DOI 10.1109/TSP.2017.2692740

This is a post-peer-review, pre-copyedit version of an article published in IEEE Transactions on Signal Processing. The final authenticated version is available online at: https://doi.org/10.1109/TSP.2017.2692740

New semi-causal and noncausal techniques for detection of impulsive disturbances in multivariate signals with audio applications

Maciej Niedźwiecki, *Senior Member, IEEE,* and Marcin Ciołek, *Member, IEEE,*

***Abstract*—This paper deals with the problem of localization of impulsive disturbances in nonstationary multivariate signals. Both unidirectional and bidirectional (noncausal) detection schemes are proposed. It is shown that the strengthened pulse detection rule, which combines analysis of one-step-ahead signal prediction errors with critical evaluation of leave-one-out signal interpolation errors, allows one to noticeably improve detection results compared to the prediction-only based solutions. The proposed general purpose approach is illustrated with two examples of practical applications – elimination of impulsive disturbances (such as clicks, pops and record scratches) from archive audio files and robust parametric spectrum estimation.**

I. Introduction

ANALYSIS of multivariate time series plays an important role in many research areas such as medicine, economics, seismology, audio processing etc. In many cases the recorded data are contaminated with short-time impulsive disturbances such as breathing artifacts in polysomnographic recordings (caused by patients rapid body movements during sleep) [1], event-related potentials in multichannel electroencephalographic (EEG) signals [2], or clicks, pops and scratches in stereo archive audio recordings (caused by aging and/or mishandling of the recording medium, e.g. the vinyl or shellac record) [3]. Localization of noise pulses is needed for diagnostic purposes (in the first two cases) or to reconstruct the original, undistorted signal (in the third case).

The signal processing task of localization, and possibly removal, of impulsive disturbances is conceptually related to the problem of detection of outliers, i.e., observations that are inconsistent with the remainder of the analyzed data set, broadly discussed in the statistical literature. Detection of outliers in time series, which can be traced back to the pioneering work of Fox [4], can be carried out using different approaches, both parametric (model-based) and nonparametric (model-free) – see Gupta et al. [5] and references therein.

Most of the parametric methods proposed in the statistical literature were developed for univariate time series and are based on critical examination of residuals generated by autoregressive (AR), autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models of the analyzed signals, with known or unknown parameters [6], [7], [8], [9], [10], [11], [12]. In the latter case model parameters and outliers are estimated jointly.

This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772. Computer simulations were carried out at the Academic Computer Centre in Gdańsk. Both authors are with the Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, ul. Narutowicza 11/12, Gdańsk, Poland, maciekn@eti.pg.gda.pl , marcin.ciolek@pg.gda.pl

Even though outliers in multivariate processes can be handled by applying univariate techniques to the component series, such approach is not the best one as it does not properly take into account the joint dynamics of the series – the advantages of studying outliers directly under a multivariate framework are discussed e.g. in [13], [14] and [3]. The multivariate extension of the univariate methods are presented in [14] and [15]. In [16] detection of multivariate outliers is carried out by testing the judiciously chosen signal projections – the data are projected in the direction that either maximizes or minimizes the kurtosis coefficient of the projections. In [17] a similar task is achieved by applying the independent component analysis (ICA) technique.

In majority of approaches proposed in the statistical literature cited above, additive outliers – constituting the class of disturbances considered in this study – are explicitly or implicitly assumed to be of the Kronecker-delta type, which means that they corrupt isolated samples. This stays in contrast with the signal processing practice where noise pulses are usually made up of a certain number of outliers, forming the anomalous waveforms (called outlier patches in [12]) that corrupt many consecutive signal samples, often measured in tens or even hundreds.

The second limitation is due to the fact that the proposed procedures are iterative. Since only one outlier is detected and removed during an iteration, it may take many iterations to detect multiple outliers, making the approach unsuitable for on-line (real-time) processing. Even when operated in the off-line (batch) processing mode, such sequential detection procedures become impractical when the length of the processed data files is counted in millions of samples (as in the case of EEG or audio recordings), rather than in hundreds of samples (as in typical economic time-series analyzed in the statistical literature). Additionally, in the presence of multiple outliers, sequential detection may suffer from some negative effects known as outlier masking and outlier smearing (or swamping) [12], resulting in the false negative and false positive detection errors, respectively.

To overcome limitations mentioned above, many authors usually propose to evaluate multi-step-ahead signal prediction errors, instead of residual errors, to tackle the detection of impulsive disturbances in univariate signals – see e.g. [18], [19], [20], [21] and references therein. In contrast, there are only a few studies of detection techniques for multivariate signals, such as those presented in the papers [13] and [3], devoted to the removal of clicks from stereo archive gramophone recordings.

In this paper we propose a new general purpose pulse detection technique, applicable to multivariate signals, which is based on joint evaluation of signal prediction errors and

leave-one-out signal interpolation errors. We show that the resulting decision rules, which check both forward consistency of the analyzed signal (consistency of the examined sample with the preceding samples) and its backward consistency (consistency of the examined sample with the succeeding samples), allow one to design detection schemes that are more sensitive to outliers while, at the same time, less prone to raising false detection alarms. The advantages of the new detection technique are demonstrated in two practical cases – elimination of clicks from archive audio signals and robust spectrum estimation.

II. Problem statement

The measured multivariate signal will be denoted by $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$, where $t = \ldots, -1, 0, 1, \ldots$, denotes normalized (dimensionless) discrete time and m denotes the number of signal components (channels).

We will assume that the signal $\mathbf{y}(t)$ can be written down in the form

$$\mathbf{y}(t) = \mathbf{s}(t) + \boldsymbol{\delta}(t) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^{\mathrm{T}}$ denotes the noiseless signal and $\boldsymbol{\delta}(t) = [\delta_1(t), \ldots, \delta_m(t)]^{\mathrm{T}}$ is the noise pulse sample. No specific assumption about the noise pulse sequence will be made except that it is statistically independent of $s(t)$, and that the pulses, varying in length and shape, are sparsely distributed in time. In particular, we will not make an idealistic (and highly impractical) assumption that pulses are Kronecker-delta type, i.e., that each one corrupts only one signal sample.

Denote by $d(t)$ the pulse location function

$$d(t) = \begin{cases} 1 & \text{if} \quad \boldsymbol{\delta}(t) \neq \mathbf{0} \\ 0 & \text{if} \quad \boldsymbol{\delta}(t) = \mathbf{0} \end{cases} . \tag{2}$$

Our goal will be to precisely localize noise pulses, i.e., to obtain a good estimate $\widehat{d}(t)$ of the function $d(t)$.

III. Non-adaptive detection

A. Vector autoregressive processes

Suppose, for the time being, that the signal $\mathbf{s}(t)$ is a zero-mean stationary vector autoregressive (VAR) process of order n with known characteristics [22]. Such a process has two equivalent descriptions: the forward-time (causal) description

$$\mathbf{s}(t) = \sum_{i=1}^{n} \mathbf{A}_i^f \mathbf{s}(t-i) + \boldsymbol{\eta}_f(t), \quad \mathrm{cov}[\boldsymbol{\eta}_f(t)] = \boldsymbol{\rho}_f \tag{3}$$

and the backward-time (anti-causal) one

$$\mathbf{s}(t) = \sum_{i=1}^{n} \mathbf{A}_i^b \mathbf{s}(t+i) + \boldsymbol{\eta}_b(t), \quad \mathrm{cov}[\boldsymbol{\eta}_b(t)] = \boldsymbol{\rho}_b \tag{4}$$

where $\{\boldsymbol{\eta}_f(t)\}$ and $\{\boldsymbol{\eta}_b(t)\}$ denote the vector white noise sequences with covariance matrices $\boldsymbol{\rho}_f$ and $\boldsymbol{\rho}_b$, respectively. Unlike the univariate case ($m = 1$), the $m \times m$ matrices of forward and backward autoregressive coefficients differ, i.e., generally it holds that $\mathbf{A}_i^f \neq \mathbf{A}_i^b$, $i = 1, \ldots, n$. Similarly, $\boldsymbol{\rho}_f \neq \boldsymbol{\rho}_b$.

If the sequence of autocorrelation matrices $\mathbf{R}_i = \mathrm{E}[\mathbf{s}(t)\mathbf{s}^{\mathrm{T}}(t-i)], i = 0, \ldots, n$ of the process $\mathbf{s}(t)$ is known, which is assumed in this section, parameters of the forward/backward VAR models can be obtained by solving the set of Yule-Walker equations given in the form

$$\begin{bmatrix} \mathbf{I} & -\mathbf{A}_1^f & \ldots & -\mathbf{A}_{n-1}^f & -\mathbf{A}_n^f \\ -\mathbf{A}_n^b & -\mathbf{A}_{n-1}^b & \ldots & -\mathbf{A}_1^b & \mathbf{I} \end{bmatrix} \boldsymbol{\mathcal{R}} = \begin{bmatrix} \boldsymbol{\rho}_f & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \ldots & \mathbf{O} & \boldsymbol{\rho}_b \end{bmatrix} \tag{5}$$

where $\mathbf{I}$ and $\mathbf{O}$ denote respectively the identity and null matrices of appropriate dimensions, and $\boldsymbol{\mathcal{R}}$ is the block Toeplitz matrix of the form

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \mathbf{R}_0 & \ldots & \mathbf{R}_n \\ \vdots & \ddots & \vdots \\ \mathbf{R}_n^{\mathrm{T}} & \ldots & \mathbf{R}_0 \end{bmatrix}.$$

The spectral density (matrix) function of a VAR process is given by

$$\begin{aligned} \mathbf{S}(\omega) &= \mathbf{A}_f^{-1}(e^{j\omega}) \boldsymbol{\rho}_f \mathbf{A}_f^{-\mathrm{T}}(e^{-j\omega}) \\ &= \mathbf{A}_b^{-1}(e^{j\omega}) \boldsymbol{\rho}_b \mathbf{A}_b^{-\mathrm{T}}(e^{-j\omega}) \end{aligned} \tag{6}$$

where $j = \sqrt{-1}$, $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency and

$$\mathbf{A}_f(z) = \mathbf{I} - \sum_{i=1}^{n} \mathbf{A}_i^f z^i$$

$$\mathbf{A}_b(z) = \mathbf{I} - \sum_{i=1}^{n} \mathbf{A}_i^b z^i.$$

B. Causal detection

The popular noise pulse detection scheme used in the univariate case is based on monitoring signal prediction errors: detection alarm is raised at the instant $t_0 + 1$, after a no detection period of length n or longer [$\widehat{d}_f(t_0 - n + 1) = \ldots = \widehat{d}_f(t_0) = 0$] if the absolute value of the one-step-ahead prediction error $e_f(t_0 + 1)$ exceeds μ times its standard deviation [18]

$$\widehat{d}_f(t_0+1) = \begin{cases} 1 & \text{if} \quad |e_f(t_0+1|t_0)| > \mu \sigma_f(t_0+1|t_0) \\ 0 & \text{elsewhere} \end{cases} \tag{7}$$

where the multiplier μ is chosen so as to guarantee that

$$P\left(\widehat{d}(t_0+1) = 1 \mid d(t_0+1) = 0\right) = \epsilon \tag{8}$$

and ϵ, $0 < \epsilon \ll 1$ denotes the significance level. Under Gaussian assumptions, for $\epsilon = 0.003$ one obtains $\mu = 3$, which corresponds to the well-known "3-sigma" rule used to detect outliers in Gaussian signals.

