eq. (4)

$$\begin{aligned} p(\{m\}|\mathbf{x},M) &\approx \log(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{x}^{\mathbf{T}}\mathbf{G}_{\mathbf{E}}(\mathbf{G}_{\mathbf{E}}^{\mathbf{T}}\mathbf{G}_{\mathbf{E}})^{-1}\mathbf{G}_{\mathbf{E}}^{\mathbf{T}}\mathbf{x}) - \log(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{x}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}}(\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{G}_{\mathbf{A}})^{-1}\mathbf{G}_{\mathbf{A}}^{\mathbf{T}}\mathbf{x}) \\ &= \log(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{f}_{\mathbf{E}}^{\mathbf{T}}\mathbf{f}_{\mathbf{E}}) - \log(\mathbf{x}^{\mathbf{T}}\mathbf{x}-\mathbf{f}_{\mathbf{A}}^{\mathbf{T}}\mathbf{f}_{\mathbf{A}}), \end{aligned} \quad (4)$$

where $\mathbf{G}_{\mathbf{A}}$ represents the matrix of basis function and $\mathbf{G}_{\mathbf{E}}$ the matrix of Bayesian evidence. Subsequently, the general linear model can be described using the formula $\mathbf{f} = \mathbf{G}\cdot\hat{\mathbf{b}} = \mathbf{G}\cdot(\mathbf{G}^{\mathbf{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathbf{T}}\mathbf{x}$. After multiplying the general linear model by its transposition, we obtain the relationship $\mathbf{f}^{\mathbf{T}}\mathbf{f} = \mathbf{x}^{\mathbf{T}}\mathbf{G}\cdot(\mathbf{G}^{\mathbf{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathbf{T}}\mathbf{x}$. Fig. 1 shows the principle of the designed BACD.

The most significant changes in the signal are determined by the maximum a posteriori probability of the changepoint position. In our study, the tracking algorithm for abrupt changes can be considered as Bayesian because the result is a posteriori probability of the parameters instead of a hypothesis test or an estimated value. As well as other Bayesian methods, BACD requires subjective prior probabilities. The resulting relationship is dependent solely on samples of the input signal.

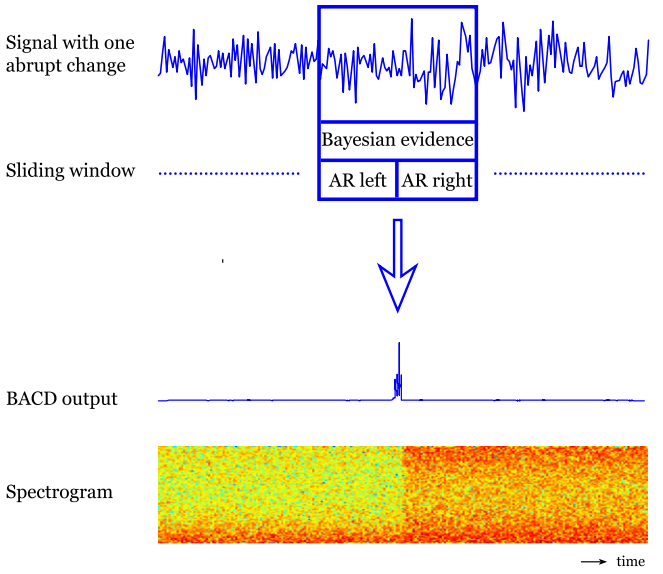

Fig. 1. Principle of the normalized Bayesian changepoint detector. Computing of a posteriori probability of changepoint position supposes two piecewise AR models with the same length and Bayesian evidence evaluation. Models are shifted with each new data sample in the sliding window.

In Eq. (4), the term $\mathbf{x}^{\mathbf{T}}\mathbf{x}$ represents the energy of the speech signal while the term $\mathbf{f}^{\mathbf{T}}\mathbf{f}$ demonstrates the energy of the signal model, and therefore the difference of both terms can be interpreted as the prediction error energy. If we maximize the likelihood function for the general linear model (detailed description can be found at page 17 in O Ruanaidh and Fitzgerald (1996)), assuming uniform priors of linear coefficients, we also obtain the expression $\mathbf{x}^{\mathbf{T}}\mathbf{x}$ - $\mathbf{f}^{\mathbf{T}}\mathbf{f}$, which represents the estimate of the variance independent of the linear coefficients. Although the full derivation of the detector was designed strictly on the basis of the Bayesian approach, the resulting Eq. (4) can be seen within the meaning of the criteria log likelihood ratio, where likelihood for the first model without change and likelihood for the second model with one change are compared. A similar approach can be found in (Ajmera et al., 2004), where the number of parameters is the same in the numerator and the denominator, and hence there is no need for regularization in the likelihood test.

However, the calculation of the changepoint positions places high demands on computation, which is the reason behind the development of the recursive BACD algorithm. The inputs of the recursive BACD include window length and AR model order $M = M_1 + M_2$. A posteriori probability $p(\{m\}|\mathbf{x}, M)$ is computed for the changepoint position m in the middle of the window length. A detailed mathematical description of recursive BACD as well as the $\mathbf{G}_{\mathbf{A}}$ and $\mathbf{G}_{\mathbf{E}}$ matrixes can be found in Table 1. The algorithm code in the ©Matlab environment can also be found online on http://sami.fel.cvut.cz/bacd/.

The recursive algorithm can be summarized as follows (see Table 1 for details):

Table 1
The recursive Bayesian changepoint detection algorithm for the calculation of a posteriori probability in the middle of the window.

I. Adjustment of both initial matrix of basis function $\mathbf{G_A}$ for changepoint $m = N/2$ and Bayesian evidence $\mathbf{G_E}$ for normalized estimate in window of the length wl, and initialization of energy $D = \mathbf{x^T x}$, correlation vectors $\mathbf{g_A} = \mathbf{x^T G_A}$, $\mathbf{g_E} = \mathbf{x^T G_E}$, and inverse of correlation matrices $\Phi_A = \left(\mathbf{G_A^T G_A}\right)^{-1}$, $\Phi_E = \left(\mathbf{G_E^T G_E}\right)^{-1}$ for changepoint position in the middle of the window.

