Postprint of: Niedźwiecki M., Ciołek M., Cisowski K., Elimination of Impulsive Disturbances From Stereo Audio Recordings Using Vector Autoregressive Modeling and Variable-order Kalman Filtering, IEEE Transactions on Audio Speech and Language Processing, Vol. 23, Iss. 6 (2015), pp. 970-981, DOI: 10.1109/TASLP.2015.2414823

Elimination of Impulsive Disturbances From Stereo Audio Recordings Using Vector Autoregressive Modeling and Variable-order Kalman Filtering

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

***Abstract*—This paper presents a new approach to elimination of impulsive disturbances from stereo audio recordings. The proposed solution is based on vector autoregressive modeling of audio signals. On-line tracking of signal model parameters is performed using the exponentially weighted least squares algorithm. Detection of noise pulses and model-based interpolation of the irrevocably distorted samples is realized using an adaptive, variable-order Kalman filter. The proposed approach is evaluated on a set of clean audio signals contaminated with real click waveforms extracted from old gramophone recordings.**

***Index Terms*—Elimination of impulsive disturbances, vector autoregressive models, adaptive Kalman filtering.**

I. Introduction

ARCHIVE audio files, such as old gramophone recordings, are often degraded by impulsive disturbances. Clicks, pops, ticks and record scratches are caused by aging and/or mishandling of the surface of gramophone records, specks of dust and dirt, faults in the record stamping process etc. In the case of magnetic tape recordings, impulsive disturbances can be usually attributed to transmission or equipment artifacts (e.g. electric or magnetic pulses). Elimination of such disturbances from archive audio documents is an important element of saving our cultural heritage.

Most of the known approaches to elimination of impulsive disturbances from archive audio signals are based on adaptive prediction – the autoregressive (AR) model of the analyzed signal is continuously updated and used to predict consecutive signal samples [1]–[10]. Whenever the absolute value of the one-step-ahead prediction error becomes too large, namely when it exceeds a prescribed multiple of its estimated standard deviation, a "detection alarm" is raised, and the predicted sample is scheduled for reconstruction. The test is then extended to multiple-step-ahead prediction errors – detection alarm is terminated when a given number of samples in a row remain sufficiently close to the predicted signal trajectory (or when the length of detection alarm reaches its maximum allowable value). As shown in [11], detection results can be further improved if the results of forward-time detection are combined with the analogous results of backward-time detection. The latter can be obtained by means of processing audio signal backward in time, provided, of course, that the entire recording is available. In addition to reducing the number and length of false alarms, bidirectional processing allows one to carve detection alarms more carefully (smaller number of overlooked noise pulses, better front/end matching of noise pulses).

An alternative approach to detection of noise pulses, based on double thresholding, was proposed in [12]. Unlike sequential prediction-based methods mentioned before, the double threshold approach incorporates block processing. The signal is divided into blocks (possibly overlapping), each of which is analyzed separately. For each block the best-fitting AR signal model is determined and used to compute the sequence of residual errors. The detection procedure is two-step. The aim of the first, prescreening step is to find the abnormally large values of residual errors (attributed to the presence of noise pulses). This is achieved by means of using an outlier detector equipped with a relative high detection threshold. The purpose of the second step is to precisely localize the beginning and end points of each preliminary detection alarm found during prescreening. Localization is performed using outlier detector equipped with a small detection threshold.

Once the impulsive disturbance is localized, the corrupted samples are reconstructed using the AR-model based projection technique [13] or its Bayesian extension known as Gibbs sampling [4], [5].

Although two tracks of a stereophonic audio signal can be split and processed separately, this is certainly not the best approach to restoration of stereo recordings. We will show that both detection and reconstruction (interpolation) of irrevocably distorted samples can be performed more reliably when two channels are analyzed jointly using the vector autoregressive modeling technique. Fig. 1 compares one-step-ahead prediction errors obtained – for a typical clean stereo audio signal – using scalar and vector modeling, respectively. In both cases model parameters were estimated using the method of exponentially weighted least squares with forgetting factors chosen so as to equalize estimation memory of the compared approaches (for more details see Section 2.D). Note that the joint left/right channel analysis allows one to model audio signal more accurately – in the case considered the variance of the prediction errors was reduced by the factor of 1.7 for the left channel, and by the factor of 2.1 for the right channel.

When it comes to signal restoration, more accurate modeling

This work was supported by the National Science Centre under the agreement UMO-2013/09/B/ST7/ 01582.

The authors are with the Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk, Poland (e-mails: maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl, krci@eti.pg.gda.pl).

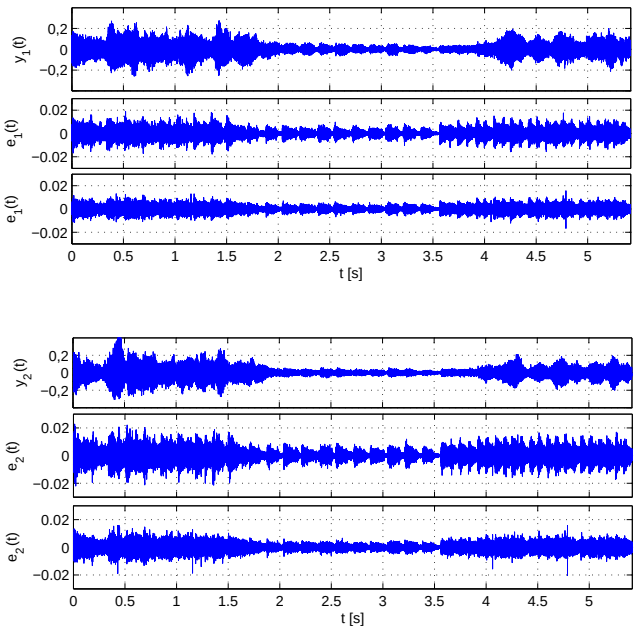

Fig. 1: Comparison of the one-step-ahead prediction errors obtained – for a typical stereo audio signal (top plots in each group) – using scalar signal modeling (middle plots in each group) and vector signal modeling (bottom plots in each group). The upper group of three plots corresponds to the left stereo channel, and the lower group – to the right stereo channel. Note the scale difference between top plots and lower plots in each group.

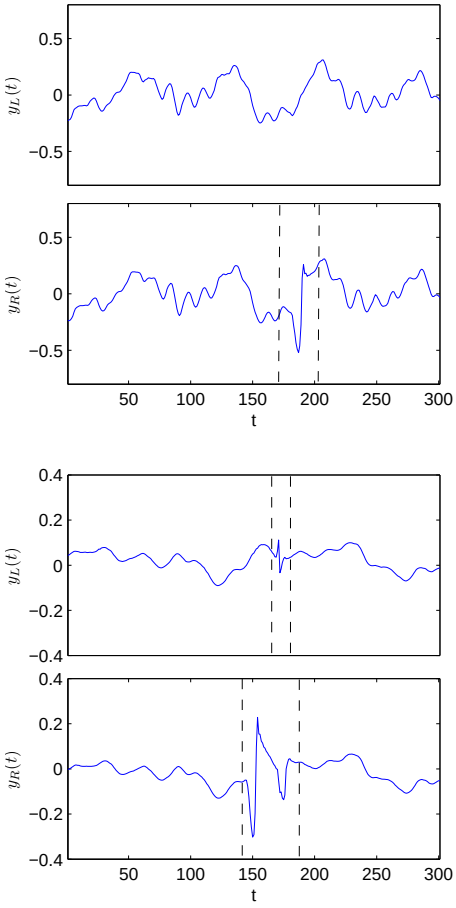

Fig. 2: Typical impulsive noise patterns encountered in archive stereo gramophone recordings: impulsive disturbance corrupting only one of two stereo tracks (the upper two plots), and impulsive disturbance corrupting both tracks (the lower two plots). Broken vertical lines show the beginning and the end of each noise pulse.

has two potential benefits. First, since most of the existing noise pulse detection procedures are based on adaptive thresholding of signal prediction errors, vector processing makes them more sensitive to abnormal signal patterns. Second, more accurate models usually guarantee better interpolation of irrevocably distorted samples called in question by the outlier detector.

When the restored audio material originates from stereo gramophone recordings, there is an additional incentive to use the vector approach. In the case of archive stereo gramophone recordings, the local groove damages or imperfections, causing clicks, are often asymmetric, i.e., they are more strongly emphasized on the left or right edge of the groove, or they affect only one side of the groove. Since the typical stereo recording/playback system is half vertical - half horizontal, i.e., it preserves asymmetry mentioned above(see Fig. 2), restoration can be performed more reliably when two channels are analyzed jointly, simply because the uncorrupted material in one channel may be helpful in detecting and interpolating corrupted samples in the other channel.

II. Signal Identification

The measured stereo audio signal will be denoted by $\mathbf{y}(t) = [y_1(t), y_2(t)]^{\mathrm{T}}$, where $t = \ldots, -1, 0, 1, \ldots$, denotes normalized (dimensionless) discrete time and $y_1(t)/y_2(t)$ denotes the left/right audio track.

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 $\mathbf{s}(t) = [s_1(t), s_2(t)]^{\mathrm{T}}$ denotes the clean audio signal and $\boldsymbol{\delta}(t) = [\delta_1(t), \delta_2(t)]^{\mathrm{T}}$ is a signal made up of sparsely distributed noise pulses (such as clicks, pops and record scratches). To keep the analysis simple, we will assume that the measured signal is not contaminated with an additive wideband noise (the so-called surface noise), i.e., that impulsive noise is the only disturbance that should be eliminated.