The multivariate extension of (7) has the form

$$\widehat{d}_f(t_0+1) = \begin{cases} 1 & \text{if} \quad \alpha_f(t_0+1|t_0) > \mu^2 \\ 0 & \text{elsewhere} \end{cases} \tag{9}$$

where

$$\begin{aligned}\alpha_f(t+1|t) &= \mathbf{e}_f^{\mathrm{T}}(t+1|t)\boldsymbol{\Sigma}_f^{-1}(t+1|t)\mathbf{e}_f(t+1|t)\\ \mathbf{e}_f(t+1|t) &= \mathbf{y}(t+1) - \widehat{\mathbf{y}}_f(t+1|t)\\ \widehat{\mathbf{y}}_f(t+1|t) &= \sum_{i=1}^{n}\mathbf{A}_i^f\mathbf{y}(t-i+1)\end{aligned} \tag{10}$$

and $\boldsymbol{\Sigma}_f(t+1|t) = \boldsymbol{\rho}_f$.

Note that the decision rule (9) reduces down to (7) when $m = 1$. Under Gaussian assumptions it also holds that

$$P\left(\alpha_f(t_0+1|t_0) > \mu^2 \mid d(t_0+1) = 0\right) = \epsilon \tag{11}$$

where μ and ϵ are the same quantities as those appearing in (9). Once the detection alarm is triggered, the test is extended to multi-step-ahead predictions. This can be carried out in two different ways.

1) Open-loop detection (a scalar version of this algorithm was presented in [21]): Assuming that the measurements taken at the instants $t-n+1,\ldots,t$ are undistorted [which means that $\mathbf{y}(t-i) = \mathbf{s}(t-i)$ for $i = 0,\ldots,n-1$], the predicted value of $\mathbf{y}(t_0+k)$ can be computed recursively using the formula

$$\widehat{\mathbf{y}}_f(t+k|t) = \sum_{i=1}^{n}\mathbf{A}_i^f\widehat{\mathbf{y}}_f(t+k-i|t), \quad k = 1,2,\ldots \tag{12}$$

where $\widehat{\mathbf{y}}_f(i|t) = \mathbf{y}(i)$ for $i \le t$. The k-step-ahead predictor (12) is a concatenation of k one-step-ahead predictors – note that the unknown "future" samples appearing on the right hand side of (12) are successfully replaced by their VAR-model based predictions.

Denote by $\mathbf{e}_f^o(t+k|t)$ the k-step-ahead prediction error

$$\mathbf{e}_f^o(t+k|t) = \mathbf{y}(t+k) - \widehat{\mathbf{y}}_f(t+k|t_0).$$

The covariance matrix of the prediction error, further denoted by $\boldsymbol{\Sigma}_f^o(t_0+k|t)$, can be easily evaluated using the following recursive algorithm see Chapter 2 in [22])

$$\begin{aligned}\boldsymbol{\Sigma}_f^o(t+k|t) &= \boldsymbol{\Sigma}_f^o(t+k-1|t) + \mathbf{H}_{k-1}^f\boldsymbol{\rho}_f[\mathbf{H}_{k-1}^f]^{\mathrm{T}}\\ \mathbf{H}_k^f &= \sum_{i=1}^{\min(k,n)}\mathbf{H}_{k-i}^f\mathbf{A}_i^f\end{aligned} \tag{13}$$

$$k = 1,2,\ldots$$

with initial conditions set to $\mathbf{H}_0^f = \mathbf{I}$ and $\boldsymbol{\Sigma}_f^o(t|t) = \mathbf{O}$.

Let

$$\alpha_f^o(t|t_0) = [\mathbf{e}_f^o(t|t_0)]^{\mathrm{T}}[\boldsymbol{\Sigma}_f^o(t|t_0]^{-1}\mathbf{e}_f^o(t|t_0)$$

Detection alarm is raised at the instant t_0+1 if $\alpha_f^o(t_0+1|t_0) > \mu^2$, and it is terminated at the instant $t_0+k_0^o+1$, yielding $\widehat{d}(t_0+1) = \ldots = \widehat{d}(t_0+k_0^o) = 1$ and $\widehat{d}(t_0+k_0^o+1) = \ldots = \widehat{d}(t_0+k_0^o+n) = 0$, if the prediction error at the instant $t_0+k_0^o$ is excessive $\alpha_f^o(t_0+k_0^o|t_0) > \mu^2$ while the next n consecutive prediction errors are sufficiently small

$$\alpha_f^o(t|t_0) \le \mu^2 \tag{14}$$

$$t = t_0+k_0^o+1,\ldots,t_0+k_0^o+n$$

or if k_0^o reaches its maximum allowable value $k_{\max}$ (which plays the role of a "safety valve"). As a result of applying the detection procedure summarized above, one localizes a block of k_0^o corrupted samples $\mathbf{y}(t_0+1),\ldots,\mathbf{y}(t_0+k_0^o)$ preceded and succeeded by at least n samples classified as uncorrupted.

2) Decision-feedback detection (see [3] for a specialized two-dimensional version of this algorithm): While the open-loop scheme detects an entire block of corrupted samples, the decision-feedback scheme approves/rejects samples in a sequential way, one by one. In this case the multi-step-ahead signal prediction and the corresponding covariance matrix of prediction error at the instant t depend on earlier decisions of the outlier detector, i.e., on decisions made prior to t. The decision-feedback algorithm is a robust variant of Kalman filter derived for the state space description of the VAR signal

$$\begin{aligned}\mathbf{x}_f(t+1) &= \boldsymbol{\mathcal{A}}_f\mathbf{x}_f(t) + \boldsymbol{\mathcal{C}}_f\boldsymbol{\eta}_f(t+1)\\ \mathbf{y}(t) &= \boldsymbol{\mathcal{C}}_f^{\mathrm{T}}\mathbf{x}_f(t) + \boldsymbol{\delta}(t)\end{aligned}$$

where

$$\boldsymbol{\mathcal{A}}_f = \begin{bmatrix}\mathbf{A}_1^f & \mathbf{A}_2^f & \ldots & \mathbf{A}_{n-1}^f & \mathbf{A}_n^f\\ \mathbf{I} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O}\\ \mathbf{O} & \mathbf{I} & \ldots & \mathbf{O} & \mathbf{O}\\ \vdots & & \ddots & & \\ \mathbf{O} & \mathbf{O} & \ldots & \mathbf{I} & \mathbf{O}\end{bmatrix}, \quad \boldsymbol{\mathcal{C}}_f = \begin{bmatrix}\mathbf{I}\\ \mathbf{O}\\ \mathbf{O}\\ \vdots\\ \mathbf{O}\end{bmatrix}$$

and $\mathbf{x}_f(t) = [\mathbf{s}^{\mathrm{T}}(t),\ldots,\mathbf{s}^{\mathrm{T}}(t-n+1)]^{\mathrm{T}}$ denotes the state vector. The algorithm, which should be started at the instant t_0+1, can be summarized as follows

$$\begin{aligned}\widehat{\mathbf{x}}_f(t|t-1) &= \boldsymbol{\mathcal{A}}_f\widehat{\mathbf{x}}_f(t-1|t-1)\\ \mathbf{Q}_f(t|t-1) &= \boldsymbol{\mathcal{A}}_f\mathbf{Q}_f(t-1|t-1)\boldsymbol{\mathcal{A}}_f^{\mathrm{T}} + \boldsymbol{\mathcal{C}}_f\boldsymbol{\rho}_f\boldsymbol{\mathcal{C}}_f^{\mathrm{T}}\\ \boldsymbol{\Sigma}_f^d(t|t-1) &= \boldsymbol{\mathcal{C}}_f^{\mathrm{T}}\mathbf{Q}_f(t|t-1)\boldsymbol{\mathcal{C}}_f\\ \mathbf{e}_f^d(t|t-1) &= \mathbf{y}(t) - \boldsymbol{\mathcal{C}}_f^{\mathrm{T}}\widehat{\mathbf{x}}_f(t|t-1)\\ \alpha_f^d(t|t_0) &= [\mathbf{e}_f^d(t|t-1)]^{\mathrm{T}}[\boldsymbol{\Sigma}_f^d(t|t-1)]^{-1}\mathbf{e}_f^d(t|t-1)\end{aligned}$$

$$\text{if } \alpha_f^d(t|t_0) \le \mu^2 \text{ then}$$

$$\begin{aligned}\mathbf{L}_f(t) &= \mathbf{Q}_f(t|t-1)\boldsymbol{\mathcal{C}}_f[\boldsymbol{\Sigma}_f^d(t|t-1)]^{-1}\\ \widehat{\mathbf{x}}_f(t|t) &= \widehat{\mathbf{x}}_f(t|t-1) + \mathbf{L}_f(t)\mathbf{e}_f^d(t|t-1)\\ \mathbf{Q}_f(t|t) &= \mathbf{Q}_f(t|t-1) - \mathbf{L}_f(t)\boldsymbol{\Sigma}_f^d(t|t-1)\mathbf{L}_f^{\mathrm{T}}(t)\end{aligned}$$

$$\text{if } \alpha_f^d(t|t_0) > \mu^2 \text{ then}$$

$$\begin{aligned}\widehat{\mathbf{x}}_f(t|t) &= \widehat{\mathbf{x}}_f(t|t-1)\\ \mathbf{Q}_f(t|t) &= \mathbf{Q}_f(t|t-1)\end{aligned} \tag{15}$$

$$t = t_0+1, t_0+2,\ldots$$

where initial conditions should be set to $\widehat{\mathbf{x}}_f(t_0|t_0) = [\mathbf{y}^{\mathrm{T}}(t_0),\ldots,\mathbf{y}^{\mathrm{T}}(t_0-n+1)]^{\mathrm{T}}$ and $\mathbf{Q}_f(t_0|t_0) = \mathbf{O}$.