$$\mathbf{G_A} = \begin{bmatrix} x[0] & x[-1] & \cdots & 0 & 0 & \cdots \\ x[1] & x[0] & \cdots & 0 & 0 & \cdots \\ x[2] & x[1] & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ x[N/2-1] & x[N/2-2] & \cdots & 0 & 0 & \cdots \\ 0 & 0 & \cdots & x[N/2] & x[N/2-1] & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & x[N-1] & x[N-2] & \cdots \end{bmatrix} \quad \mathbf{G_E} = \begin{bmatrix} x[0] & x[-1] & \cdots \\ x[1] & x[0] & \cdots \\ x[2] & x[1] & \cdots \\ \vdots & \vdots & \cdots \\ x[N/2-1] & x[N/2-2] & \cdots \\ x[N/2] & x[N/2-1] & \cdots \\ \vdots & \vdots & \cdots \\ x[N-1] & x[N-2] & \cdots \end{bmatrix}$$

II. Updating data

A. New observed signal sample x_{new} (in the end of the window wl)

$\mathbf{G_{A_{new}}} = [\underbrace{0\,0\,\cdots}_{M_1}\ \underbrace{x_{new}\,x_{new-1}\,\cdots}_{M_2}]$, $D_{new} = D + (x_{new})^2$, $\mathbf{w}_1 = \Phi_A \mathbf{G}^T_{\mathbf{A_{new}}}$,

$\mathbf{g}_{A_{new}} = \mathbf{g}_A + x_{new}\,\mathbf{G}_{\mathbf{A}_{new}}$, $\lambda = 1 + \mathbf{G}_{\mathbf{A}_{new}}\mathbf{w}_1$, $\Phi_{A_{new}} = \Phi_A - \mathbf{w}_1 \lambda^{-1} \mathbf{w}_1^T$

B. Removing old sample (in the begin of the window)

$\mathbf{Z} = [\underbrace{x_{new-wl-1}\,x_{new-wl-2}\,\cdots}_{M_1}\ \underbrace{0\,\cdots}_{M_2}]$, $D_{new2} = D_{new}$ - $(x_{new\text{-}wl})^2$, $\mathbf{w}_2 = \Phi_{Anew}\mathbf{Z}^T$, $\lambda = 1 - \mathbf{Z}\,\mathbf{w}_2$,

$\mathbf{g}_{A_{new2}} = \mathbf{g}_{A_{new}} - x_{new-wl}\,\mathbf{Z}$, $\Phi_{A_{new2}} = \Phi_{A_{new}} + \mathbf{w}_2 \lambda^{-1} \mathbf{w}_2^T$

III. Position update (in the middle of the window $wl/2$)

A. Replacing with row of zeros

$\mathbf{R} = [\underbrace{0\,0\,\cdots}_{M_1}\ \underbrace{x_{new-wl/2}\,x_{new-wl/2-1}\,\cdots}_{M_2}]$, $\mathbf{g}_{A_{new3}} = \mathbf{g}_{A_{new2}} - x_{new-wl/2}\,\mathbf{R}$,

$\mathbf{w}_3 = \Phi_{Anew2}\mathbf{R}^T$, $\lambda = 1 - \mathbf{R}\,\mathbf{w}_3$, $\Phi_{A_{new3}} = \Phi_{A_{new2}} + \mathbf{w}_3\,\lambda^{-1}\mathbf{w}_3^T$

B. Replacing with new data

$\mathbf{Q} = [\underbrace{x_{new-wl/2}\,x_{new-wl/2-1}\,\cdots}_{M_1}\ \underbrace{0\,0\,\cdots}_{M_2}]$, $\mathbf{g}_{A_{new4}} = \mathbf{g}_{A_{new3}} - x_{new-wl/2}\,\mathbf{Q}$

$\mathbf{w}_4 = \Phi_{Anew3}\mathbf{Q}^T$, $\lambda = 1 - \mathbf{R}\,\mathbf{w}_4$, $\Phi_{A_{new4}} = \Phi_{A_{new3}} + \mathbf{w}_4\,\lambda^{-1}\mathbf{w}_4^T$

IV. Bayesian evidence update

A. Adding new sample

$\mathbf{G}_{E_{new}} = [\underbrace{x_{new}\,x_{new-1}\,\cdots}_{(M_1+M_2)/2}]$, $\mathbf{g}_{E_{new}} = \mathbf{g}_E + x_{new}\,\mathbf{G}_{E_{new}}$, $\mathbf{w}_E = \Phi_E \mathbf{G}^T_{E_{new}}$,

$\lambda = 1 + \mathbf{G}_{E_{new}}\mathbf{w}_E$, $\Phi_{E_{new}} = \Phi_E - \mathbf{w}_E\,\lambda^{-1}\mathbf{w}_E^T$,

B. Removing old sample

$\mathbf{Z}_E = [\underbrace{x_{new-wl}\,x_{new-wl-1}\,\cdots}_{(M_1+M_2)/2}]$, $\mathbf{g}_{E_{new2}} = \mathbf{g}_{Enew} + x_{new}\,\mathbf{Z}_E$, $\mathbf{w}_{E_2} = \Phi_E \mathbf{Z}_E^T$,

$\lambda = 1 - \mathbf{Z}_{E_{new}}\mathbf{w}_E$, $\Phi_{E_{new}} = \Phi_E + \mathbf{w}_E\,\lambda^{-1}\mathbf{w}_{E_2}^T$

V. Calculate changepoint position probabilities

$$p(\{m\}|\mathbf{d}, M) \approx \log\left(D_{new2} - \mathbf{g}_{E_{new2}}\Phi_{E_{new2}}\mathbf{g}^T_{E_{new2}}\right) - \log\left(D_{new2} - \mathbf{g}_{A_{new4}}\Phi_{A_{new4}}\mathbf{g}^T_{A_{new4}}\right)$$

VI. Return to II. for new sample x and compute new change position $m + 1$

I. First, the initial matrices of basis function $\mathbf{G_A}$ and Bayesian evidence $\mathbf{G_E}$ for changepoint position in the middle of the window are assembled. Subsequently, the signal energy, correlation vectors, and inverse of correlation matrices are computed.

II. The designed recursive algorithm is initiated at this point. First, a new sample of the observed signal is added to the end of the window (II.A.). An increase in the quantity of data leads to an increase in the number of rows in matrices $\mathbf{G_A}$. Second, we have to update key matrix terms through extension of the matrices. Third, the old sample at the beginning of the window has to be removed (II.B.) and key matrix terms must be updated again.

III. In this step, the position in the middle of the window for the subsequent new sample will be updated. Updating is carried out in two stages. The first stage consists in replacing the m-th row of $\mathbf{G_A}$ with a row of zeroes **R.** The second stage then consists in replacing the row of zeros introduced into $\mathbf{G_A}$ with the matrix **Q** and modifying all relevant terms.

IV. In this step, the new sample must be added for Bayesian evidence (IV.A.), while the old sample has to be removed (IV.B.), and the terms updated.

V. Finally, having modified all the relevant terms, the logarithm of posterior density is computed for a new data sample and a new changepoint position.