The clean audio signal will be modeled as a two-dimensional vector autoregressive (VAR) process of order r [14], [15], [16]

$$\mathbf{s}(t) = \sum_{i=1}^{r} \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{n}(t) \tag{2}$$

where

$$\mathbf{A}_i = \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_{1i}^{\mathrm{T}} \\ \boldsymbol{\alpha}_{2i}^{\mathrm{T}} \end{bmatrix}, \quad i = 1, \ldots, r$$

are the 2×2 matrices of AR coefficients and $\{\mathbf{n}(t)\}$, $\mathbf{n}(t) = [n_1(t), n_2(t)]^{\mathrm{T}}$, denotes two-dimensional zero-mean white noise with a covariance matrix

$$\mathrm{cov}[\mathbf{n}(t)] = \begin{bmatrix} \rho_1^2 & \rho_{12} \\ \rho_{12} & \rho_2^2 \end{bmatrix} = \boldsymbol{\rho}.$$

Denote by $\boldsymbol{\theta}_j = [\boldsymbol{\alpha}_{j1}^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_{jr}^{\mathrm{T}}]^{\mathrm{T}}$ the vector of coefficients characterizing the j-th channel, and by $\boldsymbol{\varphi}(t) = [\mathbf{y}^{\mathrm{T}}(t-1), \ldots, \mathbf{y}^{\mathrm{T}}(t-r)]^{\mathrm{T}}$ – the corresponding regression vector (the same for both channels). Denote by $\mathbf{0}_r$ the $2r \times 1$ null vector, and by $\mathbf{O}_r$ and $\mathbf{I}_r$ – the $2r \times 2r$ null and identity matrices, respectively. Furthermore, let

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}(t) & \mathbf{0}_r \\ \mathbf{0}_r & \boldsymbol{\varphi}(t) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}.$$

Using these shorthands, in the absence of noise pulses the model (2) can be rewritten in the form

$$\mathbf{y}(t) = \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\theta} + \mathbf{n}(t). \tag{3}$$

Finally, to account for nonstationarity of audio signals, the following time-varying version of (3) will be used

$$\mathbf{y}(t) = \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + \mathbf{n}(t), \quad \mathrm{cov}[\mathbf{n}(t)] = \boldsymbol{\rho}(t) \tag{4}$$

where $\boldsymbol{\theta}(t)$ denotes the slowly time varying vector of autoregressive coefficients, and $\boldsymbol{\rho}(t)$ is the time-variant noise covariance matrix. Estimation (tracking) of $\boldsymbol{\theta}(t)$ will be carried out using the method of exponentially weighted least squares (EWLS), namely

$$\widehat{\boldsymbol{\theta}}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{t} \lambda^{t-k} \parallel \mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\boldsymbol{\theta} \parallel^2 \tag{5}$$

where λ, $0 < \lambda < 1$, denotes the so-called forgetting constant which decides upon the effective estimation memory of the EWLS estimator, given by

$$l(t) = \sum_{k=1}^{t} \lambda^{t-k} = \frac{1-\lambda^t}{1-\lambda}. \tag{6}$$

The value of λ should be chosen so as to trade off the bias and variance components of the mean-squared parameter tracking error $\mathrm{E}[\parallel \boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t) \parallel^2]$. Short-memory algorithms are "fast" (yield small tracking bias) but "inaccurate" (yield large tracking variance), whereas long-memory algorithms are "slow" but "accurate". The best results are obtained if the estimation memory of a tracking algorithm "matches" the degree of nonstationarity of the identified process [17], [18].

Evaluation of (5) is straightforward and leads to

$$\widehat{\boldsymbol{\theta}}(t) = \left[\sum_{k=1}^{t} \lambda^{t-k} \boldsymbol{\Phi}(k)\boldsymbol{\Phi}^{\mathrm{T}}(k)\right]^{-1} \left[\sum_{k=1}^{t} \lambda^{t-k} \boldsymbol{\Phi}(k)\mathbf{y}(k)\right]. \tag{7}$$

Due to the block-diagonal structure of $\boldsymbol{\Phi}(k)$, the estimator (7) can be rewritten in a decomposed form as

$$\widehat{\boldsymbol{\theta}}_j(t) = \left[\sum_{k=1}^{t} \lambda^{t-k} \boldsymbol{\varphi}(k)\boldsymbol{\varphi}^{\mathrm{T}}(k)\right]^{-1} \left[\sum_{k=1}^{t} \lambda^{t-k} \boldsymbol{\varphi}(k) y_j(k)\right] \quad j = 1, 2. \tag{8}$$

A similar technique can be used to track the covariance matrix of the forming noise $\mathbf{n}(t)$. The local estimate of $\boldsymbol{\rho}(t)$ can be obtained from

$$\widehat{\boldsymbol{\rho}}(t) = \frac{\mathbf{R}(t)}{l(t)} \tag{9}$$

where $\mathbf{R}(t)$ denotes the exponentially weighted sum of "squared" residual errors

$$\mathbf{R}(t) = \sum_{k=1}^{t} \lambda^{t-k} \left[\mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\widehat{\boldsymbol{\theta}}(t)\right] \times \times \left[\mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\widehat{\boldsymbol{\theta}}(t)\right]^{\mathrm{T}}. \tag{10}$$

A. Recursive Estimation Algorithms

Both $\widehat{\boldsymbol{\theta}}(t)$ and $\widehat{\boldsymbol{\rho}}(t)$ can be computed recursively. The recursive algorithm for computation of $\widehat{\boldsymbol{\theta}}_j(t)$ has a well-known form [17]

$$\begin{aligned} \varepsilon_j(t|t-1) &= y_j(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_j(t-1) \\ \widehat{\boldsymbol{\theta}}_j(t) &= \widehat{\boldsymbol{\theta}}_j(t-1) + \mathbf{k}(t)\varepsilon_j(t|t-1) \\ j &= 1, 2 \\ \mathbf{k}(t) &= \frac{\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)}{\lambda + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)} \\ \mathbf{Q}(t) &= \frac{1}{\lambda}\left[\mathbf{I}_r - \mathbf{k}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]\mathbf{Q}(t-1). \end{aligned} \tag{11}$$

Note that the gain vector $\mathbf{k}(t)$ does not depend on j, i.e., it is the same for both channels. Recursive computation of $\widehat{\boldsymbol{\rho}}(t)$ is based on the following relationships

$$l(t) = \lambda l(t-1) + 1 \tag{12}$$

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{\boldsymbol{\varepsilon}(t|t-1)\boldsymbol{\varepsilon}^{\mathrm{T}}(t|t-1)}{\lambda + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{Q}(t-1)\boldsymbol{\varphi}(t)} \tag{13}$$

where $\boldsymbol{\varepsilon}(t|t-1) = [\varepsilon_1(t|t-1), \varepsilon_2(t|t-1)]^{\mathrm{T}}$.

B. Relationship to Maximum Likelihood Estimation

Suppose that the identified process is stationary, i.e., that it obeys (3). Under Gaussian assumptions the statistically efficient estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, given $\mathcal{Y}(t) = \{\mathbf{y}(1), \ldots, \mathbf{y}(t)\}$, can be obtained from

$$\{\boldsymbol{\theta}^*(t), \boldsymbol{\rho}^*(t)\} = \arg\max_{\{\boldsymbol{\theta}, \boldsymbol{\rho}\}} p(\mathcal{Y}(t)|\mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho})$$

where $\mathcal{Y}_0 = \{\mathbf{y}(1-r), \ldots, \mathbf{y}(0)\}$ denotes the set of initial conditions and

$$p(\mathcal{Y}(t)|\mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{k=1}^{t} p(\mathbf{y}(k)|\mathbf{y}(k-1), \ldots, \mathbf{y}(1), \mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho})$$
$$= (2\pi)^{-t} \{\det[\boldsymbol{\rho}]\}^{-t/2} \times$$
$$\times \exp\left\{-\frac{1}{2}\sum_{k=1}^{t} \| \mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\boldsymbol{\theta} \|^2_{\boldsymbol{\rho}^{-1}}\right\}$$

denotes the so-called conditional likelihood function. The resulting conditional maximum likelihood (CML) estimators can be evaluated iteratively using the following relationships [14]

$$\boldsymbol{\theta}^*_{i+1}(t) = \left\{\sum_{k=1}^{t} \boldsymbol{\Phi}(k) \left[\boldsymbol{\rho}^*_i(t)\right]^{-1} \boldsymbol{\Phi}^{\mathrm{T}}(k)\right\}^{-1} \times$$
$$\times \left\{\sum_{k=1}^{t} \boldsymbol{\Phi}(k) \left[\boldsymbol{\rho}^*_i(t)\right]^{-1} \mathbf{y}(k)\right\}$$
$$\boldsymbol{\rho}^*_{i+1}(t) = \frac{1}{t}\sum_{k=1}^{t} \left[\mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\boldsymbol{\theta}^*_{i+1}(t)\right] \times$$
$$\times \left[\mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\boldsymbol{\theta}^*_{i+1}(t)\right]^{\mathrm{T}}$$
$$\boldsymbol{\theta}^*(t) = \lim_{i\to\infty} \boldsymbol{\theta}^*_i(t), \quad \boldsymbol{\rho}^*(t) = \lim_{i\to\infty} \boldsymbol{\rho}^*_i(t).$$

Kashyap and Rao [14] have proved that in the special case where all channels share the same regression vector (which is the situation considered here), the CML estimators $\boldsymbol{\theta}^*(t)$ and $\boldsymbol{\rho}^*(t)$ coincide with the LS (least squares) estimators $\widehat{\boldsymbol{\theta}}(t)$ and $\widehat{\boldsymbol{\rho}}(t)$ obtained from (5) after setting $\lambda = 1$ – see Theorem 6a.1 in [14]. This is an intriguing result since, according to (8), the estimator $\widehat{\boldsymbol{\theta}}_j(t)$ is obtained by considering only the j-th equation in (3) without reference to the other equation. The collection of such "decoupled" estimators yields the CML estimator of $\boldsymbol{\theta}$.

