RECURSIVE BAYESIAN AUTOREGRESSIVE CHANGEPOINT DETECTOR FOR SEQUENTIAL SIGNAL SEGMENTATION

Roman Čmejla, and Pavel Sovka

Dept. of Circuit theory, Faculty of Electrical Engineering, Czech Technical University in Prague
Technická 2, 166 27 Prague, Czech Republic (Europe)
phone: + 420 2 2435 2291, fax: + 420 2 311 98 05, email: cmejla@feld.cvut.cz
web: http://amber.feld.cvut.cz/user/cmejla/

ABSTRACT

The contribution addresses a sliding window modification of the Bayesian autoregressive change-point detector (BACD) enabling the sequential localization of signal changes (change-point detection). The modification consists in using the simplified data-dependent Bayesian evidence normalizing the classical BACD formula and in the recursive evaluation of these two functions. The suggested approach seems to be computationally effective and numerical stable as shown by experiments. Apart from the evaluation of the algorithm accuracy two illustrative examples with modelled signals are given. One application to the violin signal segmentation demonstrates the algorithm performance – even relatively weak and gradual signal changes can be detected.

1. INTRODUCTION

The signal segmentation has been intensively studied for decades. A great number of methods and applications have been published. Robust algorithms using recursive identification, likelihood and Bayesian approaches have been designed for the batch or sequential detection of multiple change-points detection [1], [2], [4], [6].

The first step in signal segmentation is the change-point detection. One of the methods is the Bayesian autoregressive change-point detector [2] which assumes piecewise constant parameters of AR signal and one change-point in one signal segment. The numerical behaviour of the BACD strongly depends on a signal length (number of samples). Greater number of samples gives more consistent estimation of correlations used in the BACD formula but on the other hand causes numerical difficulties. The condition requiring only one change in one signal segment is too limiting for the analysis of real signals as speech, music or biological signals. The segmentation of these signals requires the sequential detection of multiple change-points. Therefore the recursive BACD algorithm based on the sequential Bayesian approach has been suggested [8]. This growing window algorithm uses the data and change-point position updates [2], [7]. Another very simple idea for the sequential change-point detection seems to be the signal segmentation and repeatedly using the BACD. This method requires a proper choice of the segment length and suffers from the sensitivity to a noise. Moreover the BACD results gained from different segments can not be compared. Nevertheless if a normalization of the BACD is used together with the data and change-point position updates then a simple sliding window algorithm suitable for the sequential change-point detection can be obtained.

2. METHOD DESCRIPTION

Firstly, the brief definition of the BACD will be given. Secondly, the BACD use for a segmented signal will be analysed. Finally, the BACD normalization and the recursive algorithm together with the decreasing of computational costs will be described.

2.1 BACD definition

The signal model for the BACD consists of two parts: "left" generated by AR model with M_1 parameters a_k and "right" generated by another AR model with M_2 parameters b_k [2]

$$d[n] = \begin{cases} \sum_{k=1}^{M_1} a_k \cdot d[n-k] + e[n], & n \le m \\ \sum_{k=1}^{M_2} b_k \cdot d[n-k] + e[n], & n > m \end{cases} \qquad n = 1, \cdots, N . \tag{1}$$

In matrix form: $\mathbf{d} = \mathbf{G}_\mathrm{A} \cdot \mathbf{b}_\mathrm{A} + \mathbf{e}$. The matrix $\mathbf{G}_\mathrm{A}$ has the Jordan form and depends on the unknown change-point index $m = 1,...,N$ determined by the maximum of the posterior (MAP) probability density function (pdf) [2]

$$p(m \mid \mathbf{d}, \mathrm{M}) \propto \frac{\left[\left(D - \mathbf{g}_\mathrm{A} \boldsymbol{\Phi}_\mathrm{A} \mathbf{g}_\mathrm{A}^\mathrm{T}\right)\right]^{-\frac{N-M}{2}}}{\sqrt{\Delta_\mathrm{A}}}, \; M = M_1 + M_2 . \tag{2}$$

Matrix $\boldsymbol{\Phi}_\mathrm{A} = (\mathbf{G}_\mathrm{A}^\mathrm{T} \mathbf{G}_\mathrm{A})^{-1}$ is the inverse correlation matrix, $D = \mathbf{d}^\mathrm{T}\mathbf{d}$ is the signal energy, $\mathbf{g}_\mathrm{A} = \mathbf{d}^\mathrm{T}\mathbf{G}_\mathrm{A}$ is the correlation vector, and $\Delta_\mathrm{A} = \det(\mathbf{G}_\mathrm{A}^\mathrm{T}\mathbf{G}_\mathrm{A})$. Equation (2) giving the posterior pdf is referred to as the classical BACD formula.