VI. For computing the posterior density of the newly observed speech sample x and the new changepoint position $m+1$, we have to return to step II.

2.4. Assessment of fluency disorder

In typical disfluent speech, when compared to the same utterance by a healthy person, there are more silences and irregular prolongations, which are demonstrated by the decreasing number of spectral changes per time unit. Thus, we can assume that the number of spectral changes in the speech signal will decrease with respect to the severity of speech disfluency. In contrast, the variability of distances between individual spectral changes will increase as a consequence of the higher degree of disfluency. Based upon these assumptions, we introduce two novel measurements of speech disfluency. The *extent of speech fluency* (ESF) is calculated as the overall number of spectral changes divided by the entire duration of the speech sample. The *speech fluency variability* (SFV) is determined as a logarithm of standard deviation applied to the distances between two following spectral changes. In this case, our motivation behind using the logarithm results from the following phenomena. First, there are general suggestions in the literature that a number of speech features are better represented in the logarithmic domain (see for example Asgari and Shafran, 2010a). Second, perceptual judgements performed by the experts show a logarithmic dependence with respect to the severity of speech disfluency. The effectiveness of using a logarithm is highlighted by Fig. 2, which shows a nonlinear relationship between the number of disfluencies and the corresponding class according to the clinical experts' evaluation.

Although we can expect a large number of smaller local maxima representing spectral changes from the typical output of the BACD, only abrupt changes can be considered as significant. Therefore, a robust detector of abrupt changes is introduced for tracking the severity of speech disfluency. Fig. 3 shows the individual steps describing the algorithm for the abrupt change detector. First, the speech signal (see Fig. 3a) is down-sampled to 16 kHz and the output obtained from the normalized recursive BACD is filtered by a low-pass filter with a break frequency at 20 Hz. As a result, we gain the smooth filtered signal of the BACD output (see Fig. 3b). Second, the local minima of the BACD output are calculated and candidates of abrupt changes are then detected as the local maxima between two consecutive minima in the appropriate segments (see Fig. 3c). However, as we mentioned earlier, not every detected maximum necessarily corresponds to an abrupt change. Thus, as we will discuss further, there is a need to set up a threshold distinguishing abrupt changes from other, in this case insignificant changes. Finally, the significant abrupt changes retained for evaluation are those candidates greater than the threshold (see Fig. 3d). Fig. 3e shows the final output of the abrupt change detector plotted in the spectrogram of speech utterance.

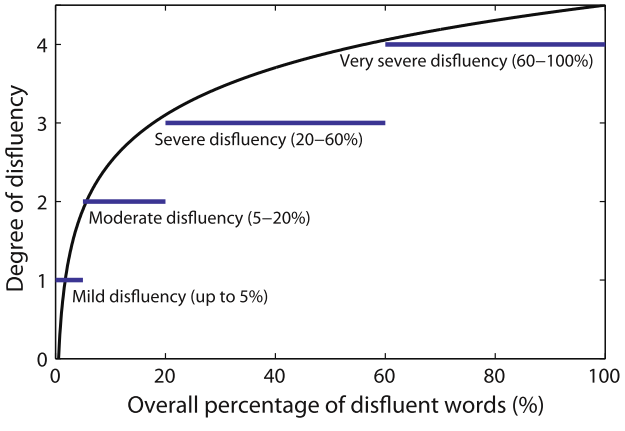

Fig. 2. The black line shows the logarithmic relationship between the number of disfluencies and the corresponding class according to the clinical experts' judgement. The blue lines represent individual classes related to the severity of disfluency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

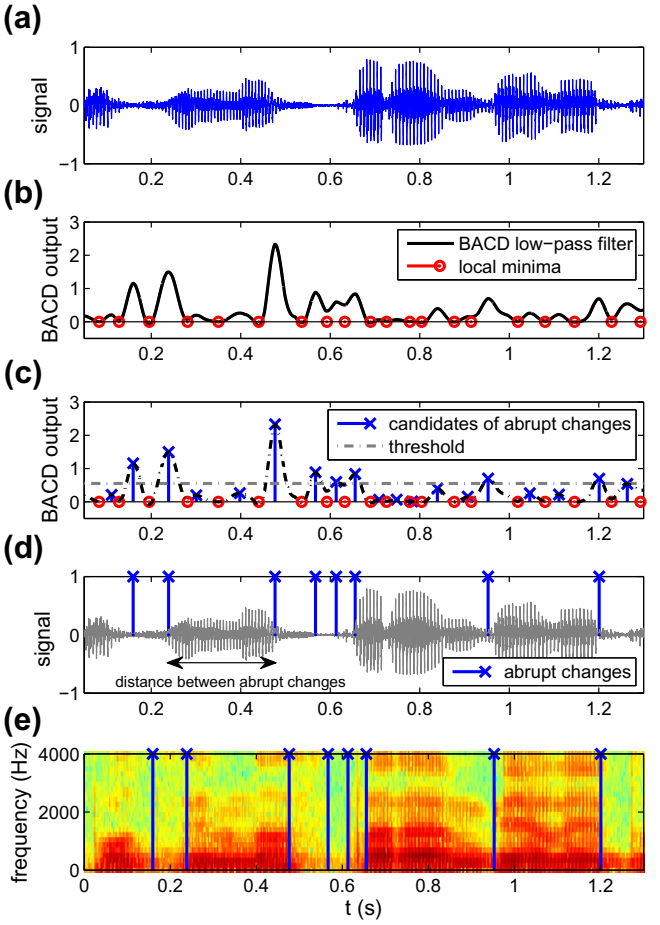

Fig. 3. Principle of the algorithm for the abrupt change detector. (a) Selected speech utterance of a stutterer; (b) Output of the normalized recursive Bayesian changepoint detector filtered through a low pass filter; (c) Candidates for abrupt changes; (d) Abrupt changes retained; (e) Abrupt changes highlighted in the speech spectrogram.