When process coefficients are time-varying, they can be tracked using the finite-memory variant of the CML estimator, obtained by maximizing the following exponentially weighted likelihood function

$$\prod_{k=1}^{t} \left[p(\mathbf{y}(k)|\mathbf{y}(k-1), \ldots, \mathbf{y}(1), \mathcal{Y}_0, \boldsymbol{\theta}, \boldsymbol{\rho})\right]^{\lambda^{t-k}}$$
$$= (2\pi)^{-l(t)} \{\det[\boldsymbol{\rho}]\}^{-l(t)/2} \times$$
$$\times \exp\left\{-\frac{1}{2}\sum_{k=1}^{t} \lambda^{t-k} \| \mathbf{y}(k) - \boldsymbol{\Phi}^{\mathrm{T}}(k)\boldsymbol{\theta} \|^2_{\boldsymbol{\rho}^{-1}}\right\}.$$

Since the equivalence proof given in [14] can be easily extended to such exponentially weighted conditional maximum likelihood (EWCML) estimators, the EWLS estimators (8) and (9) can be also regarded as EWCML estimators.

C. Estimation in the Presence of Outliers

The estimates EWLS were obtained under the assumption that $\boldsymbol{\delta}(t) \equiv \mathbf{0}$, i.e., that the measured signal is free of impulsive disturbances. A simple modification will be used to make it work in the presence of noise pulses. Denote by $\mathbf{d}(t) = [d_1(t), d_2(t)]^{\mathrm{T}}$ the pulse location function

$$d_j(t) = \begin{cases} 0 & \text{if} \quad \delta_j(t) = 0 \\ 1 & \text{if} \quad \delta_j(t) \neq 0 \end{cases}, \quad j = 1, 2$$

and by $\widehat{\mathbf{d}}(t) = [\widehat{d}_1(t), \widehat{d}_2(t)]^{\mathrm{T}}$ – the output of the outlier detector (which will be described later)

$$\widehat{d}_j(t) = \begin{cases} 0 & \text{if noise pulse not detected} \\ 1 & \text{if noise pulse detected} \end{cases}, \quad j = 1, 2.$$

To make parameter estimates insensitive to noise pulses, estimation of model parameters is stopped each time when detection alarm is raised, i.e., when $\widehat{\mathbf{d}}(t) \neq \mathbf{0}$. Estimation is resumed once the reconstruction of the questioned fragment is finished (using interpolated samples in place of the corrupted ones).

D. Fair Comparison of VAR and AR Models

Since the purpose of this paper is to compare detection/reconstruction results based on vector signal modeling with those obtained using scalar modeling, one must be sure that, under time-invariant conditions, the corresponding vector/scalar signal identification algorithms have the same estimation capabilities – otherwise one would compare "apples with oranges" [17].

As a measure of the algorithm's estimation capability one can adopt the variance of the excess prediction error. Consider a stationary VAR signal governed by (2). Note that the one-step-ahead prediction error yielded by the EWLS tracker (11) can be written down in the form

$$\varepsilon_j(t+1|t) = \eta_j(t+1|t) + n_j(t), \quad j = 1, 2$$

where

$$\eta_j(t+1|t) = \boldsymbol{\varphi}^{\mathrm{T}}(t+1)[\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j(t)]$$

denotes the so-called excess prediction error, i.e., this component of the prediction error which can be solely attributed to parameter estimation errors.

When the parameter tracking algorithm has a "sufficiently long" estimation memory, the variance of the excess prediction error can be approximately evaluated using the averaging technique – since variations of the parameter estimation errors $\widehat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j$ are in the case considered much slower than variations of the components of the regression vector $\boldsymbol{\varphi}(t)$, it holds that $\overline{\eta^2_j(t+1|t)} \cong [\widehat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]^{\mathrm{T}}\, \overline{\boldsymbol{\varphi}(t+1)\boldsymbol{\varphi}^{\mathrm{T}}(t+1)}\, [\widehat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]$ where $\overline{(\cdot)}$ denotes local time averaging. This leads to the following approximation

$$\mathrm{E}[\eta^2_j(t+1|t)] \cong \mathrm{E}\left\{[\widehat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]^{\mathrm{T}} \boldsymbol{\Phi}_0 [\widehat{\boldsymbol{\theta}}_j(t) - \boldsymbol{\theta}_j]\right\}$$
$$= \mathrm{tr}\left[\mathrm{cov}[\widehat{\boldsymbol{\theta}}_j(t)] \boldsymbol{\Phi}_0\right]$$

where $\boldsymbol{\Phi}_0 = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]$. Furthermore, since it holds that [17]

$$\lim_{t\to\infty} \mathrm{cov}[\widehat{\boldsymbol{\theta}}_j(t)] \cong \frac{1-\lambda}{1+\lambda}\, \rho^2_j \boldsymbol{\Phi}_0^{-1}$$

the steady state value of $\mathrm{E}[\eta_j^2(t+1|t)]$ can be expressed in the form

$$\lim_{t\to\infty} \mathrm{E}[\eta_j^2(t+1|t)] \cong \frac{2r(1-\lambda)}{1+\lambda}\,\rho_j^2. \tag{14}$$

In order to obtain the analogous formula in the case where both audio channels are modeled separately, suppose that $s_1(t)$ and $s_2(t)$ are stationary AR signals governed by

$$\begin{aligned} s_j(t) &= \sum_{i=1}^{r} b_{ji}s_j(t-i) + n_j(t) \\ &= \boldsymbol{\psi}_j^{\mathrm{T}}(t)\boldsymbol{\beta}_j + n_j(t), \quad j=1,2 \end{aligned} \tag{15}$$

where $\boldsymbol{\beta}_j = [b_{j1},\ldots,b_{jr}]^{\mathrm{T}}$ denotes the vector of AR coefficients and $\boldsymbol{\psi}_j(t) = [s_j(t-1),\ldots,s_j(t-r)]^{\mathrm{T}}$ denotes the corresponding regression vector. Note that the VAR model (2) reduces down to (15) if all off-diagonal elements of the matrices $A_i, i=1,\ldots,r$, are equal to zero. Suppose that the parameter vector $\boldsymbol{\beta}_j$ is estimated/tracked using the scalar version of the EWLS algorithm

$$\widehat{\boldsymbol{\beta}}_j(t) = \arg\min_{\boldsymbol{\beta}} \sum_{k=1}^{t} \lambda_j^{t-k}\left[y_j(k) - \boldsymbol{\psi}_j^{\mathrm{T}}(t)\boldsymbol{\beta}\right]^2 \tag{16}$$

where λ_j, $0<\lambda_j<1$, denotes forgetting constant used for identification of the j-th track. Using the averaging technique, one can show that

$$\begin{aligned} \mathrm{E}[\eta_j^2(t+1|t)] &\cong \mathrm{tr}\left[\mathrm{cov}[\widehat{\boldsymbol{\beta}}_j(t)]\boldsymbol{\Psi}_j\right] \\ \lim_{t\to\infty}\mathrm{cov}[\widehat{\boldsymbol{\beta}}_j(t)] &\cong \frac{1-\lambda_j}{1+\lambda_j}\,\rho_j^2\boldsymbol{\Psi}_j^{-1} \end{aligned}$$

where $\boldsymbol{\Psi}_j = \mathrm{E}[\boldsymbol{\psi}_j(t)\boldsymbol{\psi}_j^{\mathrm{T}}(t)]$.

This leads to the following formula

$$\lim_{t\to\infty} \mathrm{E}[\eta_j^2(t+1|t)] \cong \frac{r(1-\lambda_j)}{1+\lambda_j}\,\rho_j^2 \tag{17}$$

which should be compared with (14). Requiring that the variance of the excess prediction errors should be in both cases the same, one arrives at the following condition of "fair comparison"

$$\frac{2(1-\lambda)}{1+\lambda} = \frac{1-\lambda_j}{1+\lambda_j}\ . \tag{18}$$

Since, under normal operating conditions, the forgetting constants λ and λ_j are close to one, i.e., $1+\lambda \cong 1+\lambda_j \cong 2$, the condition (18) is approximately equivalent to

$$l(\infty) \cong 2l_j(\infty) \tag{19}$$

where $l(\infty) = 1/(1-\lambda)$ and $l_j(\infty) = 1/(1-\lambda_j)$ denote the steady state values of the effective memory spans of the VAR and AR trackers, respectively. Note that since in the vector case the number of estimated coefficients is equal to $2r$ per one audio track, i.e., it is two times larger than the analogous quantity in the scalar case, under the condition (19) the average effective number of samples used to estimate one model coefficient is in both cases the same.

III. Detection of noise pulses and signal interpolation

A. State space problem formulation

We will start from solving a simpler problem of recovering an isolated block of m irrevocably distorted samples of a stationary AR process governed by (2). The block, which starts at the instant t_0+1 and ends at the instant t_0+m (i.e., $\mathbf{d}(t_0+1) = \ldots = \mathbf{d}(t_0+m) = \mathbf{1}$, where $\mathbf{1} = [1,1]^{\mathrm{T}}$), is preceded and succeeded by undistorted samples (i.e., $\mathbf{d}(t) = \mathbf{0}$ for $t \le t_0$ and $t > t_0+m$). We will assume that the location of the sequence of noise pulses is known exactly [i.e., $\widehat{\mathbf{d}}(t) \equiv \mathbf{d}(t)$]. We will also assume that noise pulses $\boldsymbol{\delta}(t_0+1)$, $\ldots, \boldsymbol{\delta}(t_0+m)$ can be modeled as a sequence of mutually uncorrelated Gaussian variables, independent of $\{\mathbf{n}(t)\}$, with known covariance matrices

$$\boldsymbol{\Delta}(t) = \mathrm{cov}[\boldsymbol{\delta}(t)], \quad t_0+1 \le t \le t_0+m.$$

The solution, based on Kalman filtering [19], will be a starting point for derivation of a more realistic algorithm combining adaptive detection of arbitrarily shaped noise pulses with AR-model based signal interpolation.

To design Kalman filter we need a state space equivalent of the input-output description (1)-(2). Let $q = 2r+m$. Define the $2q\times 1$ state vector $\mathbf{x}_q(t) = [\mathbf{s}^{\mathrm{T}}(t),\ldots,\mathbf{s}^{\mathrm{T}}(t-q+1)]^{\mathrm{T}}$ made up of the q most recent signal samples.