Similar to the open-loop scheme, detection alarm started at the instant t_0+1 is terminated at the instant $t_0+k_0^d+1$ if $\alpha_f^d(t_0+k_0^d|t_0) > \mu^2$ and

$$\alpha_f^d(t|t_0) = [\mathbf{e}_f^d(t|t-1)]^{\mathrm{T}}[\boldsymbol{\Sigma}_f^d(t|t-1)]^{-1}\mathbf{e}_f^d(t|t-1) \le \mu^2 \tag{16}$$

$$t = t_0 + k_0^d + 1, \ldots, t_0 + k_0^d + n$$

or if k_0^d reaches $k_{\max}$.

When $\alpha_f^d(t|t_0) > \mu^2$ for $t = t_0+1, \ldots, t_0+k_0^d$, the results yielded by the decision-feedback scheme are identical with those provided by the open-loop scheme, i.e., $k_0^o = k_0^d = k_0$ and $\mathbf{e}_f^d(t|t-1) = \mathbf{e}_f^o(t|t_0)$, $\mathbf{\Sigma}_f^d(t|t-1) = \mathbf{\Sigma}_f^o(t|t_0)$ for $t = t_0+1, \ldots, t_0+k_0$. The results differ when $\alpha_f^d(t|t_0) \leq \mu^2$ for some $t \in [t_0+1, t_0+k_0^d]$. Unlike the open-loop scheme, when $\alpha_f^d(t|t_0) \leq \mu^2$ the sample $\mathbf{y}(t)$ is provisionally approved and incorporated in prediction of succeeding signal samples. However, to avoid negative effects of "accidental approvals", when detection alarm is terminated according to (16), both rejected and provisionally approved samples (if any) are scheduled for reconstruction, i.e., the final form of the detection alarm is, similarly as in the open-loop case, $\widehat{d}(t_0+1) = \cdots = \widehat{d}(t_0+k_0^d) = 1$.

C. Signal reconstruction

Corrupted samples can be replaced with their least squares estimates. When the process is governed by the VAR model (3), such estimates can be obtained from

$$\{\widetilde{\mathbf{y}}_f(t_0+1), \ldots, \widetilde{\mathbf{y}}_f(t_0+k_0)\} = \arg \min_{\mathbf{y}(t_0+1), \ldots, \mathbf{y}(t_0+k_0)} \sum_{l=t_0+1}^{t_0+k_0+n} \| \mathbf{y}(l) - \sum_{i=1}^{n} \mathbf{A}_i^f \mathbf{y}(l-i) \|^2 . \tag{17}$$

Denote by $\mathbf{y}_f^c = [\mathbf{y}^{\mathrm{T}}(t_0+1), \ldots, \mathbf{y}^{\mathrm{T}}(t_0+k_0)]^{\mathrm{T}}$ the vector of corrupted samples and by $\mathbf{y}_f^u = [(\mathbf{y}_f^-)^{\mathrm{T}}, (\mathbf{y}_f^+)^{\mathrm{T}}]^{\mathrm{T}}$ - the vector of neighboring uncorrupted samples made up of n samples preceding the corrupted block

$$\mathbf{y}_f^- = [\mathbf{y}^{\mathrm{T}}(t_0-n+1), \ldots, \mathbf{y}^{\mathrm{T}}(t_0)]^T$$

and n samples succeeding the block

$$\mathbf{y}_f^+ = [\mathbf{y}^{\mathrm{T}}(t_0+k_0+1), \ldots, \mathbf{y}^{\mathrm{T}}(t_0+k_0+n)]^T.$$

According to [23], the closed-form solution to (17) is given by

$$\begin{aligned} \widetilde{\mathbf{y}}_f^c &= [\widetilde{\mathbf{y}}_f^{\mathrm{T}}(t_0+1), \ldots, \widetilde{\mathbf{y}}_f^{\mathrm{T}}(t_0+k_0)]^{\mathrm{T}} \\ &= -\left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \right]^{-1} (\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^u \mathbf{y}_f^u \end{aligned} \tag{18}$$

where $\boldsymbol{\mathcal{B}}_f^c$ is the $(n+k_0) \times k_0$ block matrix obtained after removing the first n block columns and the last n block columns from the $(n+k_0) \times (2n+k_0)$ block matrix

$$\boldsymbol{\mathcal{B}}_f = \begin{bmatrix} -\mathbf{A}_n^f & -\mathbf{A}_{n-1}^f & \ldots & \mathbf{I} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & -\mathbf{A}_n^f & \ldots & -\mathbf{A}_1^f & \mathbf{I} & \ldots & \mathbf{O} & \mathbf{O} \\ \vdots & & & & & & & \\ \mathbf{O} & \mathbf{O} & \ldots & & & & -\mathbf{A}_1^f & \mathbf{I} \end{bmatrix}$$

and $\boldsymbol{\mathcal{B}}_f^u$ is the $(n+k_0) \times 2n$ block matrix obtained after removing from $\boldsymbol{\mathcal{B}}_f$ its k_0 central block columns.

We note that the reconstruction formula, analogous to (18), can be based on the backward-time VAR model (4). In this case the matrix $\boldsymbol{\mathcal{B}}_f$, made up of the matrices of forward autoregressive coefficients $\mathbf{A}_i^f$, should be replaced with the analogously defined matrix $\boldsymbol{\mathcal{B}}_b$ combining matrices of backward coefficients $\mathbf{A}_i^b$. Reconstruction based on the backward-time model yields results identical with those based on the forward-time model, i.e., $\widetilde{\mathbf{y}}_f(t_0+i) = \widetilde{\mathbf{y}}_b(t_0+i)$, $i = 1, \ldots, k_0$.

It can be easily shown that $\widetilde{\mathbf{y}}_f^c$ is an unbiased estimate of $\mathbf{y}_f^c$. Actually, note that it holds

$$\boldsymbol{\mathcal{B}}_f^c \mathbf{y}_f^c + \boldsymbol{\mathcal{B}}_f^u \mathbf{y}_f^u = \boldsymbol{\mathcal{B}}_f \mathbf{y}_f = \boldsymbol{\zeta}_f \tag{19}$$

where

$$\begin{aligned} \mathbf{y}_f &= [\mathbf{y}^{\mathrm{T}}(t_0-n+1), \ldots, \mathbf{y}^{\mathrm{T}}(t_0+k_0+n)]^{\mathrm{T}} \\ \boldsymbol{\zeta}_f &= [\boldsymbol{\eta}^{\mathrm{T}}(t_0+1), \ldots, \boldsymbol{\eta}_f^{\mathrm{T}}(t_0+k_0+n)]^{\mathrm{T}}. \end{aligned}$$

Combining (18) with (19), one obtains

$$\widetilde{\mathbf{y}}_f^c = -\left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \right]^{-1} \left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\zeta}_f - (\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \mathbf{y}_f^c \right]$$

leading to

$$\mathbf{y}_f^c - \widetilde{\mathbf{y}}_f^c = \left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \right]^{-1} (\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\zeta}_f.$$

Consequently, $\mathrm{E}[\mathbf{y}_f^c - \widetilde{\mathbf{y}}_f^c] = \mathbf{0}$ and

$$\begin{aligned} \mathrm{cov}[\widetilde{\mathbf{y}}_f^c] &= \mathrm{E}\left[(\widetilde{\mathbf{y}}_f^c - \mathbf{y}_f^c)(\widetilde{\mathbf{y}}_f^c - \mathbf{y}_f^c)^{\mathrm{T}} \right] \\ &= \left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \right]^{-1} (\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\Lambda}_f \boldsymbol{\mathcal{B}}_f^c \left[(\boldsymbol{\mathcal{B}}_f^c)^{\mathrm{T}} \boldsymbol{\mathcal{B}}_f^c \right]^{-1} \end{aligned} \tag{20}$$

where $\boldsymbol{\Lambda}_f = \mathrm{cov}[\boldsymbol{\zeta}_f] = \mathrm{diag}\{\boldsymbol{\rho}_f, \ldots, \boldsymbol{\rho}_f\}$ denotes the $(k_0 + n) \times (k_0+n)$ block diagonal matrix.

We note that although derivation of (18) and (20) is analogous to that given in [23], the obtained formulas are different (since in the multivariate case the least squares solution, presented above, differs from the maximum likelihood solution, presented in [23]).

D. Semi-causal detection

The causal detection scheme is based on monitoring of signal prediction errors. To reduce the number of false alarms we will introduce an additional condition for triggering detection alarm, based on analysis of signal interpolation errors. Denote by $\widetilde{\mathbf{y}}_f^*(t+1)$ the forward-time VAR model based interpolation of the sample $\mathbf{y}(t+1)$, evaluated in terms of n preceding and n succeeding signal samples. According to (18) the interpolation formula can be written down in the form

$$\begin{aligned} \widetilde{\mathbf{y}}_f^*(t+1) &= -\left[\sum_{i=0}^{n} [\mathbf{A}_i^f]^{\mathrm{T}} \mathbf{A}_i^f \right]^{-1} \sum_{i=0}^{n} [\mathbf{A}_i^f]^{\mathrm{T}} \mathbf{v}_i^f(t+1) \\ \mathbf{v}_i^f(t+1) &= \sum_{\substack{l=0 \\ l \neq i}}^{n} \mathbf{A}_i^f \mathbf{y}(t+i-l+1), \quad i = 0, \ldots, n \end{aligned} \tag{21}$$

where $\mathbf{A}_0^f = \mathbf{I}$.

