International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:9, No:6, 2015

Adaptive Nonparametric Approach for Guaranteed Real-Time Detection of Targeted Signals in Multichannel Monitoring Systems

Andrey V. Timofeev

***Abstract*—**An adaptive nonparametric method is proposed for stable real-time detection of seismoacoustic sources in multichannel C-OTDR systems with a significant number of channels. This method guarantees given upper boundaries for probabilities of Type I and Type II errors. Properties of the proposed method are rigorously proved. The results of practical applications of the proposed method in a real C-OTDR-system are presented in this report.

***Keywords*—**Adaptive detection, change point, interval estimation, guaranteed detection, multichannel monitoring systems.

I. Introduction

MONITORING systems for control of super-extended objects (oil & gas pipelines, railways, national borders) are always multichannel complexes without exception. It is true for brand new C-OTDR monitoring systems [1]-[4] as well as for convenient large-scale monitoring multi-point sensor systems (seismic braids etc.). At different times and different places the conditions of observation are dramatically different. These circumstances influence sensors systems very strongly. The influence implies an increase in Type I and Type II errors. While noise level may be dramatically different in various time intervals for one and the same channel, the industrial noise power and its spectral characteristics as a rule are stable for extended periods of time (no less than a few minutes or sometimes even hours). Unlike the industrial noises, targeted signals have high power and short duration (no more than a few minutes). So, targeted signals have a shorter stability period with respect to stability period of noises. Using this assumption we can build an adaptive real-time detector which will guarantee prescribed level for Type I and Type II errors. That approach will be described in this paper on an example of C-OTDR monitoring system with a high number of channels (more than 20,000).

II. Statement of the Problem

Let us assume that we have a multichannel monitoring system. There are array of statistically independent channels, which are used for getting targeted signals. Indexes of system channels in conjunction form a set **Z**= {1,2,…}. Observations are made at successive times, which form a set $T=\{t_1,t_2,...\},\forall i>0:t_{i+1}-t_i=\Delta t>0$ thus, the observations are form the following sets $\mathbf{S}_j=\{S_j(t)|t\in T\}$.

Let us denote:

1) τ_j is random moment time, $\tau_j\in T$. So, τ_j is the moment of abrupt change of the observations distributions in j-th channel. This change happened due signal appearance. Actually, τ_j is the **change-point moment** [5], [6] of observation distributions**.**
2) t_0 is time of observation start;
3) h is the sample size for adaptation to noise (this parameter is selected a priori);
4) P_c - confidence coefficient, $0<P_c<1$;
5) $\Delta(h)=2\left(\left(1-P_c\right)h\right)^{-0.5}$;
6) hypothesis H_0: in channel do not signal (background model);
7) hypothesis H_1: in channel is signal (signaling model);
8) $\alpha\in]0,1[$ is a predetermined upper bound for the probability of making type I errors;
9) $\beta\in]0,1[$ is a predetermined upper bound for the probability of making type II errors.

For each channel $j\in\mathbf{Z}$ observations are described by:

$$\forall t<\tau_j,t\in\Delta:S_j(t)=\theta_j+\sigma_j(t)\varsigma_j(t)$$

$$\forall t\geq\tau_j,t\in\Delta:S_j(t)=\theta_j+\sigma_j(t)\varsigma_j(t)+\theta_{s,j}(t)+\pi_j(t)\xi_j(t).$$

where $\{\varsigma_j(t)\},\{\xi_{s,j}(t)\}$ are mutually independent random variables, $\mathbf{E}\varsigma_j(t)=0,\mathbf{E}\varsigma_j^2(t)=1,\mathbf{E}\xi_{s,j}(t)=0,\mathbf{E}\xi_j^2(t)=1,$

$$\sigma_j\leq L_N,\pi_j\leq L_S.$$

$$\forall_{i\neq j}\theta_i\neq\theta_j,\Xi_{s,j}(t)=\theta_{s,j}(t)+\pi_j(t)\xi_j(t)$$

is equation of target signal in j-*th* channel, $\theta_{s,j}(t)>0$. The constants L_N,L_S are given; noise parameters $\{\theta_j\}$ are unknown a priori.