The overdetermined state space model of (1)-(2) can be written down in the augmented companion form [to describe (1)-(2), it is sufficient to set $q=r$; the adopted higher-order (non-minimal) model is needed to solve the signal interpolation problem].

$$\begin{aligned} \mathbf{x}_q(t+1) &= \boldsymbol{\mathcal{A}}_q\mathbf{x}_q(t) + \boldsymbol{\mathcal{C}}_q\mathbf{n}(t+1) \\ \mathbf{y}(t) &= \boldsymbol{\mathcal{C}}_q^{\mathrm{T}}\mathbf{x}_q(t) + \boldsymbol{\delta}(t) \end{aligned} \tag{20}$$

where

$$\boldsymbol{\mathcal{A}}_q = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_r & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \mathbf{I} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} & & \mathbf{O} & \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} \\ \vdots & & & & & \ddots & & \vdots \\ \mathbf{O} & \mathbf{O} & & \mathbf{O} & \mathbf{O} & \ldots & \mathbf{I} & \mathbf{O} \end{bmatrix}$$

is the $2q\times 2q$ state transition matrix and $\boldsymbol{\mathcal{C}}_q = [\mathbf{I},\mathbf{O},\ldots,\mathbf{O}]^{\mathrm{T}}$ denotes the $2q\times 2$ output matrix, and $\mathbf{O} = \mathbf{O}_1$ and $\mathbf{I} = \mathbf{I}_1$ denote 2×2 null and identity matrices, respectively.

Based on (20) and on the available prior knowledge, the Kalman filter (KF) recursions can be written down as follows

$$\begin{aligned} \widehat{\mathbf{x}}_q(t|t-1) &= \boldsymbol{\mathcal{A}}_q\widehat{\mathbf{x}}_q(t-1|t-1) \\ \mathbf{P}_q(t|t-1) &= \boldsymbol{\mathcal{A}}_q\mathbf{P}_q(t-1|t-1)\boldsymbol{\mathcal{A}}_q^{\mathrm{T}} + \boldsymbol{\mathcal{C}}_q\boldsymbol{\rho}\boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \\ \mathbf{e}(t) &= \mathbf{y}(t) - \boldsymbol{\mathcal{C}}_q^{\mathrm{T}}\widehat{\mathbf{x}}_q(t|t-1) \\ \mathbf{S}(t) &= \boldsymbol{\mathcal{C}}_q^{\mathrm{T}}\mathbf{P}_q(t|t-1)\boldsymbol{\mathcal{C}}_q + \boldsymbol{\Delta}(t) \\ \mathbf{L}_q(t) &= \mathbf{P}_q(t|t-1)\boldsymbol{\mathcal{C}}_q\mathbf{S}^{-1}(t) \\ \widehat{\mathbf{x}}_q(t|t) &= \widehat{\mathbf{x}}_q(t|t-1) + \mathbf{L}_q(t)\mathbf{e}(t) \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - \mathbf{L}_q(t)\mathbf{S}(t)\mathbf{L}_q^{\mathrm{T}}(t). \end{aligned} \tag{21}$$

Since we have assumed that $\boldsymbol{\delta}(t) = \mathbf{0}$ for $t \leq t_0$, the algorithm should be started at the instant $t_0 + 1$, with initial conditions $\widehat{\mathbf{x}}_q(t_0|t_0) = [\mathbf{y}^{\mathrm{T}}(t_0), \ldots, \mathbf{y}^{\mathrm{T}}(t_0 - q + 1)]^{\mathrm{T}}$, $\mathbf{P}_q(t_0|t_0) = \mathbf{O}_q$, and stopped at the instant $t_0 + m + r$, after reading r undisturbed signal samples at the end of the corrupted fragment. The filtered state vector at the termination point t_0+m+r has the form $\widehat{\mathbf{x}}_q(t_0+m+r|t_0+m+r) = [\mathbf{y}(t_0+m+r), \ldots, \mathbf{y}(t_0+m+1), \widehat{\mathbf{s}}(t_0+m), \ldots, \widehat{\mathbf{s}}(t_0+1), \mathbf{y}(t_0), \ldots, \mathbf{y}(t_0-r+1)]^{\mathrm{T}}$ where $\widehat{\mathbf{s}}(t_0+1), \ldots, \widehat{\mathbf{s}}(t_0+m)$ is the block of interpolated samples. Since, in the case considered, the signal estimates yielded by the Kalman algorithm do not depend on measurements collected at instants $t_0 + m + r + 1, t_0 + m + r + 2$, etc., there is no point in continuing operation of the Kalman filter after reaching the point $t_0 + m + r$.

B. Signal Prediction and Detection of Noise Pulses

Similar to [6], our pulse detection scheme will be based on monitoring signal prediction errors. In the univariate (mono) case considered in [6], where the signal is governed by

$$s(t) = \sum_{i=1}^{r} a_i s(t-i) + n(t), \quad \mathrm{var}[n(t)] = \rho$$

detection alarm is raised at the instant t_0+1 (i.e., $\widehat{d}(t_0+1)$ is set to 1) if the magnitude of the one-step-ahead signal prediction error $\varepsilon(t_0 + 1|t_0) = y(t_0 + 1) - \boldsymbol{\varphi}^{\mathrm{T}}(t_0 + 1)\boldsymbol{\theta}$, where $\boldsymbol{\varphi}(t) = [y(t-1), \ldots, y(t-r)]^{\mathrm{T}}$ and $\boldsymbol{\theta} = [a_1, \ldots, a_r]^{\mathrm{T}}$, exceeds μ times its standard deviation

$$|\varepsilon(t_0 + 1|t_0)| > \mu\sigma_\varepsilon(t_0 + 1|t_0) \tag{22}$$

where $\sigma^2_\varepsilon(t_0+1|t_0) = \rho$ and μ is a constant multiplier, usually chosen in the range [3,5][1].

The test is then extended to multi-step-ahead prediction errors. Detection alarm is terminated at the instant $t = t_0 + m$ if r consecutive prediction errors are sufficiently small, namely if

$$\begin{gathered} |\varepsilon(t|t_0)| \leq \mu\sigma_\varepsilon(t|t_0) \\ t = t_0 + m + 1, \ldots, t_0 + m + r \end{gathered} \tag{23}$$

or if $t - t_0$ reaches its maximum allowable value $m_{\max}$. The output of the outlier detector is in this case equal to: $\widehat{d}(t_0 + 1) = \ldots = \widehat{d}(t_0 + m) = 1$, $\widehat{d}(t_0 + m + 1) = \ldots = \widehat{d}(t_0 + m + r) = 0$.

The detection technique briefly summarized above can be extended to the multivariable case. The detection triggering condition (22) has the following multivariate equivalent

$$\boldsymbol{\varepsilon}^{\mathrm{T}}(t_0 + 1|t_0)\boldsymbol{\Sigma}^{-1}_\varepsilon(t_0 + 1|t_0)\boldsymbol{\varepsilon}(t_0 + 1|t_0) > \mu^2 \tag{24}$$

where $\boldsymbol{\Sigma}_\varepsilon(t_0+1|t_0) = \boldsymbol{\rho}$ denotes the covariance matrix of the one-step-ahead prediction error. The stopping condition (23) can be reformulated in an analogous way

$$\begin{gathered} \boldsymbol{\varepsilon}^{\mathrm{T}}(t|t_0)\boldsymbol{\Sigma}^{-1}_\varepsilon(t|t_0)\boldsymbol{\varepsilon}(t|t_0) \leq \mu^2 \\ t = t_0 + m + 1, \ldots, t_0 + m + r \end{gathered} \tag{25}$$

[1]When μ is set to 3, condition (22) is usually referred to as "3-sigma" outlier detection rule.

where $\boldsymbol{\varepsilon}(t|t_0)$ denotes the $(t-t_0)$-step-ahead signal prediction error and $\boldsymbol{\Sigma}_\varepsilon(t|t_0)$ denotes the corresponding error covariance matrix. Both quantities can be easily computed using the Kalman filtering algorithm (21). In order to do this, one should set $q = 2r + m_{\max}$ and enforce

$$\boldsymbol{\Delta}(t) = \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}, \quad \gamma \to \infty, \quad \text{for} \quad t > t_0. \tag{26}$$

The latter condition means that the samples $\mathbf{y}(t_0 + 1), \mathbf{y}(t_0 + 2), \ldots$ should be regarded as corrupted with infinite-variance noise and – as such – completely eliminated from the estimation process. It is easy to check that in the case considered $\mathbf{S}^{-1}(t) = \mathbf{O}$, which results in $\widehat{\mathbf{x}}_q(t|t) = \widehat{\mathbf{x}}_q(t|t-1)$, $\mathbf{P}_q(t|t) = \mathbf{P}_q(t|t-1)$ for all $t > t_0$. Under such conditions Kalman filter works as a multi-step-ahead predictor yielding $\boldsymbol{\varepsilon}(t|t_0) = \mathbf{e}(t)$ and

$$\begin{aligned} \boldsymbol{\Sigma}_\varepsilon(t|t_0) &= \mathrm{cov}[\mathbf{e}(t)] = \boldsymbol{\mathcal{C}}_q^{\mathrm{T}}\mathbf{P}_q(t|t-1)\boldsymbol{\mathcal{C}}_q \\ &= \begin{bmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{bmatrix} = \boldsymbol{\Sigma}(t). \end{aligned} \tag{27}$$

Unfortunately, the solution presented above, does not allow one to differentiate between audio channels (both tracks are analyzed jointly) and for this reason it is not suitable for our purposes. We will replace it with the following decoupled decision rule

$$\widehat{d}_j(t) = \begin{cases} 0 & \text{if} \quad |e_j(t)| \leq \mu\sigma_j(t) \\ 1 & \text{if} \quad |e_j(t)| > \mu\sigma_j(t) \end{cases}, \quad j = 1, 2 \tag{28}$$

and more selective noise covariance scheduling

$$\boldsymbol{\Delta}(t) = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 0 \\ \begin{bmatrix} 0 & 0 \\ 0 & \gamma \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 0 \wedge \widehat{d}_2(t) = 1 \\ \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 1 \wedge \widehat{d}_2(t) = 0 \\ \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 1 \end{cases} \\ \gamma \to \infty. \tag{29}$$