Denote by

$$\mathbf{e}_f^*(t+1) = \mathbf{y}(t+1) - \widetilde{\mathbf{y}}_f^*(t+1)$$

the interpolation error. The additional condition that will be required to trigger detection alarm at the instant t_0+1 can be stated as follows

$$\beta_f(t_0+1) > \mu^2 \tag{22}$$

where

$$\begin{aligned}\beta_f(t+1) =&[\mathbf{e}_f^*(t+1)]^{\mathrm{T}}[\mathbf{\Sigma}_f^*(t+1)]^{-1}\mathbf{e}_f^*(t+1)\\ &\mathbf{\Sigma}_f^*(t+1) = \boldsymbol{\rho}_f^*\end{aligned} \tag{23}$$

and, according to (20),

$$\begin{aligned}\boldsymbol{\rho}_f^* = &\left[\sum_{i=0}^{n}[\mathbf{A}_i^f]^{\mathrm{T}}\mathbf{A}_i^f\right]^{-1}\\ &\times\left[\sum_{i=0}^{n}[\mathbf{A}_i^f]^{\mathrm{T}}\boldsymbol{\rho}_f\mathbf{A}_i^f\right]\left[\sum_{i=0}^{n}[\mathbf{A}_i^f]^{\mathrm{T}}\mathbf{A}_i^f\right]^{-1}.\end{aligned} \tag{24}$$

The strengthened alarm triggering rule can be defined as follows

$$\widehat{d}_f(t_0+1) = \begin{cases} 1 & \text{if } \alpha_f(t_0+1|t_0) > \mu^2 \\ & \text{and } \beta_f(t_0+1) > \mu^2 \\ 0 & \text{elsewhere}\end{cases}. \tag{25}$$

Since the detection scheme based on (25) incorporates n "future" signal samples $\mathbf{y}(t_0+2),\ldots,\mathbf{y}(t_0+n+1)$, it will be referred to as semi-causal.

Once detection alarm is triggered, both prediction error based and interpolation error based statistics are updated and combined to decide upon the termination point. The alarm started at the instant $t_0{+}1$ is terminated at the instant $t_0{+}k_0^*{+}1$ if

$$\begin{aligned}&\alpha_f(t_0+k_0^*|t_0) > \mu^2\\ &\alpha_f(t_0+k_0^*+i|t_0) \le \mu^2, \quad i=1,\ldots,n\\ &\text{or}\\ &\beta_f(t_0+k_0^*) > \mu^2\\ &\beta_f(t_0+k_0^*+i) \le \mu^2, \quad i=1,\ldots,n\end{aligned} \tag{26}$$

or if $k_0^* = k_{\max}$. Depending on whether the open-loop or decision-feedback detection scheme is used, one should set $\alpha_f(t|t_0)$ in (26) to $\alpha_f^o(t|t_0)$ or to $\alpha_f^d(t|t_0)$, respectively. According to (26), when the prediction alarm lasts longer than the interpolation one, the termination point $t_0{+}k_0^*{+}1$ coincides with the end of the interpolation alarm.

Note that, similar to (14) and (16), the stopping rule (26) introduces latency of $k_{\max}+n$ sampling intervals and therefore the semi-causal detector, based on joint evaluation of prediction errors and leave-one-out signal interpolation errors, does not introduce an extra processing delay. Owing to this, it is suitable for real-time (or, strictly speaking, nearly real-time) processing.

E. Anticausal and semi-anticausal detection

If the analyzed signal is pre-recorded, i.e., its entire history $\mathbf{y}(1),\ldots,\mathbf{y}(N)$ is available, detection of impulsive disturbances can be carried out in reverse time, based on the backward-time model (4). In this case detection process is started at the instant $t=N$, i.e., at the end of the recording, and stopped at $t=1$, so this solution is obviously not suitable for real-time processing. The backward-time decision rules (anticausal and semi-anticausal) can be defined analogously as their forward-time (causal and semi-causal), the only difference being that all quantities such as prediction and interpolation errors, and the corresponding covariance matrices are calculated in terms of the matrices $\mathbf{A}_1^b,\ldots,\mathbf{A}_n^b$ and $\boldsymbol{\rho}_b$ that characterize the backward-time model (4). For example, the backward-time triggering rule, which is the counterpart of (9), can be written down as follows

$$\widehat{d}_b(t_0-1) = \begin{cases} 1 & \text{if} \quad \alpha_b(t_0-1|t_0) > \mu^2 \\ 0 & \text{elsewhere}\end{cases} \tag{27}$$

where

$$\begin{aligned}\alpha_b(t-1|t) &= \mathbf{e}_b^{\mathrm{T}}(t-1|t)\mathbf{\Sigma}_b^{-1}(t-1|t)\mathbf{e}_b(t-1|t)\\ \mathbf{e}_b(t-1|t) &= \mathbf{y}(t-1) - \sum_{i=1}^{n}\mathbf{A}_i^b\mathbf{y}(t+i-1)\end{aligned} \tag{28}$$

and $\mathbf{\Sigma}_b(t-1|t) = \boldsymbol{\rho}_b$.

F. Noncausal detection

Consider any causal or semi-causal detection alarm of the form

$$\widehat{d}_f(t) = 1 \quad \text{for} \quad t \in \left[\underline{t}_f^i, \overline{t}_f^i\right] = T_f^i.$$

We note that due to the adopted stopping rules, each alarm is preceded and succeeded by at least n no-alarm decisions. One should realize that the beginning $\underline{t}_f^i$ of the forward alarm is usually determined more precisely than its end $\overline{t}_f^i$. This effect is caused by the fact that the accuracy of multi-step-ahead signal predictions gradually decreases as the prediction horizon increases, which makes the outlier detector more and more "tolerant", i.e., prone to accept large departures of the signal from its expected (predicted) path.

The opposite happens in the case of anticausal or semi-anticausal detection alarms given by

$$\widehat{d}_b(t) = 1 \quad \text{for} \quad t \in \left[\underline{t}_b^j, \overline{t}_b^j\right] = T_b^j.$$

This time the ending point $\overline{t}_b^j$ of the detection alarm (i.e., its starting point in the reverse-time framework) is determined more precisely than its starting point $\underline{t}_b^j$. Therefore, if the forward and backward alarms at least partially overlap, i.e.

$$T_f^i \cap T_b^j \neq \oslash \tag{29}$$

the combined forward-backward detection alarm can be defined as follows [21]

$$\widehat{d}_{fb}(t) = 1 \quad \text{for} \quad t \in \left[\underline{t}_f^i, \overline{t}_b^j\right] = T_{fb}^{ij}. \tag{30}$$

Similar arguments can be used to justify cancellation of all isolated detection alarms, i.e., all forward/backward alarms that do not obey (29) for some choice of j/i.

IV. Adaptive detection

So far we have been assuming that parameters of the forward/backward VAR models are constant and known. If the analyzed VAR signal is nonstationary, but its parameters slowly vary with time, detection of noise pulses can be based on local signal models obtained via signal identification. Denote by $\widehat{\mathbf{A}}_i^f(t), i=1,\ldots,n$, the local estimates of the forward-time VAR model parameters based on the past and present measurements $\{\mathbf{y}(i), i \le t\}$, and by $\widehat{\mathbf{A}}_i^b(t), i=1,\ldots,n$ – the analogous estimates based on the present and future measurements $\{\mathbf{y}(i), i \ge t\}$ (if available).

A. Estimation of signal parameters

The least squares type estimates of VAR parameters can be easily derived after vectorizing the models (3) and (4). Denote by $\boldsymbol{\alpha}_{ij}^f$ the $1 \times m$ vector made up of the autoregressive coefficients that constitute the j-th row of the matrix $\mathbf{A}_i^f$

$$\mathbf{A}_i^f = \begin{bmatrix} a_{11,i}^f & \dots & a_{1m,i}^f \\ \vdots & & \vdots \\ a_{m1,i}^f & \dots & a_{mm,i}^f \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_{i1}^f \\ \vdots \\ \boldsymbol{\alpha}_{im}^f \end{bmatrix}.$$

Let $\boldsymbol{\theta}_j^f = [\boldsymbol{\alpha}_{1j}^f, \dots, \boldsymbol{\alpha}_{nj}^f]^{\mathrm{T}}$ be the $mn \times 1$ vector of parameters characterizing the j-th channel, and let $\boldsymbol{\varphi}_f(t) = [\mathbf{y}^{\mathrm{T}}(t-1), \dots, \mathbf{y}^{\mathrm{T}}(t-n)]^{\mathrm{T}}$ denote the vector of regression variables (the same for all channels). Note that (3), in the absence of outliers, can be rewritten in the form

$$y_j(t) = \boldsymbol{\varphi}_f^{\mathrm{T}}(t)\boldsymbol{\theta}_j^f + \eta_j^f(t), \quad j = 1, \dots, m \tag{31}$$

where $y_j(t)$ and $\eta_j^f(t)$ are the j-th components of the vectors $\mathbf{y}(t)$ and $\boldsymbol{\eta}_f(t)$, respectively.

Parameters of the j-th channel can be tracked using the method of exponentially weighted least squares (EWLS) – see e.g. [24], [25]

$$\begin{aligned} \varepsilon_j^f(t|t-1) &= y_j(t) - \boldsymbol{\varphi}_f^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_j^f(t-1) \\ \mathbf{k}_f(t) &= \frac{\mathbf{P}_f(t-1)\boldsymbol{\varphi}_f(t)}{\lambda + \boldsymbol{\varphi}_f^{\mathrm{T}}(t)\mathbf{P}_f(t-1)\boldsymbol{\varphi}_f(t)} \\ \widehat{\boldsymbol{\theta}}_j^f(t) &= \widehat{\boldsymbol{\theta}}_j^f(t-1) + \mathbf{k}_f(t)\varepsilon_j^f(t|t-1) \\ \mathbf{P}_f(t) &= \frac{1}{\lambda}\left[\mathbf{I} - \mathbf{k}(t)\boldsymbol{\varphi}_f^{\mathrm{T}}(t)\right]\mathbf{P}_f(t-1) \end{aligned} \tag{32}$$

where λ, $0 < \lambda < 1$, denotes the so-called forgetting constant which determines the steady state value of the effective memory of the estimator: $N_{\mathrm{ef}} = 1/(1-\lambda)$. The value of N_{ef} should be chosen in accordance with the rate of process nonstationarity. Small values of N_{ef} (short memory) make the estimation algorithm "fast" (yielding small tracking bias) but "inaccurate" (yielding large tracking variance), whereas large values of N_{ef} (long memory) make it "slow" but "accurate". The best results are obtained if N_{ef} is selected so as to match the degree of nonstationarity of the identified process, trading off the bias and variance components of the mean square parameter tracking error [24], [25].