The research objective is to build the **signals detection procedure**, which will be guarantee prescribed level for Type

Andrey V. Timofeev is with the LPP "Equali Zoom", Astana, 010000, Kazakhstan(phone: +7-911-191-42-67; e-mail: timofeev.andrey@gmail.com).

I (α) and Type II (β) errors. In solving this problem for each channel we can use observations of this channel only. So, we have not possibility for use the cross-channel information, in contrast to [7]. This is due presence of huge number of channels (more than 20,000), data from which we have to process in real time.

III. Solution Method

Let us consider the following simple statistic:

$$\overset{=}{\theta}_j\left(t_0,h\right)=\sum_{p=t_0}^{t_0+h}S_j\left(p\right)/\left(L_N h\right)+\left(\left(1-P_c\right)h\right)^{-0.5}.$$

The $\overset{=}{\theta}_j\left(t,h\right)$ is non-parametric confidence upper bound for θ_j, and it is easy to see that $\mathbf{P}\left(\theta_j\le\overset{=}{\theta}_j\left(t_0,h\right)\right)\ge P_c$ and $\underset{j}{\forall}P\left(\lim\limits_{h\to\infty}\overset{=}{\theta}_j\left(t_0,h\right)\to\theta_j\right)=1$. So, interval $\left(t_0,t_0+h\right)$ is called initial interval adaptation to noise (IIAN); calculation of $\overset{=}{\theta}_j\left(t,h\right)$ we will call as "adaptation to noise" (AN-procedure).

Let us consider following cycle statistics:

$$Y_j(t\mid h,H)=\left(H\left(\Delta(h)+\varepsilon\right)\right)^{-1}\left(\sum_{k=t-z}^{t-1}\frac{S_j(k)}{L_N^2+L_S^2}\right)-\frac{\overset{=}{\theta}_j\left(t_0,h\right)}{\Delta(h)+\varepsilon}+\frac{\alpha' S_j(t)}{(\Delta(h)+\varepsilon)H}.$$

$$z=\inf\left\{t\ge t_0+h\mid t\ge H\left(L_N^2+L_S^2\right)\right\},$$

$$\alpha'=H-(z-1)/\left(L_N^2+L_S^2\right),\varepsilon>0.$$

Statistics $Y_j(t\mid h,H)$ are defined on the sequence of the intervals:

$$U(z,\delta_0,\delta_1,...)=\left(u(t_i,z,\delta_i)\mid i\ge 1,u(t_i,z,\delta_i)=(t_i,t_i+z+\delta_i),\right.$$
$$\left.t_i=t_{i-1}+z+\delta_{i-1},\delta_i\ge 0\right).$$

Those cycle statistics $Y_j(t\mid h,H)$ will be used for guarantee detection of signals by reducing the task of detection signals to the task of the moment τ interval estimation. As confidence interval we will consider any interval $u(t_i,z,\delta_i)$ from sequence $U(z,\delta_0,\delta_1,...)$. Once $Y_j(t\mid h,H)$ is calculated, we have to decide what is right: $\tau\in u(t_i,z,\delta_i)$ or $\tau\notin u(t_i,z,\delta_i)$. Simply put, the accuracy of moment τ estimation is z. The conditions which guarantee the prescribed quality of the algorithm are determined by the following theorem:

Theorem 1. Let

1. $1-P_c<\alpha,1-P_c<\beta$.
2. $\underset{j}{\forall}\underset{t\ge\tau}{Inf}\,\theta_{s,j}(t)\ge 2\Delta(h)+\varepsilon$.
3. $H=P_c\left[\left(\beta-1+P_c\right)\left(1-b\right)^2\left(\Delta(h)+\varepsilon\right)^2\right]^{-1}$, *where*

$$b=\left(\frac{cm+1}{c+1}\right),c=\left(\frac{\alpha-1+P_c}{\beta-1+P_c}\right)^{0.5},m=\frac{\Delta(h)}{\Delta(h)+\varepsilon}.$$