It is straightforward to check that under (29) the corresponding values of $\mathbf{S}^{-1}(t)$ are given by

$$\mathbf{S}^{-1}(t) = \begin{cases} \boldsymbol{\Sigma}^{-1}(t) & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 0 \\ \begin{bmatrix} \frac{1}{\sigma_1^2(t)} & 0 \\ 0 & 0 \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 0 \wedge \widehat{d}_2(t) = 1 \\ \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma_2^2(t)} \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 1 \wedge \widehat{d}_2(t) = 0 \\ \mathbf{O} & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 1 \end{cases} \tag{30}$$

allowing one to: accept both components of $\mathbf{y}(t) = [y_1(t), y_2(t)]^{\mathrm{T}}$ if both channels are regarded as outlier-free [$\widehat{d}_1(t) = \widehat{d}_2(t) = 0$], reject $y_2(t)$ if only the second channel is corrupted [$\widehat{d}_1(t) = 0 \wedge \widehat{d}_2(t) = 1$], reject $y_1(t)$ if only the first channel is corrupted [$\widehat{d}_1(t) = 1 \wedge \widehat{d}_2(t) = 0$], or reject both components of $\mathbf{y}(t)$ if both channels are corrupted [$\widehat{d}_1(t) = \widehat{d}_2(t) = 1$].

C. Adaptive Detection and Interpolation

The adaptive version of the detection/interpolation procedure described above can be obtained by combining the KF algorithm (21) with the EWLS algorithm (11) - (13), i.e., by replacing the true model parameters $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, which were previously assumed to be constant and known, with their most recent estimates $\widehat{\boldsymbol{\theta}}(t)$ and $\widehat{\boldsymbol{\rho}}(t)$, respectively. According to [13], the AR-model based reconstruction of samples called in question by the outlier detector can be carried out independently – without any information loss – for each local analysis frame starting and ending with r undistorted samples $\mathbf{y}(t)$. For this reason we will focus our attention on a single detection episode which starts when at least one of two prediction errors evaluated for a stereo signal takes an excessive value, and ends when r consecutive prediction errors take for both channels sufficiently small values.

Suppose that the outlier detector is triggered at the instant t_0+1, i.e.,

$$|\varepsilon_j(t_0+1|t_0)| = |y_j(t_0+1) - \boldsymbol{\varphi}^{\mathrm{T}}(t_0+1)\widehat{\boldsymbol{\theta}}_j(t_0)| > \mu\widehat{\rho}_j(t_0) \quad \text{for } j=1 \text{ and/or } 2. \tag{31}$$

Once this happens, the parameter tracking procedure is temporarily stopped, and the KF-based detection procedure, described earlier, is started. However, we will introduce an important modification – the fixed-order Kalman filter will be replaced with the variable-order one. Such modification is possible due to the special structure of the matrices $\boldsymbol{\mathcal{A}}_q$, $\boldsymbol{\mathcal{C}}_q$ and $\mathbf{P}_q(t_0|t_0)$ incorporated in (21). Taking advantage of this structure, one can show that the order of the Kalman filter (21) can be – without affecting estimation results – gradually increased, starting from $r+1$, until the stopping condition is met. The variable-order Kalman filter offers significant computational savings over its fixed-order ($q = q_{\max} = 2r + m_{\max}$) version.

D. Algorithm

Denote by $\widehat{\boldsymbol{\Theta}}_r(t) = [\widehat{\boldsymbol{\theta}}_1(t)|\widehat{\boldsymbol{\theta}}_2(t)]$ the $2r \times 2$ matrix made up of the estimated process coefficients, and by

$$\widehat{\boldsymbol{\Theta}}_q(t) = \begin{bmatrix} \widehat{\boldsymbol{\theta}}_1(t) & \widehat{\boldsymbol{\theta}}_2(t) \\ \mathbf{0}_{q-r} & \mathbf{0}_{q-r} \end{bmatrix}, \quad q > r$$

– the analogous matrix extended with zeros. Denote by $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ the vectors made up of the first column and the second column of the matrix $\mathbf{X}$, respectively. Denote by $\mathbf{X}^{(1,2)}$ the matrix made up of the first two columns of $\mathbf{X}$. Finally, let $q(t) = r+t-t_0$. The adaptive algorithm which combines (21) with (28)-(29) can be summarized as follows:

Initialization

$$\widehat{\mathbf{x}}_r(t_0|t_0) = [\mathbf{y}^{\mathrm{T}}(t_0), \ldots, \mathbf{y}^{\mathrm{T}}(t_0-r+1)]^{\mathrm{T}}$$
$$\mathbf{P}_r(t_0|t_0) = \mathbf{O}_r$$

Time update step ($t \geq t_0+1$)

$$\widehat{\mathbf{y}}(t|t-1) = \widehat{\boldsymbol{\Theta}}^{\mathrm{T}}_{q(t)-1}(t_0)\widehat{\mathbf{x}}_{q(t)-1}(t-1|t-1)$$
$$\mathbf{e}(t) = \mathbf{y}(t) - \widehat{\mathbf{y}}(t|t-1) = [e_1(t), e_2(t)]^{\mathrm{T}}$$
$$\widehat{\mathbf{x}}_{q(t)}(t|t-1) = \begin{bmatrix} \widehat{\mathbf{y}}(t|t-1) \\ \widehat{\mathbf{x}}_{q(t)-1}(t-1|t-1) \end{bmatrix}$$
$$\mathbf{H}_{q(t)-1}(t) = \mathbf{P}_{q(t)-1}(t-1|t-1)\widehat{\boldsymbol{\Theta}}_{q(t)-1}(t_0)$$
$$\boldsymbol{\Sigma}(t) = \widehat{\boldsymbol{\Theta}}^{\mathrm{T}}_{q(t)-1}(t_0)\mathbf{H}_{q(t)-1}(t) + \widehat{\boldsymbol{\rho}}(t_0) = \begin{bmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{bmatrix}$$
$$\mathbf{P}_{q(t)}(t|t-1) = \begin{bmatrix} \boldsymbol{\Sigma}(t) & \mathbf{H}^{\mathrm{T}}_{q(t)-1}(t) \\ \mathbf{H}_{q(t)-1}(t) & \mathbf{P}_{q(t)-1}(t-1|t-1) \end{bmatrix}$$

Outlier detection step

$$\widehat{d}_j(t) = \begin{cases} 0 & \text{if} \quad |e_j(t)| \leq \mu\sigma_j(t) \\ 1 & \text{if} \quad |e_j(t)| > \mu\sigma_j(t) \end{cases}, \quad j=1,2$$

Measurement update step ($t \geq t_0+1$)

Case 1: if $\widehat{d}_1(t) = \widehat{d}_2(t) = 0$ or $t \geq t_0 + m_{\max}$ then

$$\mathbf{L}_{q(t)}(t) = \mathbf{P}^{(1,2)}_{q(t)}(t|t-1)\boldsymbol{\Sigma}^{-1}(t)$$
$$\widehat{\mathbf{x}}_{q(t)}(t|t) = \widehat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{L}_{q(t)}(t)\mathbf{e}(t)$$
$$\mathbf{P}_{q(t)}(t|t) = \mathbf{P}_{q(t)}(t|t-1) - \mathbf{L}_{q(t)}(t)\boldsymbol{\Sigma}(t)\mathbf{L}^{\mathrm{T}}_{q(t)}(t)$$

Case 2: if $\widehat{d}_1(t) = 0$ and $\widehat{d}_2(t) = 1$ then

$$\mathbf{l}_{q(t)}(t) = \frac{1}{\sigma_1^2(t)}\,\mathbf{P}^{(1)}_{q(t)}(t|t-1)$$
$$\widehat{\mathbf{x}}_{q(t)}(t|t) = \widehat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{l}_{q(t)}(t)e_1(t)$$
$$\mathbf{P}_{q(t)}(t|t) = \mathbf{P}_{q(t)}(t|t-1) - \sigma_1^2(t)\mathbf{l}_{q(t)}(t)\mathbf{l}^{\mathrm{T}}_{q(t)}(t)$$

Case 3: if $\widehat{d}_1(t) = 1$ and $\widehat{d}_2(t) = 0$ then

$$\mathbf{l}_{q(t)}(t) = \frac{1}{\sigma_2^2(t)}\,\mathbf{P}^{(2)}_{q(t)}(t|t-1)$$
$$\widehat{\mathbf{x}}_{q(t)}(t|t) = \widehat{\mathbf{x}}_{q(t)}(t|t-1) + \mathbf{l}_{q(t)}(t)e_2(t)$$
$$\mathbf{P}_{q(t)}(t|t) = \mathbf{P}_{q(t)}(t|t-1) - \sigma_2^2(t)\mathbf{l}_{q(t)}(t)\mathbf{l}^{\mathrm{T}}_{q(t)}(t)$$

Case 4: if $\widehat{d}_1(t) = \widehat{d}_2(t) = 1$ then

$$\widehat{\mathbf{x}}_{q(t)}(t|t) = \widehat{\mathbf{x}}_{q(t)}(t|t-1)$$
$$\mathbf{P}_{q(t)}(t|t) = \mathbf{P}_{q(t)}(t|t-1)$$

E. Comparison with other approaches

Apart from vector processing, which replaced scalar processing, the main difference between the approach summarized above and that described earlier [11] lies in the way the multi-step-ahead signal prediction is carried out. Unlike the open-loop prediction scheme that was used in [11], signal predictions yielded by the Kalman filter algorithm depend not only on samples collected prior to the instant t_0+1, but also on samples that were provisionally accepted afterwards – such

predictions can be called decision-feedback since they depend on detection decisions made earlier. It was observed that the approach based on open-loop prediction shows tendency to raise too short detection alarms, i.e., alarms that end well before the entire pulse waveform is complete. Fig. 3f shows a typical open-loop prediction based detection scenario. Since the primary detection alarm, raised at the beginning of the noise pulse, is terminated too soon, it causes acceptance of r corrupted signal samples. This, in turn, evokes the secondary detection alarm, triggered when outlier detection is resumed after the break. As a result, the reconstructed signal (Fig. 3g) is heavily (and audibly) distorted.