2.2 Signal segmentation and the BACD normalization

Let us consider the problem of the sequential localization of change-points. Starting from the simple idea of repeatedly using the classical BACD on signal segments the BACD modification can be explained. First, the following problems connected with using the classical BACD for the segmented signal will be discussed.

P1. Choice of the segment length. The longer segments are the more consistent estimation of correlations $\mathbf{\Phi}$ and $\mathbf{g}$ can be obtained but the occurrence more than change-point in one segment is very probably.

P2. Segments overlap. The greater overlapping is the better time resolution can be achieved then but also the greater number of operations is needed.

P3. Posterior pdf normalization. The most important disadvantage of the repeatedly used BACD formula is the impossibility to compare posterior pdfs $p_l(m\,|\,\mathbf{d},\mathsf{M})$ and $p_k(m\,|\,\mathbf{d},\mathsf{M})$ (2) belonging to different segments $l,k = 1,2,..,L$. When the BACD is repeatedly used then many candidates for one change-point are generated, and thus a logical filtration of these candidates is required [10] not to obtain false change-points.

Problem P1 cannot be generally solved because it is strongly dependent on the application (signal characteristics). The contradictory requirements P2 can be solved using the recursive evaluation of posterior pdf (2). The sliding recursion decreases the computational costs while preserving the high time resolution (because it uses the overlap $N-1$ samples for segment length N samples). The effective solution of problem P3 is included in the marginalization process leading to pdf (2). As pointed out in [2] and [8] the posterior pdf (2) is influenced only by the nominator of Bayesian formula because it is derived under the condition that the given data segment $\mathbf{d}$ is constant [2]. Thus the Bayesian evidence in the denominator of Bayesian formula is constant. But if the pdf (2) is repeatedly used for the segmented signal then the data are not constant. Therefore a modified Bayesian evidence (slightly differing from the Bayesian evidence in [2])

$$BE \propto \frac{\left[\left(D-\mathbf{g}\mathbf{\Phi}\mathbf{g}^{\mathbf{T}}\right)\right]^{-\frac{N-M}{2}}}{\sqrt{\Delta}} \tag{3}$$

has to be used to normalize the posterior pdf (2). Matrix $\mathbf{G}$ is composed of all segment samples without the division into "left" and "right" part. D, Δ, $\mathbf{g}$ are defined as given below equation (2) using $\mathbf{G}$ rather than $\mathbf{G}_{\mathrm{A}}$. The final normalized BACD (BACDN) is then given by

$$\tilde{p}(m\,|\,\mathbf{d},\mathsf{M}) \propto \frac{\left[\left(D-\mathbf{g}_{\mathrm{A}}\mathbf{\Phi}_{\mathbf{A}}\mathbf{g}^{\mathbf{T}}{}_{\mathrm{A}}\right)\right]^{-\frac{N-M}{2}}}{\sqrt{\Delta_{\mathrm{A}}}} \cdot BE^{-1}. \tag{4}$$

This approach overcomes the disadvantage of the classical BACD (2) residing in the impossibility to compare BACD results between different signal segments.