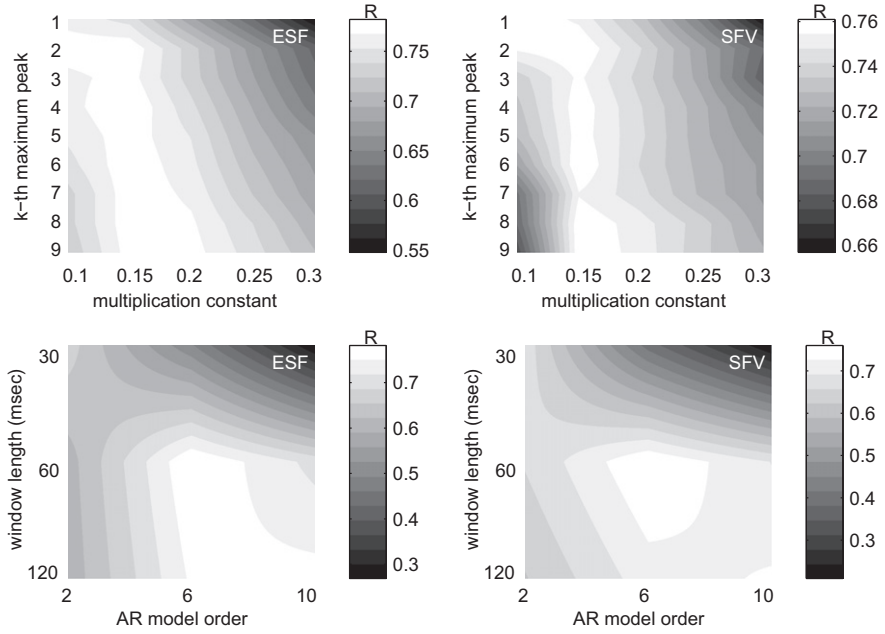

Fig. 4. Correlations between the average number of abrupt changes and the perceptual evaluation of speech fluency by experts for different threshold settings (above), as well as window length and AR model order (below).

Nevertheless, analyses performed with different speakers indicate that the amplitude of abrupt change candidates is quite variable. From this point of view, it is better to use the adaptive threshold for each speaker separately. In addition, we need to find the optimal input parameters of BACD including window length and AR model order. All those parameters can be obtained experimentally with the guidance of clinical ratings validated by developmental stuttering taxonomy. Fig. 4 shows the motivation behind the selection of representative parameter values. The threshold was calculated as a fraction from the k-th maximum peak of global BACD output and multiplication constant. Fig. 4 above shows the Pearson correlation between perceptual evaluations by experts and both measures (ESF and SFV) for different defined constants. The vertical axis represents the order of greatest detected maximum and horizontal axis represents multiplication constants (from 0.1 to 0.3) of k-th maximum peaks. As can be seen, the most significant correlation coefficients exceeding the value of 0.77 are illustrated by a white surface. In this case, the greatest coefficient of 0.78 was found for the fourth maximum and multiplication constant of 0.15. Subsequently, these values were set up as the adaptive threshold. Almost the identical procedure is applied to find the optimal window length and AR model order for the input of the BACD. Fig. 4 below shows the relationships between perceptual evaluations by experts and both measures for different window lengths and AR model orders. The testing range was set up from 2 to 10 with a step of 2 for the AR model order, and from 30 ms to 120 ms with a step of 10 ms for window length. As the result with the best classification performance, the sixth order of AR model (M_1, $M_2 = 6$) and the window length of 60 ms were used for further analysis.

2.5. Assessment of articulatory disorder

As opposed to speech disfluency, dysarthria is associated with imprecise articulation, which can be demonstrated by means of lower spectral changes. Therefore, more accurate speech articulation is expected to correspond with greater spectral changes as a consequence of consonant-to-vowel alteration. In other words, a posteriori probabilities of change positions are more proportional to the spectral distance between two adjacent segments, and therefore represent greater clarity of articulation. Thus, on the basis of the output of normalized recursive BACD, we are able to define parameters useful in the assessment of dysarthric speech. The *consonant-to-vowel transition accuracy* (CVTA) is computed as the mean of all spectral changes provided by BACD output. The *consonant-to-vowel variability* (CVV) is performed as a logarithm of standard deviation of all spectral changes. Fig. 5 shows a sample of PD and HC speech performance extracted from /pa/-/ta/-/ka/ syllable repetition using normalized recursive BACD.

Similarly to measurements of speech fluency, we need to find the optimal BACD input parameters of window length and AR model order. These parameters were obtained experimentally using the similar principle mentioned in Section 2.4. Analysis of variance (ANOVA) was applied to find the most significant differences between the PD and HC groups for a different set-up of window length and AR model order. The testing values ranged from 2 to 10 with a step of 2 for the AR model order, and from

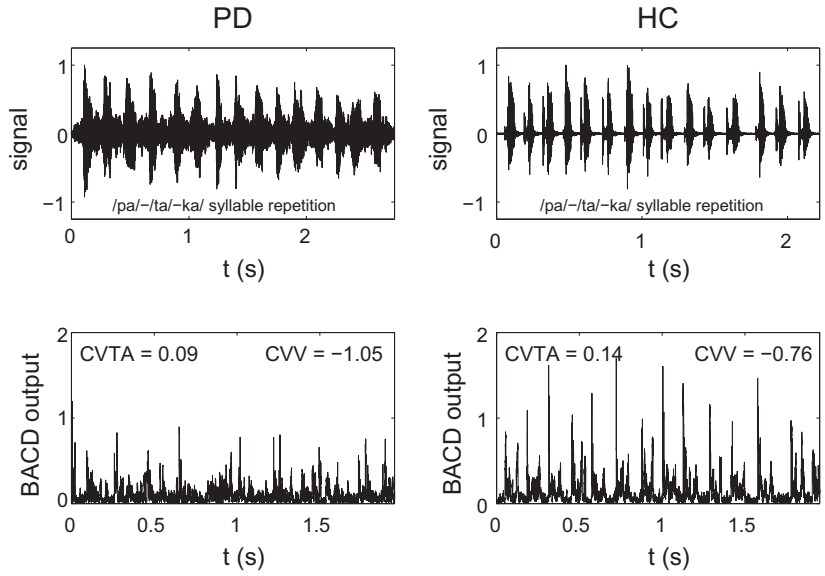

Fig. 5. Details of consonant-to-vowel measures performed on rapid steady syllable repetition. (up) Speech signals of /pa/-/ta/-/ka/ syllable repetition; (down) Output of the normalized recursive Bayesian changepoint detector. The left panel is for a person with PD, the right panel is for a HC subject.

Table 2
List of results of both fluency measures (ESF and SFV) with mean, SD values, and statistical significances according to the ANOVA with comparisons between groups by *post hoc* Bonferroni adjustment.

	ESF (1/s)	SFV (−)
	Mean ± SD	Mean ± SD
Normal healthy speech (0)	352.6 ± 92.5	2.89 ± 0.19
Mild disfluency (1)	295.3 ± 89.5	3.02 ± 0.20
Moderate disfluency (2)	245.6 ± 78.7	3.16 ± 0.20
Severe disfluency (3)	197.0 ± 47.1	3.26 ± 0.17
Very severe disfluency (4)	158.2 ± 44.1	3.60 ± 0.35
	Comparison between the groups	
ANOVA	$F(4, 117) = 38.3^{*}$	$F(4, 117) = 35.7^{*}$
0 vs. 1	NS	NS
1 vs. 2	$p < 0.001$	$p < 0.01$
2 vs. 3	$p < 0.001$	$p < 0.001$
3 vs. 4	NS	NS

NS = not significant.
* $p < 0.001$.