The results improve considerably if the decision-feedback approach is used, since samples provisionally accepted in the middle of detection alarms may significantly decrease the prediction error variance, which increases sensitivity of the outlier detector. Therefore detection alarms raised by the scheme based on decision-feedback predictions are usually longer than those yielded by the open-loop scheme – see Figs. 3h and 3i.

Unlike the prediction-based approaches, the double threshold approach shows tendency to produce overly long detection alarms. It is not difficult to explain this effect. Suppose that the pulse waveform starts at the instant $t_0 + 1$ and ends at the instant $t_0 + k_0 + 1$. Note that, even though the sample $y(t_0 + k_0 + 1)$ is outlier-free, the corresponding value of the residual error usually still remains large as it is evaluated based on r preceding signal samples, at least some of which are contaminated by outliers. It is not until the sample $y(t_0+k_0+r+1)$ is reached, that residual errors are entirely unaffected by the detected noise pulse. As a result, when the adopted order of autoregression is large ($r \geq 10$ is a recommended choice under 44.1 and 48 kHz sampling), the corresponding detection alarms are usually much longer than the "ground truth" ones – see Figs. 3d and 3e.

The common limitation of all schemes compared above is the lack of precision in determining the end points of detection alarms. This drawback, caused by the fact that detection decisions are based on the results of forward-time (i.e., unidirectional) signal analysis, can be alleviated by means of bidirectional processing – see Section IV D.

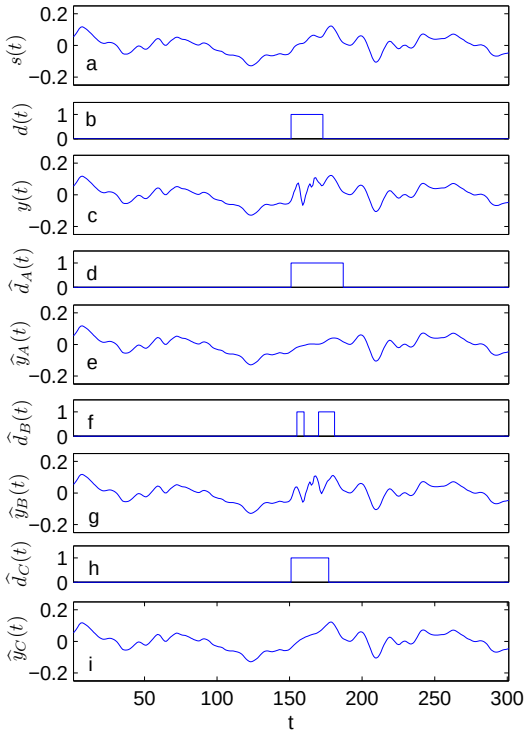

Fig. 3: Comparison of three detection schemes. The corresponding plots show: clean audio signal (a), exact location of the inserted noise pulse (b), corrupted audio signal (c), detection alarm yielded by the double-threshold approach (d) and the corresponding signal reconstruction (e), detection alarms yielded by the open-loop prediction approach (f) and the corresponding signal reconstruction (g), and detection alarm yielded by the decision-feedback prediction approach (h) and the corresponding signal reconstruction (i).

IV. Implementation issues

A. Alternative Noise Covariance Estimation Scheme

When the EWLS algorithm (11) - (13) is used for identification of the VAR model (4), both $\boldsymbol{\theta}(t)$ and $\boldsymbol{\rho}(t)$ are tracked with the same speed/accuracy, determined by the forgetting constant λ. Since experiments, incorporating real audio signals, show that the coefficients of the covariance matrix $\boldsymbol{\rho}(t)$ often vary faster than autoregressive coefficients $\boldsymbol{\theta}(t)$, for outlier detection purposes it may be beneficial to replace (9) with the following exponentially weighted estimate

$$\widehat{\boldsymbol{\rho}}(t) = \lambda_0 \widehat{\boldsymbol{\rho}}(t-1) + (1-\lambda_0)\boldsymbol{\varepsilon}(t|t-1)\boldsymbol{\varepsilon}^{\mathrm{T}}(t|t-1) \quad (32)$$

where λ_0, $0 < \lambda_0 < 1$, is a forgetting constant different from λ. When $\lambda_0 < \lambda$ (which is recommended), the effective estimation memory of the algorithm (32) is smaller than estimation memory of the EWLS tracker, allowing the outlier detector to react faster to sudden changes in $\boldsymbol{\rho}$. However, even though application of the modified covariance estimator (32) yields better detection results, signal interpolation is consistently better when the original EWLS estimator (9) is used. For this reason the best results are obtained when the KF algorithm is run twice: first to detect noise pulses, using (32), and second – to reconstruct the irrevocably distorted samples, using (9).

B. Elimination of Channel Offsets

Adopting the VAR model (2), one implicitly assumes that the modeled signal is zero-mean: $\mathrm{E}[\mathbf{s}(t)] = \mathbf{0}$. Since for a typical stereo audio recording such an assumption is not valid, the problem of non-zero channel offsets should be solved in some way.

The direct solution is to incorporate offsets into the VAR model, i.e., to use the following signal description in place of

(2)

$$\mathbf{s}(t) = \sum_{i=1}^{r} \mathbf{A}_i \mathbf{s}(t-i) + \boldsymbol{\gamma} + \mathbf{n}(t) \tag{33}$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2]^{\mathrm{T}}$, and γ_1, γ_2 denote unknown constants. Since after adopting $\boldsymbol{\theta}_j = [\boldsymbol{\alpha}_{j1}^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_{jr}^{\mathrm{T}}, \gamma_j]^{\mathrm{T}}$ and $\boldsymbol{\varphi}(t) = [\mathbf{y}^{\mathrm{T}}(t-1), \ldots, \mathbf{y}^{\mathrm{T}}(t-r), 1]^{\mathrm{T}}$ the shorthand signal description (3) remains unchanged, identification of the bias-corrected VAR model (33) can be handled in an exactly the same way as described earlier. For the same reason the variable-order KF algorithm does not need any modifications.

The indirect solution to the offset problem is to remove non-zero signal means prior to applying the detection/interpolation procedure. Such a signal "centering" operation can be easily realized by means of passing the signal $\mathbf{y}(t)$ through a high-pass filter $H(q^{-1})$ of the form

$$H(q^{-1}) = \frac{c(1-q^{-1})}{1-cq^{-1}}$$

where c, $0 < c < 1$ denotes a bandwidth-controlling constant which should be sufficiently close to 1. The advantage of the indirect solution, compared to the direct one, is its greater flexibility due to the fact that the constants λ and c can be chosen independently of each other (when the extended VAR model (33) is used, the tracking rate, determined by λ, is the same for all model coefficients).

C. Closing Detection Gaps

One of the consequences of adopting the decentralized detection rule (28), instead of (25), is that detection alarms may not form solid blocks of "ones" preceded and succeeded by at least r "zeros". While detection alarms raised for unipolar noise pulses usually have this property, for bipolar pulses, or pulses of even more complicated shapes, it often happens that the outlier detector accepts a few samples located in the transition zone between the positive and negative peaks of the click waveform – even though such measurements are not reliable. It was observed that such "accidental acceptances" of samples located in the middle of long-lasting artifacts can adversely affect reconstruction results. For this reason it is recommended that all detection gaps of length smaller than r are removed prior to reconstruction. We note that a similar technique was used in [12] to eliminate detection errors caused by "destructive interference", occurring when some residual errors take values close to zero in the middle of long noise pulses.

Of course, each time when a detection alarm is modified, the Kalman filter algorithm should be rerun to incorporate changes.

D. Bidirectional Processing

So far we have assumed that the archive audio signal is analyzed sequentially, forward in time. In such a case a sample is regarded as an outlier if it is "inconsistent" with the signal past, which is indicated by excessive values of prediction errors. When signal characteristics change abruptly, e.g. when an entirely new sound starts to build up, all causal prediction-based detection schemes are prone to generate false detection alarms, calling in question uncorrupted signal samples simply because they do not match the signal past. Since such samples are consistent with the signal "future", rather than its "past", the number of false alarms can be significantly reduced if results of forward-time detection are combined with the analogous results of backward-time detection. The latter can be obtained by means of processing audio signal, using the Kalman filtering algorithm, backward in time (provided, of course, that the entire recording is available). Kalman filter applied to time-reversed data will be further referred to as backward Kalman filter.

The set of local, case-dependent fusion rules that can be used to combine forward and backward detection alarms, denoted respectively by $\widehat{d}_j^f(t)$ and $\widehat{d}_j^b(t)$, was proposed and experimentally verified in [11]. First, the beginning of each forward/backward detection alarm is shifted back by a small fixed number of samples Δt. Then such extended alarms are combined in a way that depends on their mutual configuration called a detection pattern. For example, when forward and backward detection alarms in channel j form solid blocks that at least partially overlap (which is the most frequently encountered detection pattern)

$$\begin{gathered}\widehat{d}_j^f(t) = 1 \quad \text{for} \quad t \in [\underline{t_j^f}, \overline{t_j^f}] \\ \widehat{d}_j^b(t) = 1 \quad \text{for} \quad t \in [\underline{t_j^b}, \overline{t_j^b}] \\ [\underline{t_j^f}, \overline{t_j^f}] \cap [\underline{t_j^b}, \overline{t_j^b}] \neq \oslash\end{gathered}$$

the best results can be obtained using the "front edge / front edge" fusion rule. According to this rule, the combined alarm is started at the instant $\underline{t_j^f}$ corresponding to the front edge of the forward alarm, and terminated at the instant $\overline{t_j^b}$ corresponding to the front edge of the backward alarm (which, after time reversal, becomes its back edge)

$$\widehat{d}_j^{fb}(t) = 1 \quad \text{for} \quad t \in [\underline{t_j^f}, \overline{t_j^b}].$$

Fusion rules applicable to other detection patterns can be found in [11].