In this case, if the decision rule will be defined by the following way

$$R(b)=\begin{cases}if\ \ Y_j(t\mid h,H)\ge b\ \text{then}\ H_1\ \text{is true in j-th channel}\\ if\ \ Y_j(t\mid h,H)<\text{b}\ \ \text{then}\ H_0\ \text{is true in j-th channel}\end{cases},$$

then next inequalities will be true for prescribed α и β:

1. $\mathbf{P}\left(Y_j(t_i+z+\delta_i\mid h,H)<b\mid H_1:\tau\in u(t_i,z,\delta_i)\right)\le\alpha$
2. $\mathbf{P}\left(Y_j(t_i+z+\delta_i\mid h,H)\ge b\mid H_0:\tau\notin u(t_i,z,\delta_i)\right)\le\beta$.

Poof of Theorem 1. Let us consider the next representation

$$Y_j\left(t-z,H\right)=\begin{cases}\rho_j(1)+m_j(t,H),\ t<\tau\\ \rho_j(2)+m'_j(t,H),\ t-z\ge\tau\end{cases}$$

$$\rho_j(1)=\left(\theta-\overset{=}{\theta}_j\left(t_0,h\right)\right)\left(\Delta(h)+\varepsilon\right)^{-1},$$

$$\rho_j(2)\ge\left(\phi_j+\theta-\overset{=}{\theta}_j\left(t_0,h\right)\right)\left(\Delta(h)+\varepsilon\right)^{-1}.$$

It is obvious that

$$\mathbf{P}\left(\left|\rho_j(1)\right|<\Delta(h)/\left(\Delta(h)+\varepsilon\right)\right)\ge P_c,\tag{1}$$

$$\mathbf{P}\left(\left|\rho_j(2)\right|>1\right)\ge P_c\tag{2}$$

$$m_j(t,H)=\left(H\left(\Delta(h)+\varepsilon\right)\right)^{-1}\left[\sum_{k=t-z}^{t-1}\left(\frac{\sigma_j(k)\varsigma_j(k)}{L_N^2+L_S^2}\right)+\sigma_j(k)\varsigma_j(k)\alpha'\right]$$

$$m'_j(t,H)=\left(H\left(\Delta(h)+\varepsilon\right)\right)^{-1}\left[\sum_{k=t-z}^{t-1}\left(\frac{\sigma_j(k)\varsigma_j(k)+\pi_j(t)\xi_j(t)}{L_N^2+L_S^2}\right)+\right.$$
$$\left.+\left(\sigma_j(k)\varsigma_j(k)+\pi_j(t)\xi_j(t)\right)\alpha'\right]$$

Easy to see

$$\mathbf{Var}\left(m_j(t,H)\mid t_i+z+\delta_i<\tau\right)<$$
$$\mathbf{Var}\left(m_j'(t,H)\mid t_i\ge\tau\right)\le\left(H\left(\Delta(h)+\varepsilon\right)^2\right)^{-1}.$$

It is obvious that

$$\mathbf{P}\left(Y_j(t_i+z+\delta_i\mid h,H)>b\mid H_0:\tau\notin u(t_i,z,\delta_i)\right)=$$
$$\mathbf{P}\left(\left|\rho_j(1)+m_j(t,H)\right|>b\left|t_i+z+\delta_i<\tau\right.\right)\le$$

$$\mathbf{P}\left(\left|\rho_j(1)\right| + \left|m_j(t,H)\right| > b \middle| t_i + z + \delta_i < \tau\right) =$$
$$\mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\cdot$$