To protect the identification algorithm against outliers, parameter estimation should be suspended at the beginning of each detection alarm ($t = t_0 + 1$), and resumed at its end, as soon as the regression vector $\boldsymbol{\varphi}_f(t)$ is free of outliers ($t = t_0 + k_0 + n + 1$).

It should be noted that the gain vector $\mathbf{k}_f(t)$ does not depend on j, i.e., it is the same for all channels. Interestingly, when all channels share the same regression vector $\boldsymbol{\varphi}_f(t)$, the local EWLS estimates $\widehat{\boldsymbol{\theta}}_j^f(t)$, obtained by considering each channel separately, coincide with the global estimate obtained by considering all channels jointly – the fact noticed for the first time in [26] (for LS estimators).

The backward-time EWLS estimates can be obtained in the analogous way as the forward-time ones.

B. Detection scheme

The adaptive variants of the proposed detection rules can be obtained by replacing in the corresponding formulas the true model parameters with their estimates. For example, the adaptive counterparts of (10) and (22) can be expressed in the form

$$\begin{aligned} \alpha_f(t+1|t) &= \boldsymbol{\varepsilon}_f^{\mathrm{T}}(t+1|t)\widehat{\boldsymbol{\Sigma}}_f^{-1}(t+1|t)\boldsymbol{\varepsilon}_f(t+1|t) \\ \beta_f(t+1) &= [\boldsymbol{\varepsilon}_f^*(t+1)]^{\mathrm{T}}[\widehat{\boldsymbol{\Sigma}}_f^*(t+1)]^{-1}\boldsymbol{\varepsilon}_f^*(t+1) \end{aligned}$$

where $\boldsymbol{\varepsilon}_f(t+1|t)$ and $\boldsymbol{\varepsilon}_f^*(t+1|t)$ denote respectively the vectors of prediction and interpolation errors, obtained when the true VAR matrices $\mathbf{A}_i^f$ are replaced in (10) and (21) with their estimates $\widehat{\mathbf{A}}_i^f(t_0)$. The quantity $\widehat{\boldsymbol{\Sigma}}_f(t+1|t)$, which denotes the estimate of the covariance matrix of the one-step-ahead prediction error, can be obtained from

$$\begin{aligned} &\widehat{\boldsymbol{\Sigma}}_f(t+1|t) = \\ &= \begin{cases} \lambda_0\widehat{\boldsymbol{\Sigma}}_f(t|t-1) & \\ \quad +(1-\lambda_0)\boldsymbol{\varepsilon}_f(t|t-1)\boldsymbol{\varepsilon}_f^{\mathrm{T}}(t|t-1) & \text{if } \widehat{d}_f(t) = 0 \\ \widehat{\boldsymbol{\Sigma}}_f(t|t-1) & \text{if } \widehat{d}_f(t) = 1 \end{cases} \end{aligned} \tag{33}$$

where λ_0, $0 < \lambda_0 < 1$ denotes the forgetting constant, generally different from the forgetting constant λ used in the EWLS identification algorithm presented in the previous subsection.

The estimate of the covariance matrix of interpolation errors can be obtained using the following formula, based on (24)

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_f^*(t+1) = &\left[\sum_{i=0}^{n}[\widehat{\mathbf{A}}_i^f(t)]^{\mathrm{T}}\widehat{\mathbf{A}}_i^f(t)\right]^{-1} \\ &\times \left[\sum_{i=0}^{n}[\widehat{\mathbf{A}}_i^f(t)]^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_f(t+1|t)\widehat{\mathbf{A}}_i^f(t)\right] \\ &\times \left[\sum_{i=0}^{n}[\widehat{\mathbf{A}}_i^f(t)]^{\mathrm{T}}\widehat{\mathbf{A}}_i^f(t)\right]^{-1}. \end{aligned} \tag{34}$$

Unlike the prediction error covariance (33), which must be updated in a continuous manner, the interpolation error covariance (34) has to be computed only at the instants t where $\alpha_f(t+1|t) > \mu^2$ (to confirm or cancel the prediction based detection alarm).

Once the detection alarm is triggered at the instant $t_0 + 1$, the multi-step-ahead prediction statistics can be obtained after replacing in (13)–(14) or (15) the true values of signal parameters $\mathbf{A}_i^f, i = 1, \dots, n$ and $\boldsymbol{\rho}_f$ with their estimates $\widehat{\mathbf{A}}_i^f(t_0), i = 1, \dots, n$ and $\widehat{\boldsymbol{\Sigma}}_f(t_0+1)$, respectively. In the semi-causal case, for $t > t_0$ the following formula should be used until detection alarm is terminated

$$\beta_f(t+1) = [\boldsymbol{\varepsilon}_f^*(t+1)]^{\mathrm{T}}[\widehat{\boldsymbol{\Sigma}}_f^*(t_0+1)]^{-1}\boldsymbol{\varepsilon}_f^*(t+1).$$

C. Remarks on solutions based on disturbance models

The detection/reconstruction methods described in this paper refrain from making any assumptions about the sequence of noise pulses $\{\boldsymbol{\delta}(t)\}$, except that it is sparse and independent of $\{\mathbf{s}(t)\}$. Analyses that incorporate explicit or implicit models

of the disturbance may be carried out using the Bayesian framework, such as the Markov chain Monte Carlo (MCMC) simulation methods exploited e.g. in [27] and [28]. In this case, the problem of joint detection and reconstruction of corrupted samples can be solved by means of Gibbs sampling, i.e., by iterative approximation of the joint posterior distribution of the clean signal, its AR-model parameters and noise pulse parameters. In [27] disturbance is modeled as a Gaussian process with time-varying variance and Markov chain pulse activity (on/off) structure. Since in general a large number of sampling iterations may be needed for the posterior distribution to reach its steady state form, the method is highly computationally intensive. The computational load of the MCMC approach can be reduced if a more specific, deterministic model is incorporated. Such an approach was proposed in [28], where audio clicks were modeled as exponentially decaying pulses with adjustable location and shape parameters. However, the computational complexity is still considerable.

In contrast with the aforementioned,, the proposed methods belong to the class of non-iterative one-pass (causal) or two-pass (noncausal) procedures with a relatively low computational requirements. However, in our opinion, there is not much sense in setting the MCMC approach against the currently developed one. A much better idea is to use both approaches jointly, e.g. to use the MCMC algorithm described in [27] to *refine* the results yielded by one of the noncausal algorithms proposed in this paper. Such post-processing may be computationally affordable since, as noted in [27], only a few iterations of the MCMC algorithm are usually needed when the starting point is close to the final solution. Alternatively, Gibbs sampling can be used only to reconstruct the corrupted fragments [29]. This may make a difference in the case of long noise pulses as it allows one to avoid the covariance defect typical of maximum likelihood and least squares reconstructions [23] ("collapse" of the reconstructed material in the middle of large gaps).

Another example of a synergetic approach is given in our recent paper [30], which describes the computationally cheap disturbance localization method based on template matching. This method can be used when noise pulses have highly repetitive shapes that match several typical "patterns" or templates. To localize noise pulses, the appropriately modified disturbance templates are correlated with the sequence of one-step-ahead prediction errors. However, such a strategy can be used only in cases where the detected pulse has a typical shape, matching one of the templates. Otherwise (for atypical disturbance waveforms) the classical general purpose AR-model based localization technique must be applied. This means that, rather than competing, both techniques complement each other.

V. Simulation results

In this section we will show how the proposed methods can be used to solve two practical problems - elimination of impulsive disturbances from archive audio signals (part A), and robust parametric spectrum estimation (part B).

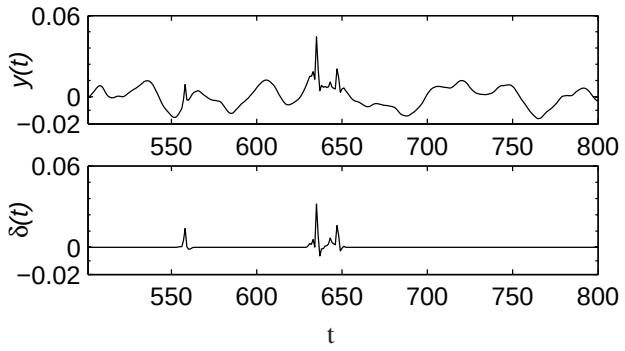

Fig. 1: A fragment of an archive gramophone recording corrupted by impulsive disturbances (clicks).

A. Restoration of audio recordings

The proposed detection rules were evaluated and compared in the two-dimensional case – they were used to localize impulsive disturbances in stereo audio signals. To obtain realistic test signals, $40 = 4 \times 10$ clean audio recordings representing 4 music categories (jazz, choir, opera, classical), lasting for about 22 seconds each, were sampled at the rate of 48 kHz with 16-bit resolution and corrupted by means of adding a sequence of real click waveforms extracted from silent parts of a real archive stereo gramophone recording – see Fig. 1. The total number of pulses added to each recording (the same sequence in all cases) was equal to 1853: 467 pulses corrupting the left channel only, 454 pulses corrupting the right channel only, and 932 pulses corrupting both channels (different pulses in different channels). The total number of corrupted samples was equal to 15650, which constitutes 1.48% of all samples. The width of noise pulses (picked at random from the stereo click database containing 710 waveforms) ranged from 3 samples to 41 samples and their location was random.

Quite obviously, detection efficiency depends on how well the decision sequence $\widehat{d}(t)$ matches the "ground truth" pulse location function $d(t)$. We will use three measures which quantify discrepancy/similarity between $\widehat{d}(t)$ and $d(t)$: degree of underfitting, degree of overfitting and degree of overlapping.