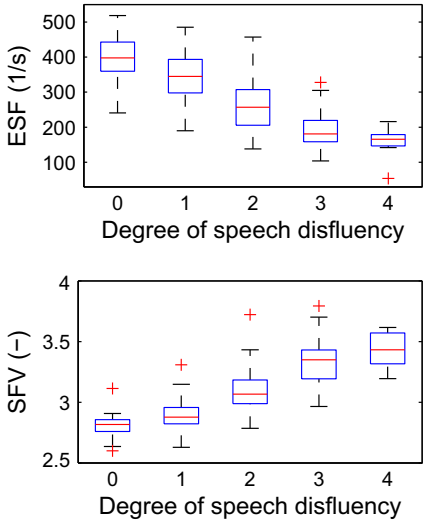

Fig. 6. Boxplots of new fluency parameters including extent of speech fluency (ESF) and speech fluency variability (SFV) for different degrees of speech disfluency according to the clinical experts' judgement.

5 ms to 40 ms with a step of 5 ms for window length. Using the above-mentioned principle, the sixth order of AR model (M_1, $M_2 = 6$) and window length of 20 ms was used for further analysis.

2.6. Statistics

Differences between groups (degrees of speech fluency and PD vs. HC subjects) were statistically compared using the ANOVA with *post hoc* Bonferroni adjustment for each acoustic variable (ESF and SFV for fluency disorders and CVTA and CVV for articulation disorders). The Kolmogorov–Smirnov test was used to test for the normality of the distribution of the data. As the acoustic variables showed a Gaussian distribution, the relationships between the acoustic variables as well as between expert evaluations and measurements of speech fluency were assessed using the Pearson product-moment correlation. The level of significance was set at $p < 0.05$.

3. Results

3.1. Assessment of fluency disorder

Table 2 details the results of the speech fluency investigations. As can be seen from Fig. 6, the number of significant spectral changes represented by ESF decreases as the severity of speech disfluency increases. In contrast, the

Table 3
List of results for both articulation measures (CVTA and CVV) with mean, SD values, and statistical significances according to the ANOVA with comparison between groups by *post hoc* Bonferroni adjustment.

	CVTA (–)	CVV (–)
	Mean ± SD	Mean ± SD
PD	0.15 ± 0.04	−0.80 ± 0.13
HC	0.18 ± 0.03	−0.68 ± 0.07
	Comparison between the groups	
ANOVA	$F(1, 115) = 35.5^{*}$	$F(1, 115) = 33.1^{*}$
PD vs. HC	$p < 0.001$	$p < 0.001$

* $p < 0.001$.

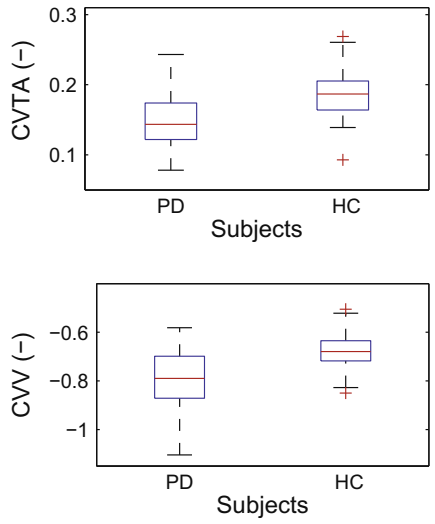

Fig. 7. Boxplots of new articulatory parameters including consonant-to-vowel transition accuracy (CVTA) and consonant-to-vowel variability (CVV) for differentiation between PD and HC groups.

variability of distances between spectral changes expressed by SFV is increased with respect to degree of disfluency.

Statistically significant differences according to ANOVA were found between the five degrees of disfluency for both measures of ESF [$F(4, 117) = 38.3$, $p < 0.001$] and SFV [$F(4, 117) = 35.7$, $p < 0.001$]. A *post hoc* Bonferroni adjustment indicated significant differences between mild and moderate degrees of disfluency (1 vs. 2) for both measures (ESF, $p < 0.001$; SFV, $p < 0.01$) and also between moderate and severe degrees of disfluency (2 vs. 3) for both measures (ESF, $p < 0.001$; SFV, $p < 0.001$). The relationships between the measures of speech disfluency and expert evaluations were found to be significant, with high correlations for ESF ($r = -0.75$, $p < 0.001$) and SFV ($r = 0.74$, $p < 0.001$). A significant correlation was also found between ESF and SFV ($r = -0.89$, $p < 0.001$).

3.2. Assessment of articulatory disorder

Table 3 lists the detailed results of articulatory disorders assessment. Fig. 7 shows that people with PD exhibit more inaccurate consonant-to-vowel articulation when compared to HC subjects. These differences are accurately captured by the measures of CVTA and CVV.

ANOVA showed statistically significant differences for both articulatory measures of CVTA [$F(1, 115) = 35.5$, $p < 0.001$] and CVV [$F(1, 115) = 33.1$, $p < 0.001$]. A *post hoc* Bonferroni adjustment indicated a significant difference between the PD and HC groups for both measurements (CVTA, $p < 0.001$; CVV, $p < 0.001$). The correlation between CVTA and CVV was found to be statistically significant ($r = 0.77$, $p < 0.001$).

4. Discussion

In this study, we present a novel, robust changepoint detection with application for automatic assessment of fluency and articulatory disorders. This method extends the standard usage of Bayesian changepoint detection, which is included in many different digital signal processing tasks. The originality of the modification consists in its normalizing of a posteriori probability using Bayesian evidence, which allows for easy processing of signals using a sliding window. Furthermore, the method is implemented in a recursive algorithm for practical use.

To show the reliability of the proposed BACD method in speech pathology assessment, we have tested our detector on two types of central nervous system disorders, including people with stuttering and PD. Although the speech was affected in two different ways in dependence on the type of disorder, our method showed its power to evaluate and differentiate various degrees of speech impairment. The stuttering group is represented by a wide range of speech performances, including participants with almost no signs up to severe levels of disfluency. Here, we have shown the efficiency of the BACD in the estimation of disfluency severity, which could be useful in support for therapists' evaluation as well as feedback in speech treatment (Van Borsel et al., 2003). On the other hand, in the PD dataset, the patients were newly diagnosed with an early stage of the disease, and most of them with only mildly impaired speech. In this case, on the basis of articulation measures, we demonstrated the suitability of BACD in catching even small alterations of speech and distinguishing PD patients from healthy speakers, with the motivation of potential early diagnosis, essential in improving the life of such affected persons (Singh et al., 2007).