Suppose that the combined forward-backward detection alarm starts at the instant $t_0 + 1$, ends at the instant $t_0 + m$, and that it is preceded and succeeded by at least r undistorted samples:

$$\begin{gathered}\widehat{d}_1^{fb}(t_0+1) = 1 \quad \text{or} \quad \widehat{d}_2^{fb}(t_0+1) = 1 \\ \widehat{d}_1^{fb}(t_0+m) = 1 \quad \text{or} \quad \widehat{d}_2^{fb}(t_0+m) = 1 \\ \widehat{d}_1^{fb}(t) = \widehat{d}_2^{fb}(t) = 0 \\ t \in [t_0 - r + 1, t_0] \cup [t_0 + m + 1, t_0 + m + r].\end{gathered}$$

Since the combined alarm differs from its forward/backward components, the samples scheduled for reconstruction should be reestimated. In this case the Kalman filter algorithm is run in a non-adaptive mode, i.e., its operation is not controlled by the internal outlier detector – the aggregated detection sequences $\widehat{d}_1^{fb}(t)$ and $\widehat{d}_2^{fb}(t)$ are used instead.

E. Model Stability Monitoring and Enforcement

In majority of audio applications, including the adaptive detection/reconstruction problem considered in this paper, stability of the signal model must be guaranteed to make the model-based analysis well-posed. The VAR model (2) is asymptotically stable iff all zeros, $z_i, i = 1, \ldots, 2r$, of the characteristic polynomial

$$\mathcal{A}(z^{-1}) = \det\left[\mathbf{I} - \sum_{i=1}^{r} \mathbf{A}_i z^{-i}\right]$$

lie inside the unit circle in the complex plane: $|z_i| < 1, i = 1, \ldots, 2r$.

Unfortunately, when true signal parameters $\mathbf{A}_i$ are replaced with their EWLS estimates $\widehat{\mathbf{A}}_i(t)$, the resulting VAR model is not guaranteed to be stable. For this reason, whenever a detection alarm is raised, the model is checked for stability. If stability conditions are not met, model coefficients are reestimated using the stability-preserving Whittle-Wiggins-Robinson (WWR) algorithm (the multivariate extension of the Levinson-Durbin algorithm) – for a detailed description of this algorithm and discussion of its properties see Complement C8.6 in [20]. The localized version of the WWR algorithm solves for $\widehat{\mathbf{A}}_1(t), \ldots, \widehat{\mathbf{A}}_r(t)$ and $\widehat{\boldsymbol{\rho}}(t)$ the set of Yule-Walker type equations of the form

$$[\mathbf{I}, -\widehat{\mathbf{A}}_1(t), \ldots, -\widehat{\mathbf{A}}_r(t)]\, \widehat{\boldsymbol{\mathcal{R}}}(t) = [\widehat{\boldsymbol{\rho}}(t), \mathbf{O}, \ldots, \mathbf{O}] \qquad (34)$$

where $\mathbf{O}$ denotes the 2×2 null matrix,

$$\widehat{\boldsymbol{\mathcal{R}}}(t) = \begin{bmatrix} \widehat{\mathbf{R}}_0(t) & \widehat{\mathbf{R}}_1(t) & \ldots & \widehat{\mathbf{R}}_r(t) \\ \widehat{\mathbf{R}}_1^{\mathrm{T}}(t) & \widehat{\mathbf{R}}_0(t) & & \vdots \\ \vdots & & & \widehat{\mathbf{R}}_1(t) \\ \widehat{\mathbf{R}}_r^{\mathrm{T}}(t) & \ldots & & \widehat{\mathbf{R}}_0(t) \end{bmatrix}$$

and $\widehat{\mathbf{R}}_k(t), k = 0, \ldots, r$, denote local estimates of the autocovariance matrices $\mathbf{R}_k = \mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-k)]$

$$\widehat{\mathbf{R}}_k(t) = \frac{1}{N} \sum_{i=0}^{N-k-1} \mathbf{y}(t-i)\mathbf{y}^{\mathrm{T}}(t-i-k). \qquad (35)$$

To comply with memory settings of the EWLS algorithm (7), the value of N is set to the equivalent width $k(t) = (1+\lambda)/(1-\lambda)$ of the exponential window [different from its effective width $l(t)$] [17].

The important property of the WWR algorithm is the guaranteed stability of the resultant VAR model provided that the matrix $\widehat{\boldsymbol{\mathcal{R}}}(t)$ is positive definite (which is always the case when the biased estimates (35) are used).

It should be stressed that the WWR algorithm is a "rescue" estimation procedure, used *only* when the EWLS-based model is not stable at the moment of triggering detection alarm (which does not happen frequently). If the WWR algorithm is used permanently, i.e., instead of the EWLS algorithm, the detection/reconstruction results deteriorate due to evidently worse predictive capabilities of the corresponding VAR models. This seems to be the price paid for the guaranteed model stability. On the other hand, if stability monitoring/enforcement is skipped, signal reconstruction errors may occasionally become very large.

F. Numerical Safeguards

When $\widehat{d}_1(t) = 0$ and/or $\widehat{d}_2(t) = 0$, the applied covariance scheduling (30) puts Kalman filter in a difficult numerical situation. For example, when $\widehat{d}_1(t) = \widehat{d}_2(t) = 0$ (Case 1), the measurement update step should set the upper 2×2 block of the Kalman gain matrix $\mathbf{L}_{r+t-t_0}(t)$ to the identity matrix. This in turn should result in setting the first two elements of the vector $\widehat{\mathbf{x}}_{r+t-t_0}(t|t)$ to $y_1(t)$ and $y_2(t)$, respectively, and zeroing the 2×2 upper-left corner block of the *a posteriori* matrix $\mathbf{P}_{r+t-t_0}(t|t)$. Since the "theoretical" *a posteriori* covariance matrix (i.e., the one evaluated with infinite precision) is in this case singular, its computed version – due to numerical errors – may easily lose nonnegative definiteness, causing erratic behavior of Kalman filter afterwards. Similar problems arise when $\widehat{d}_1(t) = 0 \wedge \widehat{d}_2(t) = 1$ (Case 2) or $\widehat{d}_1(t) = 1 \wedge \widehat{d}_2(t) = 0$ (Case 3). The ill-conditioning problem pointed out above can be solved using square-root filtering [18] but, in the specific application considered in this paper, a much simpler round-off technique proved to guarantee numerical robustness – after each cycle of computations (performed in MATLAB) the results were rounded to the 12-th decimal place. The direct consequence of rounding-off is that all "almost zero" and "almost one" elements of the computed matrices/vectors are replaced with zeros and ones, respectively. This allows one to avoid numerical problems while preserving the variable-order structure of the Kalman filtering algorithm.

V. Experimental Results

To evaluate the proposed approach, we used 20 clean audio recordings, 10 containing classical music and 10 containing jazz music (5 vocal pieces and 5 purely instrumental ones), sampled at the rate of 48 kHz with 16-bit resolution, and contaminated with real click waveforms extracted from silent parts of old gramophone recordings. Our repository of clicks was made up of 1003 pairs of click waveforms (found in the left and right channel, respectively). Clean recordings contained from 25 to 33 seconds of audio material. Prior to adding noise pulses, all audio signals were scaled so as to make their energy in the corrupted part identical. The 20 second long click template, which was added to clean audio (the same for all recordings), consisted of 3200 pairs of equally spaced noise pulses picked at random from the click database: 807 pulses corrupting the left channel only, 800 pulses corrupting the right channel only, and 1593 pulses corrupting both channels. The total number of corrupted samples was equal to 44013, which constitutes 2,3% of all samples in the analyzed fragment.

Performance evaluation was made for 4 unidirectional/bidirectional approaches: the scalar double-threshold based approach (A/A*), the scalar open-loop prediction based approach (B/B*), the scalar decision-feedback prediction based approach (C/C*), and the vector decision-feedback prediction based approach proposed in this paper (D/D*).

All compared detection/reconstruction algorithms incorporated AR/VAR models of order $r = 10$. For the residual error based double-threshold approach A/A*, the default values of internal parameters recommended in [12] were adopted. For the prediction error based approaches B/B*, C/C* and

D/D*, signal identification was carried out using the EWLS algorithm equipped with forgetting factors: $\lambda = 0.999$ – in the case of vector processing, and [in agreement with (18)], $\lambda_1 = \lambda_2 = 0.998$ – in the case of scalar processing. The detection multiplier was set to $\mu = 4.5$. The bandwidth coefficient of the high-pass filter was set to $c = 0.995$, and the forgetting constant used for covariance updating – to $\lambda_0 = 0.993$. The alarm extension parameter was set to $\Delta t = 2$.

Our evaluation of audio reconstruction results was performed using the Perceptual Evaluation of Audio Quality (PEAQ) tool [21], [22]. PEAQ scores take negative values that range from -4 (very annoying distortions) to 0 (imperceptible distortions). The PEAQ standard uses a number of psychoacoustical evaluation techniques which are combined to give a measure of the quality difference between the original audio signal and its processed version. Even though it was introduced as an objective method to measure the quality of perceptual coders, without any reference to audio restoration, we have found it useful for our purposes as it gives scores that are well correlated with the results of time consuming listening tests. We have found out experimentally that, in the case of elimination of impulsive disturbances, the PEAQ threshold above which signal distortions can be regarded as imperceptible is roughly equal to -0.1. Similarly, according to our experience, the differences between two approaches that reach or exceed the level of 0.1 in terms of the associated PEAQ scores, i.e., $|\mathrm{PEAQ}_1 - \mathrm{PEAQ}_2| \geq 0.1$, are usually audible.