2.3 Recursive evaluation of BACDN

As discussed in the preceding text it is effective to evaluate both functions (2) and (3) recursively. The consequence of the recursion is not only the computational costs reduction [2] but also the algorithm performance improvement. The recursive evaluation of pdf (2) includes two types of update for functions $D, \mathbf{g}_{\mathrm{A}}, \mathbf{\Phi}_{\mathrm{A}}, \Delta_{\mathrm{A}}$: the update of the change-point position m and the update given a new data [2], [7]. While the position update is used here without any change the data update is modified. Instead of the growing memory algorithm used in [2] and [7] the sliding window algorithm is used here. The sliding window algorithm enables to keep the same amount of data and at the same time the autocorrelation matrix $\left(\mathbf{G}_{\mathrm{A}}^{\mathbf{T}}\mathbf{G}_{\mathrm{A}}\right)$ converges from the form used in the "autocorrelation method" to the form used in "covariance method" [9]. This means that the matrix $\left(\mathbf{G}_{\mathrm{A}}^{\mathbf{T}}\mathbf{G}_{\mathrm{A}}\right)$ is transformed from the symmetric and Toeplitz matrix to the symmetric one only. Simulations confirmed that the BACDN with the symmetric correlation matrix gives better results than the BACDN with the symmetric and Toeplitz matrix. Thus the performance of the BACDN improves during a small number of update steps. The recursion of the evidence (3) includes the update for a new data only. The update of the change-point is not needed because the matrix $\mathbf{G}$ has not the Jordan form. Thus the sliding window algorithm for (3) is simpler than for (2).

2.4 Further simplification of RBACDN

Let $\tilde{p}_l(m\,|\,\mathbf{d},\mathsf{M}), l = 1,...,L,\ m = 1,...,N$ represents the sliding window evaluation of the normalized posterior density (4) for L successive segments giving NxL values. This great number of values can be further reduced. Instead of using all NxL samples of function $\tilde{p}_l(m\,|\,\mathbf{d},\mathsf{M})$ the L values can be used only. The information needed for the localization of a change-point is included in time-progress of any function $\tilde{p}_l(m\,|\,\mathbf{d},\mathsf{M}),\ m = \text{const}$. The best choice is $m = N/2$ giving the sequence $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M}), l = 1,...,L$ and ensuring the best numerical stability and the lowest sensitivity to disturbances comparing with other asymmetric choices $m \neq N/2$. It also gives the best results for the model order selection due to fact that "left" and "right" parts of data vector have the same lengths. The evaluation of (4) giving one value of pdf (4) for each segment ($\tilde{p}_l(m\,|\,\mathbf{d},\mathsf{M}),\ l = 1,...,L,\ m = konst.$) can be also seen as the construction of the longitudinal profile of the "pdgram" (like *spectrogram*) created by plotting N values of $\tilde{p}_l(m\,|\,\mathbf{d},\mathsf{M}),\ l = \text{const.}$, against the segment indexes $l = 1,...,L$. The profile (at $m = N/2$) of the pdgram is used for the change-point localization then.

2.5 Summary of RBACDN algorithm

The RBACDN algorithm can be summarized as follows:

I. Initialization of $D, \mathbf{\Phi}_{\mathbf{A}}, \mathbf{\Phi}, \mathbf{g}, \mathbf{g}_{\mathbf{A}}, \Delta_{\mathbf{A}}, \Delta$ for $l = 1$ (first data segment) and $m = N/2 \rightarrow \tilde{p}_1(N/2\,|\,\mathbf{d},\mathsf{M})$

II. Recursion=data and position updates
   - Data update for a new sample for (2), (3)
   - Removing old sample for (2), (3)
   - Position update for (2)

→ pdgram profile $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M}), l = 2,...$

III. Post-processing of $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M}), l = 2,...$
   - Searching for two adjacent minima of smoothed $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$ (low-pass filter, cut-off $\approx \pi/100$) → stationary points.
   - Searching for one maximum of $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$ between two stationary points→ final change-points.

Notes on the algorithm: Searching for minima is relatively simple because of their comparable values (close to zero) due to normalization of the BACD by the Bayesian evidence. It can be shown that the minima represent stationary parts of signal. Thus the localization of minima is more robust than the direct localization of maxima. Searching for one maximum of unsmoothed $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$ between two minima decreases the number of false alarms and reduces errors. Using the recursive Bayesian evidence also enables to implement an automatic model order selection. In the algorithm implementation the logarithm of equation (4) is used from the numerical stability reasons. The logarithm and the recursion allow using longer window lengths (even 3000 samples) without any numerical instability.