To ensure comparable results, we have chosen two standard procedures such as text reading and fast syllable repetition (DDK task) commonly used in speech pathology assessment. Subsequently, the need emerged for a valid reference with respect to evaluation of our method. Thus, the degree of speech disfluency was audited by the independent perceptual ranking of two experts with a very good inter-rater reliability. Nevertheless, to reduce the subjectivity in data and ensure the independency of designed parameters, the overall disfluency percentage was calculated in order to create a reliable score, which was treated as an appropriate reference with a highly significant relationship to the

clinical ratings ($p < 0.001$). In addition, the mean of both experts was computed to reduce the raters' subjectivity and obtain a final score for further evaluation. As a result, our speech fluency measures based on BACD were significantly correlated ($p < 0.001$) with the perceptual judgement performed by the speech therapists. In the case of the articulation disorder, the assessment by experts was not performed because minor dysarthria-related speech changes during fast syllable repetition captures by BACD were hardly evaluable by standard perceptible tests.

When assessing the group differences between dysarthric and healthy speakers, individuals with PD were found to demonstrate a significant reduction in consonant-to-vowel articulation performance compared with healthy speakers ($p < 0.001$). On the other hand, in the stuttering group, we were only able to find a significant separation ($p < 0.01$) the mild to severe levels of disfluency (1 vs. 2 vs. 3), while differences between healthy to mild (0 vs. 1) and severe to very severe (3 vs. 4) levels of disfluency were not found to be statistically significant. In the first case, the underlying reason could be the small difference between healthy to mild levels of disfluency, because even healthy people commonly exhibit some disfluencies (Goberman et al., 2010), making both groups hard to separate by perceptual ranking. In the second case, only 7 participants were categorized at the very severe stage of disfluency, which could result in a low separable performance for our method. Finally, although the correlations between two fluency measures as well as between the two articulation measures were found to be highly significant ($p < 0.001$), each measure represents one certain aspect of speech. All these aspects in total could then increase the overall accuracy of evaluation performance.

In general, our changepoint detection algorithm has been designed for automatic acoustic assessment of voice pathology with respect to the relevance of changes in spectral discontinuity and the spectral envelope of the speech signal. Accordingly, considerable effort has been invested in the previous literature towards the development of methods allowing for extraction of relevant speech features from pathological utterances. Some authors have noted the benefits of parameters extracted from speech in predicting the average symptom severity of PD (Asgari and Shafran 2010a; Asgari and Shafran 2010b; Tsanas et al., 2011). Previous studies have also revealed that people with hypokinetic dysarthria can be distinguished from healthy speakers on the basis of spectral changes (Rosen et al., 2006). Moreover, in accordance with our findings, the degree of spectral change has been shown to be higher during clear speech, which among other aspects is associated with more precise articulation (Rosen et al., 2011). In the domain of stuttering disfluencies, several researchers have designed automatic recognition stuttering systems mostly based upon Hidden Markov Models and Mel Frequency Cepstral Coefficients approaches (Noth et al., 2000; Wisniewski et al., 2007; Ravikumar et al. 2009; Hariharan et al., 2012). Although the classification performance of approximately 80–90% of these stuttering-related systems is very promising, changepoint detection could be helpful in providing more insights related to the progression of various speech disorders without requiring a speech recognizer optimized for the given task. Finally, there are also several software packages allowing detection of voice segments. The most popular of them include PRAAT (Boersma and Weenink, 2001), WaveSurfer (Sjolander and Beskow, 2000), and openSMILE (Eyben et al., 2010). These software packages offer the extraction of various speech-related features such as jitter, shimmer, noise-to-harmonics ratios, signal energy, loudness, cepstral coefficients, pitch, formants, linear predictive coding, line spectral pairs, zero-crossing, sound annotation/transcriptions, and many others. Thus, there is a great opportunity to use these features in speech pathology assessment which could yield the design of parameters measuring similar aspects of speech as presented in this study. However, to the best of our knowledge, there is no widely-used speech-analysis software that has implemented changepoint detection.

Although we have shown the wide range of possible applications of changepoint detection in the area of speech pathology, the problem of reliable changepoint detection has attracted much interest in other research fields using several varying divergence metrics. Several comparisons with other divergence metric based approaches have already been reported in our previous research (Bergl, 2006; Bergl and Cmejla, 2007). These comparisons were based upon two experiments using synthetic and real speech signals. First, the experiment performed on synthetic AR signals with sliding windows (Bergl, 2006) includes detection of abrupt changes with an increasing value of spectral divergence (single boundary) and pairs of abrupt changes with increasing distances (multiple boundaries). In this experiment, our BACD method and General Likelihood Ratio distance (Appel and Brandt, 1983) have reached excellent results for the detection of both single and multiple boundaries. Similarly, Kullback–Leiber divergence (Couvreur and Boite, 1999) and the Mahalanobis distance (Sooful and Botha, 2001) were able to detect a single boundary very accurately while failing to detect multiple boundaries. The poor quality of detection can be the result of not using the AR model. On the other hand, this aspect might be an advantage for analysis of real signals, which can be hardly described by the AR model. Second, in the experiment using real speech signals (Bergl and Cmejla, 2007), Kullback–Leibler divergence has been proven to be the best method for finding boundaries between vowels and nasals, while other detectors including the Bhattacharyya divergence (Mak and Barnard, 1996), L2 metric (Sooful and Botha, 2001), and Jeffreys-Matusita measure (Sooful and Botha, 2001), have a tendency to miss vowel-nasals boundaries. Abrupt changes between vowel and silence were detectable by most of the above-mentioned detectors, but the best results were achieved using Bhattacharyya divergence. While most detectors react in

the occlusion–burst context, our BACD was the best method for precise detection of the burst–vowel and fricative–vowel boundaries. The underlying reason could be caused by the fact that AR models are considerably more accurate for vowels than for silence.

We believe that the designed BACD will be important for several reasons. It can be helpful in diagnosis of the disease, classifying the severity of disease, monitoring treatment, and support for therapists in the evaluation. In other words, it can be easily included in the rising need for telemedicine approaches to provide useful information for therapists and patients without the need for clinical visits. Hence, usage of the tool such as BACD may result in the improvement of patients' health and quality of their life as well as in lowering the cost of treatment. In addition, the BACD can be widely used in a number of other areas requiring reliable signal processing techniques for change-point detection. Moreover, the BACD is a freely available code that can be downloaded from the project homepage on http://sami.fel.cvut.cz/bacd/.