Further, taking into account (1), we have

$$\left(\mathbf{P}\left(\mathbf{P}\left(\left|\rho_j(1)\right| \geq \mathbf{m}\right) + \left|\rho_j(1)\right| < \mathbf{m}\right)\right) =$$
$$\mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\mathbf{P}\left(\left|\rho_j(1)\right| \geq \mathbf{m}\right) +$$
$$\mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\mathbf{P}\left(\left|\rho_j(1)\right| < \mathbf{m}\right)$$
$$\leq \mathbf{P}\left(\left|\rho_j(1)\right| \geq \mathbf{m}\right) +$$
$$\mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\mathbf{P}\left(\left|\rho_j(1)\right| < \mathbf{m}\right) =$$
$$\mathbf{P}\left(\left|\rho_j(1)\right| \geq \Delta(h)\left(\Delta(h)+\varepsilon\right)^{-1}\right) +$$
$$\mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\mathbf{P}\left(\left|\rho_j(1)\right| < \mathbf{m}\right) \leq$$
$$1 - P_c + \mathbf{P}\left(\left|m_j(t,H)\right| > b - \left|\rho_j(1)\right| \middle| t_i + z + \delta_i < \tau\right)\cdot$$
$$\mathbf{P}\left(\left|\rho_j(1)\right| < \mathbf{m}\right) \leq$$
$$1 - P_c + P_c / \left(H\left(\Delta(h)+\varepsilon\right)^2 \left(b - \mathbf{m}\right)^2\right) \tag{3}$$

On the other hand, taking into account (A.2), it easy to see

$$\mathbf{P}\left(Y_j(t_i + z + \delta_i \mid h, H) < b \mid H_1 : \tau \in u(t_i, z, \delta_i)\right) =$$
$$\mathbf{P}\left(\left|\rho_j(2) + m'_j(t,H)\right| < b \middle| t_i \geq \tau\right) \leq$$

$$\mathbf{P}\left(\left|\rho_j(2)\right| - \left|m'_j(t,H)\right| < b \middle| t_i \geq \tau\right) =$$
$$\mathbf{P}\left(-\left|m'_j(t,H)\right| < b - \left|\rho_j(2)\right| \middle| t_i \geq \tau\right) =$$
$$\mathbf{P}\left(\left|m'_j(t,H)\right| \geq \left|\rho_j(2)\right| - b \middle| t_i \geq \tau\right) =$$
$$\mathbf{P}\left(\left|m'_j(t,H)\right| \geq \left|\rho_j(2)\right| - b \middle| t_i \geq \tau\right)\cdot$$
$$\mathbf{P}\left(\left|\rho_j(2)\right| \geq 1\right) + \mathbf{P}\left(\left|m'_j(t,H)\right| \geq \left|\rho_j(2)\right| - b \middle| t_i \geq \tau\right)\cdot$$
$$\mathbf{P}\left(\left|\rho_j(2)\right| < 1\right) \leq$$

$$\mathbf{P}\left(\left|m'_j(t,H)\right| \geq \left|\rho_j(2)\right| - b \middle| t_i \geq \tau\right)\cdot$$
$$\mathbf{P}\left(\left|\rho_j(2)\right| \geq 1\right) + 1 - P_c \leq$$
$$\mathbf{P}\left(\left|m'_j(t,H)\right| \geq \left|\rho_j(2)\right| - b \middle| t_i \geq \tau\right)P_c + 1 - P_c \leq$$

$$P_c / \left(H\left(\Delta(h)+\varepsilon\right)^2 \left(1 - b\right)^2\right) + 1 - P_c \tag{4}$$

Substituting in (3) and (4) expression for H and b, we immediately obtain the assertion of the theorem ■

Here, the interval $u(t_i, z, \delta_i)$ is simply regular interval from sequence $U(z, \delta_0, \delta_1, \ldots)$. At usage of suggested approach, at first we calculate the cyclic statistic $Y_j(t_i + z + \delta_i \mid h, H_j)$ for the current interval $u(t_i, z, \delta_i)$ than we use the decision rule $R(b)$. We will call this method as adaptive cyclic analysis (ACA-procedure).

IV. Usage of the Suggested Approach in the Real C-OTDR Monitoring System

The approach described in this report is used for the detection of seismoacoustic emission sources (SES) in a real C-OTDR monitoring system. The parameters of this system are: the probe pulse duration - 50 ns; frequency sensing - 3 kHz; update rate of models – 20 Hz; the probe signal power - 15 mW; laser wavelength - 1550 nm .