Denote by

$$\mathcal{D}_{true} = \{t : d(t) = 1\}$$

the set showing location of all outliers[1] added to clean signals (the same for all recordings), and by

$$\mathcal{D}_{neg} = \{t : d(t) = 1 \wedge \widehat{d}(t) = 0\}$$

the set indicating location of samples classified as not-outliers (false negatives). Similarly, let

$$\mathcal{D}_{pos} = \{t : d(t) = 0 \wedge \widehat{d}(t) = 1\}$$

be the set showing location of samples incorrectly classified as outliers (false positives). The degrees of underfitting ξ_u and overfitting ξ_o can be defined as follows

$$\xi_u = \frac{\mathrm{card}\{\mathcal{D}_{neg}\}}{\mathrm{card}\{\mathcal{D}_{true}\}} \tag{35}$$

[1]Since each noise pulse is a sequence of outliers, the number of outliers is several times greater than the number of noise pulses.

$$\xi_o = \frac{\mathrm{card}\{\mathcal{D}_{pos}\}}{\mathrm{card}\{\mathcal{D}_{true}\}} \tag{36}$$

where $\mathrm{card}\{\mathcal{X}\}$ denotes cardinality (number of elements) of the set $\mathcal{X}$. Both performance measures quantify inaccuracy of the detection process (the smaller, the better), including such detection errors as failure to detect existing noise pulses (ξ_u), detection of nonexistent noise pulses (ξ_o) and imprecise localization of the pulse beginning and/or end points (ξ_u, ξ_o).

Finally, denote by $\mathcal{D}_{int}$ and $\mathcal{D}_{union}$ the sets corresponding to the "intersection" and "union" of $\widehat{d}(t)$ and $d(t)$

$$\mathcal{D}_{int} = \{t : d(t) = 1 \wedge \widehat{d}(t) = 1\}$$
$$\mathcal{D}_{union} = \{t : d(t) = 1 \vee \widehat{d}(t) = 1\}.$$

The degree of overlapping, defined as follows

$$\kappa = \frac{\mathrm{card}\{\mathcal{D}_{int}\}}{\mathrm{card}\{\mathcal{D}_{union}\}} \tag{37}$$

takes its maximum value, equal to 1.0, when $\widehat{d}(t) \equiv d(t)$, i.e., when both sequences coincide. Note that the degree of overlapping also takes into account detection errors such as false positives and false negatives [both increase the denominator in (37)].

Localization of impulsive disturbances in archive audio signals is not the ultimate goal of processing – the real goal is to restore the original (undistorted) audio material or, more precisely, to create audio that *sounds* like the original one. This means that the differences between the original (clean) audio signal and its restored version that are not audible are tolerable. One should realize that, depending on the audio context (i.e., on local characteristics of the corrupted audio signal), some even very "small" disturbances may be audible (e.g. when the signal is narrowband), while some relatively "large" ones may be not audible (e.g. when the corrupted signal is wideband). This means that overlooking (not detecting) noise pulses may have but must not have perceptual consequences. Similar is the situation with detection of nonexistent pulses – when the AR-model based signal reconstruction is of high quality, many false alarms do not yield audible signal distortions (although, in general, large number of false alarms – especially the long ones – causes degradation of the restored signal). For this reason simply counting false positives and false negatives does not make much sense.

In the case considered the most meaningful test of pulse detection efficiency is via assessment of the perceptual quality of the restored audio recording. Audio quality can be checked using listening tests but this is usually a very time-consuming procedure. Fortunately, our experience has shown that similar results can be obtained using the PEAQ (Perceptual Evaluation of Audio Quality) tool (originally designed to measure the quality of perceptual audio coders) – a specialized software which scores the restored audio (by comparing it with the original, noiseless recording) using several perceptual (psychoacoustic) criteria [31], [32]. PEAQ scores take negative values that range from -4 (very annoying distortions) to 0 (imperceptible distortions). In the impulsive noise removal context, improvement of the PEAQ score by 0.1 (or more) is usually audible, i.e., perceptually significant.

The results of quantitative and qualitative evaluation of different detection schemes are summarized in Tables II and III, respectively. All detectors incorporated VAR models of order $n = 10$. Estimation of autoregressive coefficients $\mathbf{A}_1^f(t), \ldots, \mathbf{A}_{10}^f(t)$ and $\mathbf{A}_1^b(t), \ldots, \mathbf{A}_{10}^b(t)$ was carried out using the EWLS algorithm with forgetting constant $\lambda = 0.999$. The remaining parameters were set to $\lambda_0 = 0.993$ and $k_{\max} = 50$.

Table II shows the average detection scores, evaluated for all 40 recordings, obtained for 4 unidirectional algorithms (A, B, C, D) and 4 bidirectional (A^*, B^*, C^*, D^*) algorithms specified in Table I.

Detection threshold was restricted to the range $[3, 4.5]$. For the values of μ higher than 4.5 the restoration quality significantly drops due to a large number of overlooked noise pulses and/or undersized detection alarms. A similar effect can be observed for $\mu < 3$ due to a large number of false and/or oversized detection alarms.

In addition to sample-level statistics, such as degree of underfitting (Tab. IIa), degree of overfitting (Tab. IIb), and degree of overlapping (Tab. IIc), we show two pulse-level statistics: the number of undetected noise pulses and the number erroneously detected (nonexistent) pulses – see Tab. IId. Note that each pulse-level decision corresponds to a sequence of elementary sample-level decisions.

Table III shows the qualitative results (divided into 4 music categories) obtained using the PEAQ tool. In addition to the algorithms A, B, C, D, and their bidirectional counterparts A^*, B^*, C^*, D^*, described above, comparison includes results obtained when the semi-causal detection rule is incorporated in the causal (E) and noncausal (E^*) algorithms that were designed in [3] for the processing of stereo audio files (denoted there by D and D^*, respectively). The modified algorithm E is denoted by F and the modified algorithm E^* – by F^*. The algorithms E and E^* were developed to address a specific problem – elimination of clicks from archive gramophone recordings – and as such they include several specific mechanisms, absent from the general purpose solutions presented in this paper:

1) Extension of detection alarms – each forward/backward detection alarm is shifted back by a small fixed number of samples (to improve localization of noise pulses with "soft" edges, typical of archive gramophone recordings).
2) Incorporation of a larger and more diversified set of fusion rules used to combine forward and backward detection alarms (established experimentally in [21] for different detection patterns).
3) Separate treatment of noise pulses detected in only one of two stereo channels (technique that is difficult to apply when the number of channels m is larger than 2).

The results summarized in Tables II and III, show clearly advantages of bidirectional, forward-backward processing (A, B, C, D, E and F versus A^*, B^*, C^*, D^*, E^* and F^*, respectively) and advantages of the decision-feedback strategy (A versus C, B versus D, A^* versus C^*, B^* versus D^*). Equally importantly, – and this is the main contribution of

A	–	causal, equipped with open-loop detection scheme
B	–	semi-causal, equipped with open-loop detection scheme
C	–	causal, equipped with decision-feedback detection scheme
D	–	semi-causal, equipped with decision-feedback detection scheme
E	–	causal, equipped with decision-feedback detection scheme, dedicated to the processing of stereo gramophone recordings
F	–	semi-causal, equipped with decision-feedback detection scheme, dedicated to the processing of stereo gramophone recordings
A*	–	noncausal, combining causal and anticausal, equipped with open-loop detection scheme
B*	–	noncausal, combining semi-causal and semi-anticausal, equipped with open-loop detection scheme
C*	–	noncausal, combining causal and anticausal, equipped with decision-feedback detection scheme
D*	–	noncausal, combining semi-causal and semi-anticausal, equipped with decision-feedback detection scheme
E*	–	noncausal, combining causal and anticausal, equipped with decision-feedback detection scheme, dedicated to the processing of stereo gramophone recordings
F*	–	noncausal, combining semi-causal and semi-anticausal, equipped with decision-feedback detection scheme, dedicated to the processing of stereo gramophone recordings

TABLE I: Description of the compared detection/reconstruction algorithms.

this paper – they demonstrate the benefits of applying semi-causal detection (A versus B, C versus D, A* versus B*, C* versus D*, E versus F, E* versus F*). Note that the specialized algorithms (E, E*) perform better than the general purpose ones (D, D*), and that the results further improve when semi-causal detection is applied (F, F*).

The most important design parameters are the detection threshold μ and the order of the VAR model n. In audio applications $\mu = 4$ is usually a good choice. Table IV shows dependence of the PEAQ scores on the model order. The best scores were obtained for $n = 10$. Such result may look as counterintuitive as one has the right to expect that increasing the model order should improve both detection of noise pulses and signal reconstruction. It seems that this does not happen for at least two reasons. First, since model parameters are estimated based on a finite amount of past data (determined by the degree of signal nonstationarity), high order models, obtained by means of signal identification, may have worse predictive capabilities than their lower order counterparts, even though in principle they are more "flexible" – this stems from the so-called principle of parsimony [25]. Note, for example, that in the case considered ($m = 2$) the VAR model of order $n = 40$ requires estimation of $N = nm^2 = 160$ coefficients Second, due to adoption of the alarm blocking strategy (detection alarms that are separated by less that n samples are clustered), higher order models tend to produce a larger number of unnecessarily long detection alarms than lower order models.

Another order-related factor, important from the practical viewpoint, is computational load of the compared algorithms, which in all cases is roughly proportional to N^2. The average times needed to process a 22 s recording on a laptop equipped with a single core 2.6 GHz processor are shown in Table V.

(a) degree of underfitting

μ	A	B	C	D	A*	B*	C*	D*
4.5	0.69	0.72	0.27	0.33	0.50	0.57	0.36	0.42
4	0.65	0.68	0.23	0.29	0.44	0.52	0.32	0.38
3.5	0.59	0.63	0.19	0.25	0.36	0.45	0.26	0.33
3	0.50	0.54	0.15	0.18	0.23	0.35	0.16	0.26

(b) degree of overfitting

μ	A	B	C	D	A*	B*	C*	D*
4.5	0.62	0.24	0.70	0.35	0.09	0.02	0.12	0.03
4	1.05	0.35	1.17	0.50	0.24	0.03	0.31	0.05
3.5	2.40	0.57	2.46	0.81	0.93	0.07	1.08	0.13
3	7.17	1.22	7.51	1.76	5.69	0.33	6.11	0.61

(c) degree of overlapping

μ	A	B	C	D	A*	B*	C*	D*
4.5	0.21	0.25	0.52	0.56	0.46	0.42	0.58	0.56
4	0.19	0.27	0.46	0.56	0.47	0.47	0.57	0.59
3.5	0.14	0.29	0.30	0.53	0.39	0.52	0.46	0.60
3	0.06	0.28	0.11	0.41	0.13	0.52	0.14	0.53

(d) number of erroneously detected (E) and undetected (U) noise pulses

μ	A		B		C		D	
	E	U	E	U	E	U	E	U
4.5	1294	353	465	538	808	237	211	374
4	2414	300	645	463	1858	195	387	317
3.5	5910	248	1073	384	5328	152	930	256
3	16423	209	2789	289	15213	120	3566	177

μ	A*		B*		C*		D*	
	E	U	E	U	E	U	E	U
4.5	116	583	8	816	160	354	16	498
4	306	481	19	716	370	297	43	436
3.5	989	356	69	593	1059	224	137	369
3	5029	197	334	434	5203	124	633	272

TABLE II: Detection statistics obtained for 4 unidirectional (A, B, C, D) and 4 bidirectional (A*, B*, C*, D*) detection algorithms described in the paper. All results were obtained for 40 artificially corrupted audio files.

n	A	B	C	D	A*	B*	C*	D*
10	3	5	3	6	8	12	8	13
15	4	9	4	11	12	20	12	24
20	12	20	12	22	34	49	33	54
25	19	31	20	38	49	72	50	86
30	22	44	25	95	57	98	67	180
35	44	118	52	234	88	226	121	492
40	72	220	118	534	187	510	254	1092

TABLE V: Dependence of the average processing times (in seconds) on the model order n. All results were obtained for 40 artificially corrupted audio files and $\mu = 4$.