Details of step II

Data update for function (2) are given by

1. adding new data row $\mathbf{x}$

$$\mathbf{x} = [\underbrace{0 \;\cdots\; 0}_{M_1} \quad d[N] \quad d[N-1] \quad \cdots \quad d[N+1-M_2]]$$

$$\hat{D}_l = D_l + d[N+1]\,d[N+1];\;\; \hat{\mathbf{g}}_{\mathbf{A},l} = \mathbf{g}_{\mathbf{A},l} + d[N+1]\,\mathbf{x};$$

$$\mathbf{W}_\mathbf{A} = \mathbf{\Phi}_{\mathbf{A},l}\mathbf{x}^\mathbf{T};\;\; \lambda = 1 + \mathbf{x}\mathbf{W}_\mathbf{A};\;\; \hat{\Delta}_{\mathbf{A},l} = \lambda\,\Delta_{\mathbf{A},l};$$

$$\hat{\mathbf{\Phi}}_{\mathbf{A},l} = \mathbf{\Phi}_{\mathbf{A},l} - \mathbf{W}_\mathbf{A}\mathbf{W}_\mathrm{A}^\mathbf{T}/\lambda$$

2. removing old data row $\mathbf{z}$

$$\mathbf{z} = [d[l-1] \quad d[l-2] \quad \cdots \quad d[l-M_1] \quad \underbrace{0 \;\cdots\; 0}_{M_2}]$$

$$D_{l+1} = \hat{D}_l - d[l]\,d[l];\;\; \breve{\mathbf{g}}_{\mathbf{A},l} = \hat{\mathbf{g}}_{\mathbf{A},l} - d[1]\,\mathbf{z};$$

$$\mathbf{W}_\mathbf{A} = \hat{\mathbf{\Phi}}_{\mathbf{A},l}\,\mathbf{z}^\mathbf{T};\;\; \lambda = 1 - \mathbf{z}\mathbf{W}_\mathbf{A};\; \breve{\Delta}_{\mathbf{A},l} = \lambda\,\hat{\Delta}_{\mathbf{A},l};$$

$$\breve{\mathbf{\Phi}}_{\mathbf{A},l} = \hat{\mathbf{\Phi}}_{\mathbf{A},l} + \mathbf{W}_\mathbf{A}\mathbf{W}_\mathrm{A}^\mathbf{T}/\lambda$$

Note: similar updates hold also for equation (3) using vectors $\mathbf{x}$ and $\mathbf{z}$ defined as

$$\mathbf{x} = [d[N] \quad d[N-1] \quad \cdots \quad d[N+1-M_2]]$$

$$\mathbf{z} = [d[l-1] \quad d[l-2] \quad \cdots \quad d[l-M_1]]$$

Position update for (2) and $m = N/2$

1. replacing of $m+1$ row $\mathbf{r}_A$ of $\mathbf{G}_A$ with row of zeros

$$\mathbf{r}_\mathbf{A} = [\underbrace{0 \;\cdots\; 0}_{M_1} \quad d[m] \quad d[m-1] \quad \cdots \quad d[m+1-M_2]]$$

$$\hat{\mathbf{g}}_{\mathbf{A},l} = \mathbf{g}_{\mathbf{A},l} - d[m+1]\,\mathbf{r}_\mathbf{A};\;\; \mathbf{W}_\mathbf{A} = \mathbf{\Phi}_{\mathbf{A},l}\mathbf{r}_\mathbf{A}^\mathbf{T};$$

$$\lambda = 1 - \mathbf{r}_\mathrm{A}\mathbf{W}_\mathbf{A};\;\; \hat{\Delta}_{\mathbf{A},l} = \lambda\,\Delta_{\mathbf{A},l};$$

$$\overline{\mathbf{\Phi}}_{\mathbf{A},l} = \breve{\mathbf{\Phi}}_{\mathbf{A},l} + \mathbf{W}_\mathbf{A}\mathbf{W}_\mathrm{A}^\mathbf{T}/\lambda$$