5. Conclusion

The BACD-based measurements were able to make a significant differentiation of the various degrees of speech disfluency as well as articulation deficits in mild dysarthria levels. Admittedly, the measures of speech fluency were in agreement with perceptual evaluation by the speech experts. To summarize, our method provides a cost-effective, easily applicable, freely available, and objective assessment of fluency and articulatory disorders. Future research could extend this technology for use in additional practical applications.

Acknowledgments

The authors are obliged to Miroslava Hrbkova, Libor Cerny, Hana Ruzickova, Jiri Klempir, Veronika Majerova, Jana Picmausova, Jan Roth, and Evzen Ruzicka for provision of clinical data. The study was partly supported by the Czech Science Foundation, project GACR P102/12/2230, Czech Ministry of Health, projects NT 12288-5/2011 and NT11460-4/2010, and Czech Ministry of Education, project MSM 0021620849.

References

Ackermann, H., Konczak, J., Hertrich, I., 1997. The temporal control of repetitive articulatory movements in Parkinson's disease. Brain and Language 56, 312–319.

Ajmera, J., McCowan, I., Bourlard, H., 2004. Robust speaker change detection. IEEE Signal Processing Letters 11, 649–651.

Appel, U., Brandt, V.A., 1983. Adaptive segmentation of piecewise stationary time series. Information Sciences 29, 27–56.

Asgari, M., Shafran, I., 2010a. Predicting severity of Parkinson's disease from speech. In: 32nd Ann International Conference of the IEEE Engineering in Medicine & Biology Society (EMBS 2010), Buenos Aires, Argentina, pp. 5201–5204.

Asgari, M., Shafran, I., 2010b. Extracting cues from speech for predicting severity of Parkinson's disease. In: IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2010), Kittila, Finland, pp. 462–467.

Basseville, M., Nikoforov, I., 1993. Detection of abrupt changes: theory and application. Information and System Science Series. Prentice Hall, Englewoods Cliffs.

Bergl, P., 2006. Accuracy of divergence measures for detection of abrupt changes. In: World Academy of Science, Engineering and Technology, vol. 18, pp. 33–36.

Bergl, P., Cmejla, R., 2007. Improved detection of boundaries of phonemes in speech databases. In: Proceedings of the fifth IASTED International Conference: Biomedical Engineering (BIEN '07), ACTA Press Anaheim, CA, USA, pp. 171–174.

Bergl, P., 2010. Objectification of speech disfluencies (in Czech). PhD thesis, Czech Technical University in Prague.

Bocklet, T., Riedhammer, K., Noth, E., Eysholdt, U., Haderlein, T., 2012. Automatic intelligibility assessment of speakers after laryngeal cancer by means of acoustic modelling. Journal of Voice 26, 390–397.

Boersma, P., Weenink, D., 2001. PRAAT, a system for doing phonetics by computer. Glot International 5, 341–345.

Chu, P.S., Zhao, X., 2004. Bayesian change-point analysis of tropical cyclone activity: the central North Pacific case. Journal of Climate 17, 4893–4901.

Cmejla, R., Sovka, P., 2001. Estimation of boundaries between units using Bayesian changepoints detectors. In: 4th International Conference on Text, Speech, and Dialogue (TSD 01), Zelezna Ruda, Czech Republic, pp. 291–298.

Cmejla, R., Sovka, P., 2004. Recursive Bayesian autoregressive change-point detector for sequential signal segmentation. In: EUSIPCO Proceedings, Wien, pp. 245–248.

Conture, E., 2001. Stuttering: its Nature, Diagnosis, and Treatment. Allyn & Bacon, Boston.

Couvreur, L., Boite, J.M., 1999. Speaker tracking in broadcast audio material in the framework of the THISL project. In: Procedings of ESCA ETRW Workshop on Accessing Information in Spoken Audio, Cambridge (UK), pp. 84–89.

Darley, F.L., Aronson, A.E., Brown, J.R., 1969a. Differential diagnostic patterns of dysarthria. Journal of Speech and Hearing Research 12, 249–269.

Darley, F.L., Aronson, A.E., Brown, J.R., 1969b. Cluster of deviant speech dimensions in the dysarthrias. Journal of Speech and Hearing Research 12, 462–469.

Dejonckere, P., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., van de Heyning, P., Remacle, M., Woisard, W., 2001. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. European Archives of Oto-Rhino-Laryngology 258, 77–82.

Dobigeon, N., Tourneret, J.-Y., Scargle Jeffrey, D., 2005. Change-point detection in astronomical data by using a hierarchical model and a Bayesian sampling approach. In: 13th IEEE Workshop on Statistical Signal Processing, Vordeaux, France, pp. 335–340.

Duffy, J.R., 2005. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. Mosby, New York, pp. 592.

Eyben, F., Wollmer, M., Schuller, B. 2010. openSMILE – the munich versatile and fast open-source audio feature extractor. In: Proceedings of ACM Multimedia, Florence, Italy, pp. 1459–1462.

Goberman, A.M., Blomgren, M., Metzger, E., 2010. Characteristics of speech disfluency in Parkinson's disease. Journal of Neurolinguistics 23, 470–478.

Godino-Llorente, J.I., Gomez-Vilda, P., 2004. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. IEEE Transactions on Bio-Medical Engineering 51, 380–384.

Harel, B.T., Cannizaro, M.S., Cohen, H., Reilly, N., Snyder, P.J., 2004. Acoustic characteristics of Parkinsonian speech: a potential biomarker