A. System Description

This system was installed to monitor railways (Astana area, Kazakhstan). The length of the fiber optic sensor (FOS) is 1,200 m. This sensor is buried in the vicinity of real railways (offset is 5 m, depth is 50 cm). The FOS length was divided on 1,200 logical C-OTDR channels, but in the full-scale C-OTDR system there are more than 20,000 channels. Each of those channels generated the stream of primary signals (makers). Probability distributions of those signal streams are subject to the Poisson law with high intensity. That is why detection system must process signal streams very quickly. Fig. 1 shows typical distributions of the signals energy throughout the length of FOS for three frequency bands.

Fig. 2 shows typical distributions of the signal energy for another time period. We can see that these distributions are dramatically different for different time periods. Those differences are related to dynamics of influences of targeted signals and to varying influence of industrial noises. At the same time, our research has demonstrated that these differences were mostly due to influence of dynamical industrial noises.

Fig. 3 shows the targeted signal in mix with background noises. Background noise energy is bit less in comparison with targeted signal energy. In contrast to noise the targeted signal width is considerably more. This feature is very important for detection and classification of targeted signal in C-OTDR monitoring system.

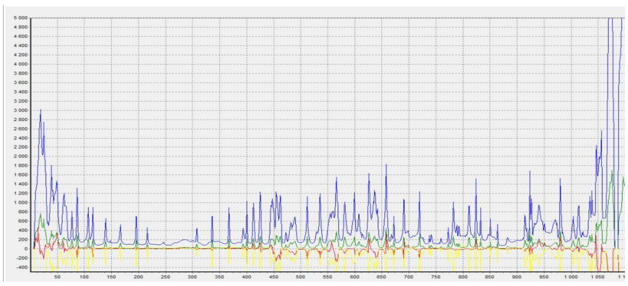

Fig. 1 The distributions of the signal strength throughout the length of FOS Sample 1

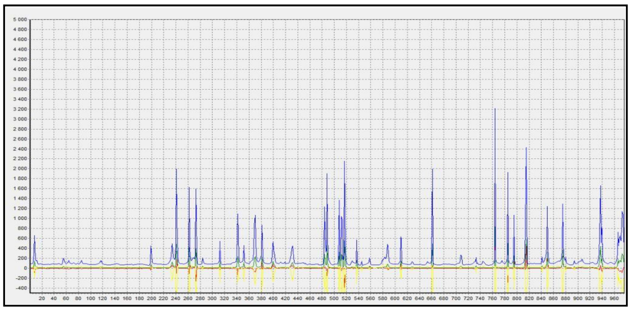

Fig 2 The distributions of the signal strength throughout the length of FOS Sample 2

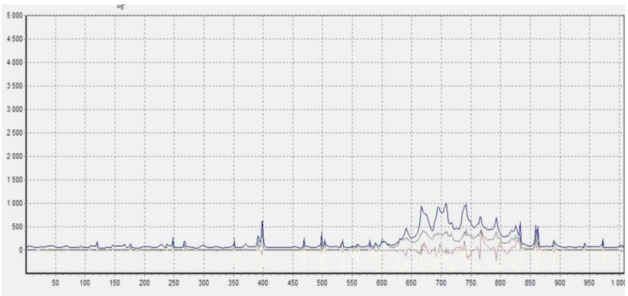

Fig. 3 Targeted Signal with mix with background noises

B. Adapting to the Dynamics of the Background Noises

So, we face a situation when there are dynamical seismoacoustic industrial noises with a large stability period in the vicinity of railways. In this situation the usage of adaptive approach for signal detection proves effective. AN-procedure has been carried out during IIAN $(t_0, t_0 + h)$. After that, the ACA-procedure is started. If for the j-th channel on the next interval $u(t_i, z, \delta_i)$ it is decided that $\tau \notin u(t_i, z, \delta_i)$, then interval $u(t_i, z, \delta_i)$ may be used for adaptation to noise. Intervals like that we will call "interval suitable for adaptation" (ISA).