B. Robust spectrum estimation

Spectral analysis of experimental data provides useful qualitative and quantitative information about their contents. Spectral density (matrix) function of a nonstationary VAR process can be defined in the following form

$$\mathbf{S}(\omega, t) = \mathbf{A}_f^{-1}(e^{j\omega}, t)\boldsymbol{\rho}_f(t)\mathbf{A}_f^{-\mathrm{T}}(e^{-j\omega}, t) \qquad (38)$$

$$\mathbf{A}_f(z, t) = \mathbf{I} - \sum_{i=1}^{n} \mathbf{A}_i^f(t) z^i$$

jazz music

n	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
4.5	-2.99	-2.15	-1.88	-1.56	-0.98	-1.06	-0.88	-0.94	-0.78	-1.01	-0.71	-0.77	-0.52
4		-2.43	-1.85	-2.01	-1.16	-1.40	-0.96	-1.36	-0.76	-1.44	-0.74	-1.19	-0.58
3.5		-2.95	-2.03	-2.72	-1.54	-1.90	-1.21	-2.27	-0.84	-2.29	-0.95	-1.73	-0.83
3		-3.52	-2.45	-3.45	-2.22	-2.86	-1.73	-3.43	-1.31	-3.44	-1.55	-3.06	-1.43

choir music

n	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
4.5	-3.82	-3.73	-3.71	-0.63	-0.69	-0.50	-0.58	-1.53	-1.74	-0.58	-0.74	-0.21	-0.21
4		-3.71	-3.66	-0.73	-0.66	-0.50	-0.52	-1.60	-1.43	-0.60	-0.65	-0.34	-0.20
3.5		-3.70	-3.59	-1.18	-0.67	-0.59	-0.50	-2.18	-1.18	-0.90	-0.61	-0.70	-0.29
3		-3.75	-3.51	-2.91	-0.95	-1.64	-0.53	-3.32	-1.02	-3.01	-0.72	-3.25	-1.38

opera music

n	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
4.5	-3.79	-3.78	-3.75	-0.90	-0.93	-0.61	-0.75	-1.64	-1.82	-0.74	-0.96	-0.54	-0.35
4		-3.78	-3.74	-1.07	-0.88	-0.62	-0.67	-1.97	-1.62	-0.76	-0.86	-0.65	-0.44
3.5		-3.77	-3.68	-1.80	-0.89	-0.89	-0.60	-2.29	-1.37	-1.21	-0.78	-1.66	-0.61
3		-3.80	-3.65	-3.52	-1.49	-2.14	-0.87	-3.63	-1.27	-3.47	-1.00	-3.62	-1.39

classical music

n	REF	A	B	C	D	E	F	A*	B*	C*	D*	E*	F*
4.5	-3.11	-2.46	-2.28	-0.69	-0.58	-0.54	-0.48	-0.87	-0.79	-0.58	-0.54	-0.56	-0.26
4		-2.48	-2.22	-0.83	-0.59	-0.57	-0.46	-0.93	-0.75	-0.63	-0.52	-0.56	-0.50
3.5		-2.75	-2.15	-1.40	-0.68	-0.76	-0.50	-1.36	-0.72	-1.14	-0.57	-1.28	-0.60
3		-3.45	-2.18	-2.94	-1.09	-1.52	-0.66	-2.93	-0.91	-2.94	-0.98	-2.88	-0.99

TABLE III: Comparison of the average PEAQ scores obtained for 6 unidirectional (A, B, C, D, E, F) and 6 bidirectional (A*, B*, C*, D*, E*, F*) detection/reconstruction algorithms described in Table I. All results were obtained for 40 artificially corrupted audio files divided into 4 music categories. REF denotes the average score of the input (corrupted) recordings.

n	A	B	C	D	A*	B*	C*	D*
10	-3.10	-2.87	-1.16	-0.82	-1.46	-1.14	-0.86	-0.69
15	-3.15	-2.92	-1.34	-0.94	-1.50	-1.18	-0.99	-0.71
20	-3.17	-2.95	-1.53	-1.23	-1.70	-1.27	-1.12	-0.78
25	-3.22	-2.98	-2.03	-1.86	-1.88	-1.57	-1.35	-1.19
30	-3.49	-3.32	-2.57	-2.59	-2.04	-1.81	-1.96	-2.01
35	-3.33	-3.30	-2.34	-2.30	-2.62	-2.39	-3.15	-3.07
40	-3.28	-3.36	-2.30	-2.48	-3.23	-2.95	-3.47	-3.43

TABLE IV: Dependence of the mean PEAQ scores, evaluated for 8 detection/reconstruction algorithms described in the paper, on the model order n. All results were obtained for 40 artificially corrupted audio files and $\mu = 4$.

where $\mathbf{A}_i^f(t), i = 1, \ldots, n$, and $\boldsymbol{\rho}_f(t)$ denote the time-varying parameters of the VAR model (an equivalent definition can be based on the backward-time VAR signal representation). The quantity (38) can be interpreted as the spectrum of a stationary process "tangent", at time t, to the nonstationary process under study. As shown by Dalhaus in his seminal work on locally stationary stochastic processes [33], [34], [35], under pretty weak assumptions imposed on the time-varying VAR model (parameter trajectories of bounded variation[2], uniform stability) such a definition of a time-varying spectrum is theoretically well justified.

The parametric spectrum estimate can be obtained by replacing in (38) the true parameters with their estimates

$$\widehat{\mathbf{S}}(\omega, t) = \widehat{\mathbf{A}}_f^{-1}(e^{j\omega}, t)\widehat{\boldsymbol{\rho}}_f(t)\widehat{\mathbf{A}}_f^{-\mathrm{T}}(e^{-j\omega}, t) \qquad (39)$$

[2]A function of bounded variation may have countably many jumps with absolutely summable jump sizes; it is continuous in between the jumps.

$$\widehat{\mathbf{A}}_f(z, t) = \mathbf{I} - \sum_{i=1}^{n} \widehat{\mathbf{A}}_i^f(t) z^i.$$

The EWLS algorithm for estimation/tracking of autoregressive coefficients $\mathbf{A}_i^f, i = 1, \ldots, n$, was already presented in Section IV. The EWLS estimator of $\boldsymbol{\rho}_f(t)$ can be expressed in the following (steady state) form (cf. equations (9), (12) and (13) in [3])

$$\widehat{\boldsymbol{\rho}}_f(t) = \lambda\widehat{\boldsymbol{\rho}}_f(t-1) + (1-\lambda)\left[1 - \mathbf{k}_f^{\mathrm{T}}(t)\boldsymbol{\varphi}_f(t)\right]\boldsymbol{\varepsilon}_f(t|t-1)\boldsymbol{\varepsilon}_f^{\mathrm{T}}(t|t-1) \qquad (40)$$

where $\boldsymbol{\varepsilon}_f(t|t-1) = [\varepsilon_1^f(t|t-1), \ldots, \varepsilon_m^f(t|t-1)]^{\mathrm{T}}$ and $\mathbf{k}_f(t)$ are the quantities updated in (32).

In the presence of impulsive disturbances the results of spectral analysis can be seriously degraded unless noise pulses are localized and removed.

Our second simulation experiment was designed to check detection efficiency of selected methods looking at the problem

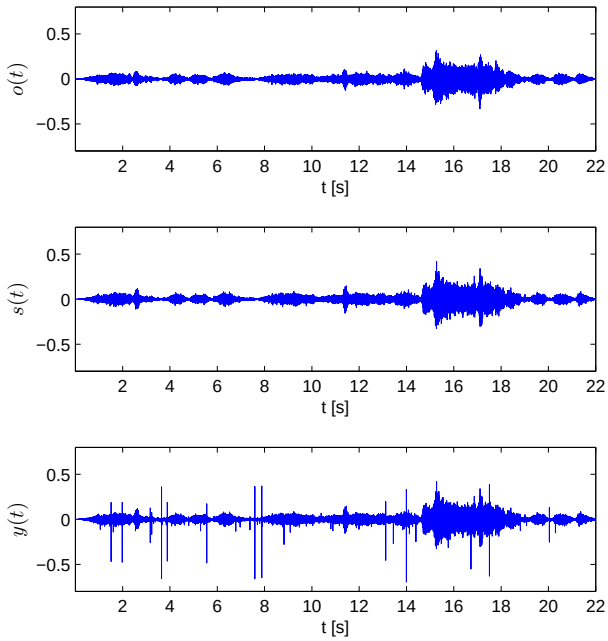

Fig. 2: Left channel of the original stereo audio file (top), synthetic audio file (middle) and its corrupted version (bottom) obtained by adding impulsive disturbances.

from the spectrum estimation perspective. To work with "realistic" data, one of the clean stereo audio recordings utilized in the previous experiment was identified using the EWLS algorithms (32) and (40). The time-varying "ground truth" VAR model obtained in this way was next used to generate 20 independent realizations of a synthetic audio signal $\mathbf{s}(t)$ with similar characteristics as the original one – see Fig. 2.