of early disease progression and treatment. Journal Neurolinguistics 17, 439–453.
Hariharan, M., Chee, L.S., Ai, O.C., Yaacob, S., 2012. Classification of speech disfluencies using LPC based parameterization techniques. Journal of Medical Systems 36, 1821–1830.
Hartelius, L., Svensson, P., 1994. Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. Folia Phoniatrica et Logopedica 46, 9–17.
Hawkins, D.M., Zamba, K.D., 2005. Statistical process control for shifts in mean or variance using a changepoint formulation. Technometrics 47, 164–173.
Henriquez, P., Alonso, J.B., Ferrer, M.A., Travieso, C.M., Godino-Llorente, J.I., Diaz-de-Maria, F., 2009. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. IEEE Transactions on Audio Speech, and Language Processing 17, 1186–1195.
Hirano, M., 1981. Clinical Examination of Voice. Springer, Vienna.
Hornykiewicz, O., 1998. Biochemical aspects of Parkinson's disease. Neurology 51 (Suppl.), S2–S9.
Jacobson, B.H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M.S., 1997. The Voice Handicap Index (VHI). American Journal of Speech-Language Pathology 6, 66–70.
Lastovka, M., Vokral, J., Cerny, L., Radilova, K., Hrdlickova, M., 1998. Speech disfluency assessment using neural networks (in Czech). Research report 237/1998/C/1.LF.
Lechta, V., 2004. Stuttering (in Czech). Praha: Portal.
Kay Elemetrics Corp., 2003. Multi-Dimensional Voice Program (MDVP): Software Instroduction Manual. Lincon Park, Kay Elemetrics.
Kent, R., Weismer, G., Kent, J., Rosenbek, J., 1989. Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders 54, 482–499.
Kent, R., Weismer, G., Kent, J., Vorperian, H., Duffy, J., 1999. Acoustic studies of dysarthric speech: methods, progress, and potential. Journal of Communication Disorders 32, 141–186.
Maier, A., Haderlein, T., Eysholdt, F., Rosanowski, F., Batliner, A., Schuster, M., Noth, E., 2009. PEAKS – A system for the automatic evaluation of voice and speech disorders. Speech Communication 51, 425–437.
Mak, B., Barnard, E., 1996. Phone clustering using the Bhattacharyya distance. In: 4th International Conference on Spoken Language Processing (ICSLP96), vol. 4.
Middag, C., Van Nuffelen, G., Martens, J.P., De Bodt, M., 2008. Objective intelligibility assessment of pathological speakers. In: Proceedings of International Conference on Spoken Language Processing (Interspeech 08), Brisbane, Australie, pp. 1745–1748.
Noth, E., Niemann, H., Haderlein, T., Decher, M., Eysholdt, U., Rosanowski, F., Wittenberg, T., 2000. Automatic stuttering recognition using hidden Markov models. In: Proceedings of International Conference on Spoken Language Processing (ICSLP 2000), Beijing, China, pp. 65–68.
O Ruanaidh, J.J., Fitzgerald, W.J., 1996, Numerical Bayesian methods applied to signal processing. Series on Statistics and Computing, Springer-Verlag.
Prochazka, A., Kukal, J., Vysata, O., 2008. Wavelet transform use for feature extraction and EEG signal segments classification. In: Proceedings of the 3rd International Conference on Communications, Control and Signal Processing, Malta, pp. 719–722.
Ravikumar, K.M., Rajagopal, R., Nagaraj, H.C., 2009. An approach for objective assessment of stuttered speech using MFCC features. In: ICGST International Journal of Digital Signal Processing, vol. 9, pp. 19–24.
de Rijk, M.C., Tzourio, C., Breteler, M.M., Dartiques, J.F., Amaduci, L., Lopez-Pousa, S., Manubens-Bertran, J.M., Alperovitch, A., Rocca, W.A., 1997. Prelevance of parkinsonism and Parkinson's disease in Europe: concerted action on the epidemiology of Parkinson's disease. Journal of Neurology, Neurosurgery and Psychiatry 62, 10–15.
Riley, G.D., 1972. A stuttering severity instrument for children and adults. Journal of Speech and Hearing Disorders 37, 314–322.
Reeves, J., Chen, J., Wang, X.L.L., Lund, R., Lu, Q.Q., 2007. A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology 46, 900–915.
Robertson, S.J., Thompson, F., 1986. Working with Dysarthrics: A Practical Guide to Therapy for Dysarthria. Winslow Press, Oxford.
Rosen, K.M., Kent, R.D., Delaney, A.L., Duffy, J.R., 2006. Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers. Journal of Speech, Language, and Hearing Research 49, 395–411.
Rosen, K.M., Folker, J.E., Murdoch, B.E., Vogel, A.P., Cahill, L.M., Delatycki, M.B., Corben, L.A., 2011. Measures of spectral change and their application to habitual, slow, and clear speaking modes. International Journal of Speech-Language Pathology 13, 165–173.
Ruben, R., 2000. Redefining the survival of the fittest: communication disorders in 21th century. Laryngoscope 110, 241–245.
Ruggiero, C., Sacile, R., Giacomini, M., 1999. Home telecare. Journal of Telemedicine and Telecare 5, 11–17.
Rusz, J., Cmejla, R., Ruzickova, H., Ruzicka, E., 2011a. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. Journal of the Acoustical Society of America 129, 350–367.
Rusz, J., Cmejla, R., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., Ruzicka, E., 2011b. Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test. Movement Disorders 26, 1951–1952.
Sapir, S., Ramig, L.O., Spielman, J.L., Fox, C., 2010. Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. Journal of Speech, Language and Hearing Research 53, 114–125.
Singh, N., Pillay, V., Choonara, Y.E., 2007. Advances in the treatment of Parkinson's disease. Progress in Neurobiology 81, 29–44.
Sjolander, K., Beskow, J. 2000. Wavesurfer – an open source speech tool. In: Proceedings of International Conference on Spoken Language Processing, pp. 464–467.
Sooful, J.J., Botha, J.C., 2001. An acoustic distance measure for automatic cross-language phoneme mapping. PRASA'01, South Africa, pp. 99–102.
Su, H.Y., Wu, C.H., Tsai, P.J., 2008. Automatic assessment of articulation disorders using confident unit-based model adaptation. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 08), Las Vegas, USA, pp. 4513–4516.
Teesson, K., Packman, A., Onslow, M., 2003. The Lindcombe behavioural data language of stuttering. Journal of Speech, Language, and Hearing Research 46, 1009–1015.
Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O., 2011. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. Journal of the Royal Society, Interface 8, 842–855.
Ureten, O., Serinken, N., 1999. Bayesian detection of radio transmitter turn-on transients. In: IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 99), Antalya, Turkey, pp. 830–834.
Van Borsel, J., Reunes, G., Van den Bergh, N., 2003. Delayed auditory feedback in the treatment of stuttering: clients as consumers. International Journal of Language and Communication Disorders 38, 119–129.
Western, B., Kleykamp, M., 2004. A Bayesian change point model for historical time series analysis. In: Ann. Meeting of the Social–Science–History-Association, St Louis, Missouri, pp. 354–374.
Wisniewski, M., Kuniszyk-Jozkowiak, W., Smolka, E., Suszynski, W., 2007. Automatic detection of disorders in continuous speech with the hidden Markov Models approach. Computer Recognition Systems 2 (45), 445–453.
Wong, H., Hu, B.Q., Lp, W.C., Xia, J., 2006. Change-point analysis of hydrological time series using grey relational method. Journal of Hydrology 324, 323–338.
Yairi, E., Ambrose, N., 1999. Early childhood stuttering I: persistence and recovery rates. Journal of Speech, Language, and Hearing Research 42, 1098–1112.