While the ACA-procedure is carried out we collect ISA-intervals until their total duration does not exceed the value *h*. In this way we form the set of ISA-intervals. Once the total duration of ISA-intervals exceeds the value *h*, we start the AN-procedure on the set of ISA-intervals. Thus we calculate the **new value** of $\overset{=}{\theta}_j(t,h)$. After that we begin to collect a set of ISA-intervals again, repeating the whole cycle. It has proved effective.

Table I contains the results of detection of the SES. Here «Distance» is an average distance at which the given class of SEV was detected, P_I - is a detection error of type I; P_{II} - an error of type II; "Type of Noise" is the type of the background industrial noise (there are two types: "noise of water drain" (type 1), and "noise from the diesel generator" (type 2). Parameters of the detection system were as follows: *h*=200 (since update rate of models is 20 Hz, the adaptation interval length is 10 seconds), $\alpha = 0.2$, $\beta = 0.1$.

Table I shows sufficiently high practical effectiveness of the described SES detection system.

TABLE I
THE PRACTICAL DETECTION RESULTS

Type of SES	Distance (m)	P_I	P_{II}	Type of noise
"hand digging the soil"	10	0.15	0.09	Type 1
"chiselling ground scrap"	5	0,18	0,09	Type 2
"walking man"	10	0,19	0,09	Type 2
"cutting frozen soil"	15	0,14	0,1	Type 1
"train"	20	0,0	0,0	Type 2

ACKNOWLEDGMENT

This investigation has been produced under the project "Development of a remote monitoring system to protect backbone communications infrastructure, oil and gas pipelines and other extended objects (project code name – OXY)", financed under the project "Technology Commercialization ", supported by the World Bank and the Government of the Republic of Kazakhstan.

REFERENCES

[1] K. N. Choi, J. C. Juarez, H. F. Taylor, "Distributed fiber optic pressure/seismic sensor for low-cost monitoring of long perimeters", Proc. SPIE 5090, Unattended Ground Sensor Technologies and Applications, 2003, pp. 134-141.

[2] J. C. Juarez, E. W. Maier, K. N. Choi, and H. F. Taylor, "Distributed Fiber-Optic Intrusion Sensor System", *Journal of Lightwave Technology*, Vol. 23, Issue 6, 2005, pp. 2081-2087.

[3] S. S. Mahmoud, Y. Visagathilagar, J. Katsifolis., "Real-time distributed fiber optic sensor for security systems: Performance, event classification and nuisance mitigation". *Photonic Sensors*, Vol.2, Issue 3, 2012, pp. 225-236.

[4] V. Korotaev, V. M. Denisov, A. V. Timofeev, and M. G. Serikova, "Analysis of seismoacoustic activity based on using optical fiber classifier," in Latin America Optics and Photonics Conference, OSA Technical Digest (online) (Optical Society of America, 2014), paper LM4A.22.

[5] Y. Mei, "Sequential change-point detection when unknown parameters are present in the pre-change distribution", *The Annals of Statistics*, Vol. 34, 2006, pp. 92-122.

[6] T.L. Lai, "Sequential Change point Detection in Quality Control and Dynamical Systems", *Journal of Royal Statistical Society*, Series B, Vol. 57, 1995, pp. 613-658.

[7] Timofeev A.V. The guaranteed detection of the seismoacoustic emission source in the C- OTDR systems, *International Journal of Mathematical, Computational, Physical and Quantum Engineering*, Vol.8, Issue 10, 2014, pp. 1213-1216.

Timofeev Andrey V. was born in Chita (Russia). He received Dr. habil. sc. ing. in Computer and Information Sciences from Tomsk State University of Control Systems and Radioelectronics, Russia, in 1994. A number of research publications in the International journals (JKSS, Stat.Methodology., Automation and Remote Control etc) and International/National conferences are at his credit. He is on the editorial board of several journals and conferences and a referee of several others. His research interests include non-asymptotic nonlinear methods of confidence estimation of multidimensional parameters of stochastic systems; machine learning, large margin classification in Banach spaces; confidence Lipschitz classifiers; technical diagnostics, C-OTDR systems; data mining; change-point problem; alpha-stable laws; statistical classification in application to biometrics and seismics.