As an instantaneous spectral distortion measure we adopted the relative entropy rate (RER) [36]

$$d_{\mathrm{RER}}(t) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big\{\mathrm{tr}\left[\left(\mathbf{S}(\omega,t)-\widehat{\mathbf{S}}(\omega,t)\right)\widehat{\mathbf{S}}^{-1}(\omega,t)\right] - \log\det\left[\mathbf{S}(\omega,t)\widehat{\mathbf{S}}^{-1}(\omega,t)\right]\Big\}\,d\omega \tag{41}$$

which is a multivariate extension of the classical Itakura-Saito measure [37].

During the simulation experiment the true time-varying spectrum of the synthetic audio signal $\mathbf{s}(t)$ and the estimates obtained by means of processing its corrupted version $\mathbf{y}(t)$ (created by adding impulsive disturbances) were evaluated at 128 equidistant frequencies using the FFT-based procedure. The mean RER scores (obtained by both time and ensemble averaging), evaluated for 4 causal/semi-causal detection algorithms (A, B, C, D), are shown in Tab. VI. Note huge improvement of estimation accuracy observed when the open-loop detection scheme (used in A and B) is replaced with the decision-feedback one (used in C and D).

The benefits of applying the robust spectral estimation technique are demonstrated in Fig. 3. Spectral estimates shown

μ	A	B	C	D
4.5	1.325	1.315	0.099	0.101
4	1.238	1.227	0.099	0.099
3.5	1.130	1.112	0.103	0.099
3	1.018	0.964	0.157	0.101

TABLE VI: Comparison of the mean RER scores evaluated for 4 causal/semi-causal detection algorithms described in the paper. All results were obtained for 20 realizations of an artificial audio signal. The mean score obtained for the input (corrupted) signal was equal to 2.630, and the mean "ground truth" score, obtained when location of noise pulses is known exactly, was equal to 0.073.

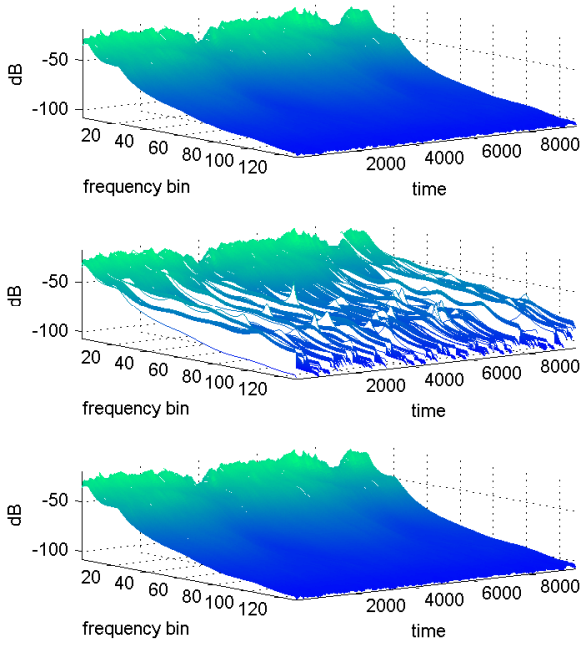

Fig. 3: Evolution of the true spectral density function of a synthetic audio signal (top) and two spectral estimates obtained by means of processing its corrupted version: without (middle) and with (bottom) removal of impulsive disturbances.

in Fig. 3 were obtained for a 2 second long excerpt (between 4 and 6 seconds) from the corrupted synthetic audio signal shown in Fig. 2.

VI. Conclusion

Several, both causal and noncausal, decision rules were presented, allowing one to localize impulsive disturbances in nonstationary multivariate autoregressive signals. It was shown that detection quality can be improved if detection is based on joint evaluation of signal prediction errors and signal interpolation errors. The paper was illustrated with the results of processing stereo audio files artificially corrupted with real impulsive disturbances extracted from archive gramophone recordings. Both detection statistics and perceptual scores

obtained using the PEAQ tool show that the strengthened detection rules, which incorporate signal interpolation errors, yield better results than the classical rules based on evaluation of signal prediction errors only. Similar conclusions can be drawn after comparing the results of spectral analysis of audio signals corrupted with outliers.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for comments that helped to improve the manuscript.

REFERENCES

[1] M. Ciołek, M. Niedźwiecki, S. Sieklicki, J. Drozdowski and J. Siebert, "Automated Detection of sleep apnea and hypopnea events based on robust airflow envelope tracking in the presence of breathing artifacts," IEEE J. Biom. and Health Inf., vol. 19, pp. 418–429, 2015.

[2] S. Sanei and J. Chambers, *EEG Signal Processing*. New York: Wiley, 2007.

[3] M. Niedźwiecki, M. Ciołek and K. Cisowski, "Elimination of impulsive disturbances from stereo audio recordings using vector autoregressive modeling and variable-order Kalman filtering," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 23, pp. 970–981, 2015.

[4] A.J. Fox, "Outliers in time series," *J. Royal Statist. Soc. Ser. B*, vol. 34, pp. 350–363, 1972.

[5] M. Gupta, J. Gao, C.C. Aggarwal and K. J. Han, "Outlier detection for temporal data: A survey," *IEEE Trans. Knowledge Data Eng.*, vol. 25, pp. 1–20, 2014.

[6] R. S. Tsay, "Time series model specification in the presence of outliers," *J. Amer. Statist. Ass.*, vol. 81, pp. 132–141, 1986.

[7] I. Chang, G. C. Tiao, and C. Chen, "Estimation of time series parameters in the presence of outliers," *Technometrics*, vol. 30, pp. 193–204, 1988.

[8] C. Chen and L.-M. Liu, "Joint estimation of model parameters and outlier effects in time series," *Journal of the American Statistical Association*, vol. 88, pp. 284–297, 1993.

[9] G. M. Ljung, "On outlier detection in time series," *J. Royal Statist. Soc. Ser. B*, vol. 55, pp. 559–567, 1993.

[10] B. Abraham and A. Chuang, "Outlier detection and time series modeling," *Technometrics*, vol. 31, pp. 241–248, 1989.

[11] A. M. Bianco, M. García Ben, E. J. Martnez, and V. J. Yohai, "Outlier detection in regression models with ARIMA errors using robust estimates," *Journal of Forecasting*, vol. 20, pp. 565–579, 2001.

[12] A. Justel, D. Peña, and R. S. Tsay, "Detection of outlier patches in autoregressive time series," *Statistica Sinica*, vol. 11, pp. 651–674, 2001.

[13] C.M. Hicks and S.J. Godsill, "A two-channel approach to the removal of impulse noise from archived recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 1994, (2) pp. 213-216.

[14] R.S. Tsay, D. Peña, and A.E. Pankratz, "Outliers in multivariate time series," *Biometrika*, vol. 87, pp. 789–804, 2000.

[15] J.-M. Helbling and R. Cléroux, "On outlier detection in multivariate time series," *Acta Math. Vietnamica*, vol 34, pp. 19–26, 2009.

[16] P. Galeano, D. Peña, and R. S. Tsay, "Outlier detection in multivariate time series by projection pursuit," *J. Amer. Statist. Ass.*, vol. 101, pp. 654–669, 2006.

[17] R. Baragona and F. Battaglia, "Outliers detection in multivariate time series by independent component analysis," *Neural Computation*, vol. 19, pp. 1962–1984, 2007

[18] M. Niedźwiecki and K. Cisowski, "Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals," *IEEE Trans. Audio, Signal Process.*, vol. 44, pp. 528–537, 1996.

[19] J.S. Godsill and J.P.W. Rayner, *Digital Audio Restoration*, Springer-Verlag, 1998.

[20] S.V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley, 2008.

[21] M. Niedźwiecki and M. Ciołek, "Elimination of impulsive disturbances from archive audio signals using bidirectional processing," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, pp. 1046–1059, 2013.

[22] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag, 2005.

[23] M. Niedźwiecki, "Statistical reconstruction of multivariate time series," *IEEE Trans. Signal Process.*, vol. 41, pp. 451–457, 1993.

[24] T. Söderström and P. Stoica, *System Identification*. Englewoods Cliffs NJ: Prentice-Hall, 1988.

[25] M. Niedźwiecki, *Identification of Time-varying Processes*. New York: Wiley, 2000.

[26] A. Zellner, "An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias," *Journ. Am. Statist. Ass.*, vol. 57, pp. 348–368, 1962.

[27] S.J. Godsill and P.J.W. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech Audio Process.*, vol. 6, pp. 352–372, 1998.

[28] F.R. Ávila and L.W.P. Biscainho, "Bayesian restoration of audio signals degraded by impulsive noise modeled as individual pulses," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, pp. 2470–2481, 2012.

[29] J.J.K.Ó Ruanaidh and W.J. Fitzgerald, "Interpolation of missing samples for audio restoration," *Electron. Lett.*, vol. 30, 1994.

[30] M. Niedźwiecki and M. Ciołek, "Localization of impulsive disturbances in audio signals using template matching," *Digital Signal Process.*, vol. 46, pp. 253–265, 2015.

[31] ITU-R Recommendation BS.1387, "Method for Objective Measurements of Perceived Audio Quality," 1998.

[32] P. Kabal, "An Examination and Interpretation of ITU-R Recommendation BS.1387: Perceptual Evaluation of Audio Quality," Department of Electrical & Computer Engineering, McGill University, Canada, 2003.

[33] R. Dahlhaus, "Fitting time series models to nonstationary processes," *Ann. Statist.*, vol. 30, pp. 351–413, 1997.

[34] R. Dahlhaus, "Local inference for locally stationary time series based on the empirical spectral measure," *Journal of Econometrics*, vol. 151, pp. 101–112, 2009.

[35] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.

[36] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Automat. Contr.*, vol. 57, pp. 2561–2575, 2012.

[37] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Jap.*, vol. 53-A, pp. 36–43, 1970.

Maciej Niedźwiecki (M'08, SM'13) received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively. He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986-1989. In 1990 - 1993 he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His main areas of research interests include system identification, statistical signal processing and adaptive systems.

Dr. Niedźwiecki is currently a member of the IFAC committees on Modeling, Identification and Signal Processing and on Large Scale Complex Systems, and a member of the Automatic Control and Robotics Committee of the Polish Academy of Sciences (PAN). He works as a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology.

Marcin Ciołek (M'17) received the M.Sc. and Ph.D. degrees from the Gdańsk University of Technology (GUT), Gdańsk, Poland, in 2010 and 2017, respectively. Since 2017, he has been working as an Adjunct Professor in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, GUT. His professional interests include speech, music and biomedical signal processing.