2. replacing of $m+1$ row $\mathbf{r}_A$ of $\mathbf{G}_A$ with new data $\mathbf{q}_A$

$$\mathbf{q}_\mathbf{A} = [d[m] \quad d[m-1] \quad \cdots \quad d[m+1-M_1] \quad \underbrace{0 \;\cdots\; 0}_{M_2}]$$

$$\mathbf{g}_{\mathbf{A},l+1} = \overline{\mathbf{g}}_{\mathbf{A},l} + d[m+1]\,\mathbf{q}_\mathbf{A};$$

$$\mathbf{W}_\mathbf{A} = \overline{\mathbf{\Phi}}_{\mathbf{A},l}\mathbf{q}_\mathbf{A}^\mathbf{T};\; \lambda = 1 + \mathbf{q}_\mathbf{A}\mathbf{W}_\mathbf{A};\; \Delta_{\mathbf{A},l+1} = \lambda\,\overline{\Delta}_{\mathbf{A},l};$$

$$\mathbf{\Phi}_{\mathbf{A},l+1} = \overline{\mathbf{\Phi}}_{\mathbf{A},l} - \mathbf{W}_\mathbf{A}\mathbf{W}_\mathrm{A}^\mathbf{T}/\lambda$$

3. EXPERIMENTS ANS RESULTS

To validate the RBACDN algorithm the Monte Carlo simulations were used. Various model orders, parameter sets and window lengths were tested. Simulations confirmed that the performances of the RBACDN and the BACD for the detection of one change-point are comparable.

For example: let us consider the signal composed of two 8-order AR processes with equal lengths 4000 samples giving the small change (cepstral distance [11] is about 3 dB - these changes are almost indistinguishable for acoustic signals). Then the RBACDN with window lengths greater than 200 samples gives the average error $|\hat{m} - m| = 5$ samples and standard deviation (std) 20 samples. These values are influenced by the improper change-point estimations falling outside the sliding window. Excluding these errors from the evaluation the RBACDN accuracy is comparable with the BACD accuracy ($|\hat{m} - m| = 3$, std=18).

Simulations show relatively low sensitivity of the RBACDN to the model order selection and the high sensitivity to the window length and signal-to-noise ratio. Used window should be longer than 200 samples, otherwise inconsistent results are obtained. For any window longer than 400 samples the RBACDN outperforms the BACD owing to the symmetric matrix $(\mathbf{G}_\mathrm{A}^\mathbf{T}\mathbf{G}_\mathrm{A})$. The RBACDN detects changes reliable when the cepstral distance greater than 2 dB.

The similarity of the RBACDN and BACD behaviour illustrates figure 1. It can be seen that the MAP of the BACD is at the same position as the maximum of $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$. Also shapes of both functions are similar.

Figure 2 shows what happens if the BACD is used for signal with three change-points. This case occurs for real non-stationary signals if the window length is greater than the distance between change-points. While the BACD gives one MAP at index 1400 (or two maxima at 1000 and 1400) the RBACDN generates function $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$ with three sharp maxima. When the right part of sliding window reaches the discontinuity in signal (at index 1000) then the left part of the window generates small discontinuity in $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$ (see index 800). But the suggested post-processing (minima searching on smoothed $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathsf{M})$) suppresses this small discontinuity. Thus the RBACDN works properly and reveals all three change-points.

Next example of segmentation using the RBACDN for real music signal is shown in figure 3 (window length 2000 samples; model orders 8). Because of the continual sliding bow movement on strings the signal does not contain any pauses, signal changes are weak (cepstral distances between segments are in the range from 1.8 to 5 dB) and gradual. Extracted segments bordered by change-points were validated by the multiple model algorithm [12], by the inspection of segmented spectrogram (see fig. 3 bottom), and by listening the isolated segments containing tones.

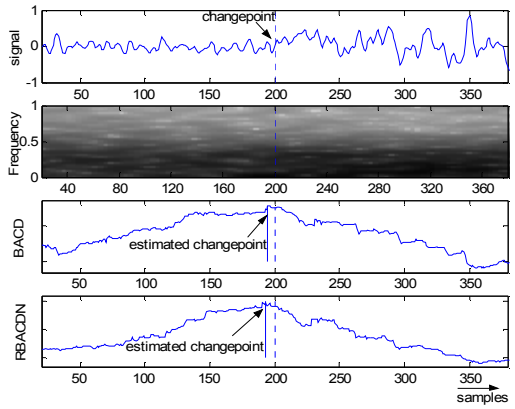

Figure 1: AR signal with one change-point. From top to bottom: signal and spectrogram with true change-point m (full line); BACD output (2) and RBACDN output $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathrm{M})$ with estimated change-point $\hat{m}$ (dashed line).

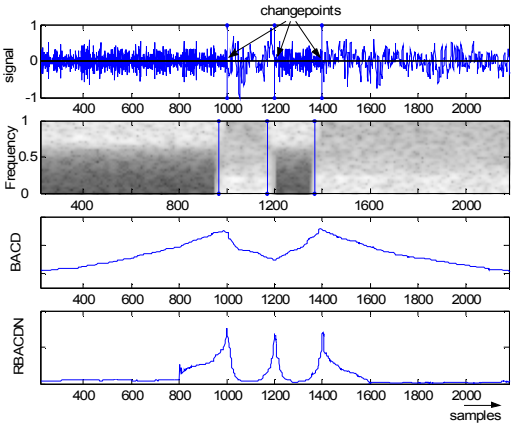

Figure 2: AR signal with three change-points. From top to bottom: signal and spectrogram with estimated change-points, BACD output (2), RBACDN output $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathrm{M})$.

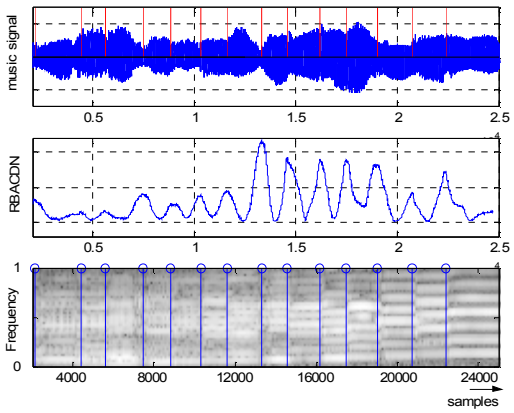

Figure 3: Violin tones separation using RBACDN. From top to bottom: signal with estimated change-points, RBACD output $\tilde{p}_l(N/2\,|\,\mathbf{d},\mathrm{M})$ and segmented spectrogram.

4. CONCLUSIONS

The sliding window autoregressive change-point detector was suggested and its properties were discussed and illustrated on synthetic and real signals. Further research will be focused on the implementation of the model order selection into the suggested algorithm.

5. ACKNOWLEDGEMENT

This work has been supported by the research program Transdisciplinary Research in Biomedical Engineering MSM210000012 of the Czech University in Prague. Authors are also grateful to Mr. M. Strupl for his valuable comments.

REFERENCES

[1] F. Gustafsson, *Adaptive filtering and change detection.* J. Wiley New York, 2000.

[2] J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian methods applied to signal processing.* Springer-Verlag New York, 1996

[3] A. Procházka, J. Uhlíř, J., P.J.W. Rayner, N.G. Kingsbury (eds.), *Signal Analysis and Prediction.* Birkhauser, Boston, 1998.

[4] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination", *Biometrika*, vol. 82, pp. 711-732, 1995

[5] J-Y. Tourneret, M Doisy, and M. Lavielle, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise; application to SAR image edge detection," *Signal Processing*, vol. 83, pp. 1871–1887, 2003.

[6] E.Punskaya, C. Andrieu, A. Doucet, and W. J. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 747–758, Mar. 2002.

[7] J.J.K.O'Ruanaidh, W.J.Fitzgerald and K.J.Pope, Recursive Bayesian location of a discontinuity in time series, in *Proc. International Conference on Acoustics, Speech and Signal Processing*, Adelaide, Australia, 1994.

[8] S. J. Godsill and J. W. Rayner, *Digital audio restoration.* Springer-Verlag New York, 1998

[9] S. M. Kay, ands. L. Marple, "Spectrum Analysis – A Modern Perspective," *Proceedings of the IEEE,* vol. 69, pp. 1380–1419, Nov. 1981

[10] R. Cmejla, and P. Sovka: Estimation of boundaries between speech units using Bayesian change-point detectors, in *Proc. International Conference on Text Speech and Dialog 2001 in series Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, Germany, 2001

[11] R. Cmejla, R. and P. Sovka: "Application of Bayesian Detector in Speech Processing" in *Signal and Image Processing* (*SIP'99*), p. 76-80, IASTED Calgary, 1999.

[12] P. Anderson, "Adaptive forgetting in recursive identification through multiple models," *International Journal of Control*, vol. 42, pp. 1175– 1193, 1985.