Tab. I summarizes performance statistics for the compared approaches. Qualitative comparison of the results given in Tab. I is presented in Tab. II, which shows the number of cases where a given approach obtained a better PEAQ score than its competitor. The number of "strong wins", i.e., the number of cases where the compared PEAQ scores differ by at least 0.1, are shown in curly brackets.

There are several conclusions that can be drawn after examining results presented in Tabs. I and II:

1) In almost all cases the vector decision-feedback prediction based approach yields better results than approaches based on double thresholding and open-loop prediction – see A vs. D, B vs. D, A* vs. D* and B* vs. D* in Tab. II. The performance gains are usually large.
2) In almost all cases the vector decision-feedback prediction based approach yields better results than its scalar counterpart – see C vs. D and C* vs. D* in Tab. II. The performance improvement is noticeable in about 30% of cases (usually those where local correlation between the left and right audio tracks is strong – see Fig. 2).
3) All approaches benefit from bidirectional processing – see A vs. A*, B vs. B*, C vs. C* and D vs. D* in Tab. II. The performance improvements are usually significant.

Random listening tests, performed on real archive audio recordings, support these findings.

VI. Conclusion

The problem of elimination of impulsive disturbances from stereo audio recordings was solved using the vector autoregressive modeling technique. The proposed approach combines an exponentially weighted least squares model identification algorithm with variable-order Kalman filter, used to detect and interpolate irrevocably distorted signal samples. It was shown that restoration results improve when both stereo channels are analyzed and processed jointly.

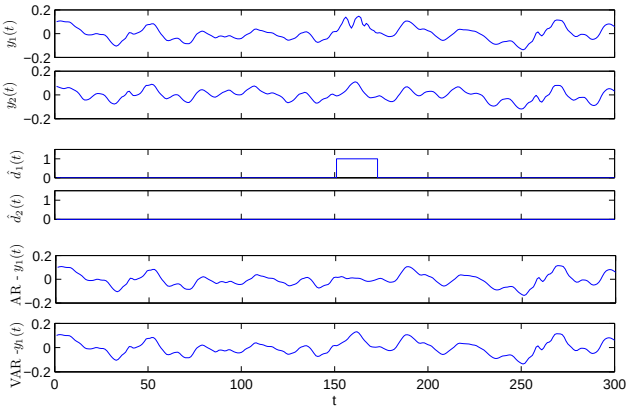

Fig. 4: Comparison of AR and VAR signal reconstructions. Two top plots show a fragment of a stereo archive audio signal with corrupted left track. Two middle plots show decisions of the outlier detector. Two bottom plots show results of signal reconstruction based on the scalar model (upper plot) and vector model (lower plot).

References

[1] S.V. Vaseghi and P.J.W. Rayner, "Detection and suppression of impulsive noise in speech communication systems," *IEE Proceedings*, vol. 137, pp. 38–46, 1990.

[2] S.V. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings," *J. Audio Eng. Soc.*, vol. 40, pp. 791–801, 1992.

[3] M. Niedźwiecki and K. Cisowski, "Adaptive scheme for elimination of broadband noise and impulsive disturbances from audio signals," *Proc. Quatrozieme Colloque GRETSI*, Juan-les-Pins, France, pp. 519–522, 1993.

[4] S.J. Godsill and P.J.W. Rayner, "A Bayesian approach to the restoration of degraded audio signals," *IEEE Trans. Speech, Audio Process.*, vol. 3, pp. 267–278, 1995.

[5] 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, 1995.

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

[7] M. Niedźwiecki, "Identification of time-varying processes in the presence of measurement noise and outliers," *Proc. 11th IFAC Symposium on System Identification*, Fukuoka, Japan, pp. 1765-1770, 1997.

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

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

[10] S. Canazza, G. De Poli, and G.A. Mian, "Restoration of audio documents by means of extended Kalman filter," *IEEE Trans. Audio, Speech Language Process.*, vol. 18, pp. 1107-1115, 2010.

[11] M. Niedźwiecki and M. Ciołek, "Elimination of impulsive disturbances from archive audio signals using bidirectional processing," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, pp. 1046–1059, 2013.

[12] P.A.A. Esquef, L.W.P. Biscainho, P.S.R Diniz, and F.P. Freeland, "A double-threshold-based approach to impulsive noise detection in audio signals," *Proc. European Signal Process. Conf.*, Tampere, Finland, pp. 2041-2044, 2000.

TABLE I: Comparison of the PEAQ scores obtained for 4 unidirectional/bidirectional approaches: the scalar double-threshold based approach (A/A*), the scalar open-loop prediction based approach (B/B*), the scalar decision-feedback prediction based approach (C/C*), and the vector decision-feedback prediction based approach proposed in this paper (D/D*). All results were obtained for 20 artificially corrupted audio files: 10 with classical music and 10 with jazz music. REF denotes the score of the input (corrupted) recording. Interpretation of PEAQ scores: 0 = imperceptible (signal distortions), -1 = perceptible but not annoying, -2 = slightly annoying, -3 = annoying, -4 = very annoying.

classical music

input file		unidirectional processing				bidirectional processing			
No.	REF	A	B	C	D	A*	B*	C*	D*
1	-3.73	-3.01	-3.88	-0.45	-0.42	-0.78	-2.24	-0.23	-0.24
2	-3.78	-0.68	-3.76	-0.25	-0.24	-0.43	-1.00	-0.09	-0.08
3	-3.50	-3.36	-3.90	-0.43	-0.31	-1.29	-2.51	-0.18	-0.18
4	-3.72	-2.31	-3.88	-0.59	-0.40	-1.33	-2.48	-0.25	-0.20
5	-3.75	-1.40	-3.78	-0.43	-0.37	-0.68	-1.23	-0.23	-0.18
6	-3.88	-3.26	-3.89	-0.31	-0.30	-1.42	-3.30	-0.14	-0.12
7	-3.83	-3.36	-3.90	-0.35	-0.33	-1.36	-3.45	-0.18	-0.15
8	-3.91	-3.56	-3.86	-1.01	-0.91	-2.21	-3.42	-0.78	-0.69
9	-3.90	-3.56	-3.90	-0.88	-0.85	-1.87	-3.59	-0.70	-0.67
10	-3.83	-3.33	-3.89	-0.35	-0.50	-1.07	-2.61	-0.19	-0.27

jazz music

input file		unidirectional processing				bidirectional processing			
No.	REF	A	B	C	D	A*	B*	C*	D*
1	-3.81	-2.06	-3.78	-0.81	-0.72	-0.95	-2.20	-0.53	-0.41
2	-3.89	-1.03	-3.67	-1.01	-1.00	-1.18	-0.87	-0.74	-0.55
3	-3.73	-0.48	-3.15	-0.60	-0.55	-0.44	-0.42	-0.33	-0.32
4	-3.52	-2.00	-3.68	-2.27	-2.30	-1.42	-1.82	-1.39	-1.13
5	-3.51	-0.57	-3.07	-0.61	-0.59	-0.52	-0.38	-0.39	-0.36
6	-3.60	-1.37	-3.49	-1.10	-1.08	-0.99	-1.17	-0.77	-0.76
7	-3.51	-1.68	-3.38	-1.16	-1.08	-2.06	-1.09	-1.10	-0.66
8	-3.37	-0.93	-2.87	-0.71	-0.53	-1.22	-0.61	-0.64	-0.46
9	-3.65	-0.74	-3.40	-0.59	-0.44	-0.84	-0.53	-0.44	-0.20
10	-3.81	-1.77	-3.63	-0.75	-0.55	-2.29	-0.99	-0.66	-0.34

TABLE II: Direct comparison of different unidirectional/bidirectional approaches: the scalar double-threshold based approach (A/A*), the scalar open-loop prediction based approach (B/B*), the scalar decision-feedback prediction based approach (C/C*), and the vector decision-feedback prediction based approach (D/D*). The competing approaches are listed in the first column. The remaining columns present the number of instances where a given approach earned a better PEAQ score than its competitor. The number of "strong wins" is shown in curly brackets.

compared approaches	classical music	jazz music	total
A vs. D	0 {0} / 10 {10}	3 {1} / 7 { 6}	3 {1} / 17 {16}
B vs. D	0 {0} / 10 {10}	0 {0} / 10 {10}	0 {0} / 20 {20}
C vs. D	1 {1} / 9 { 3}	1 {0} / 9 { 4}	2 {1} / 18 { 7}
A* vs. D*	0 {0} / 10 {10}	0 {0} / 10 {10}	0 {0} / 20 {20}
B* vs. D*	0 {0} / 10 {10}	0 {0} / 10 { 9}	0 {0} / 20 {19}
C* vs. D*	2 {0} / 8 { 0}	0 {0} / 10 { 6}	2 {0} / 18 { 6}
A vs. A*	0 {0} / 10 {10}	5 {5} / 5 { 3}	5 {5} / 15 {13}
B vs. B*	0 {0} / 10 {10}	0 {0} / 10 {10}	0 {0} / 20 {20}
C vs. C*	0 {0} / 10 {10}	0 {0} / 10 { 7}	0 {0} / 20 {17}
D vs. D*	0 {0} / 10 {10}	0 {0} / 10 { 9}	0 {0} / 20 {19}

[13] M. Niedźwiecki, "Statistical reconstruction of multivariate time series," *IEEE Transactions on Signal Processing*, vol. 41, pp. 451–457, 1993.

[14] R.L. Kashyap and R. Rao, *Dynamic Stochastic Models from Empirical Data*, Academic Press, 1976.

[15] H. Lüthepohl, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, 1991.

[16] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994.

[17] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2001.

[18] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1979.

[19] F. Lewis, *Optimal Estimation*. Wiley, 1986.

[20] T. Söderstrom and P. Stoica, *System Identification*, Prentice-Hall, 1988.

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

[22] 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.

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 Computer Science, Gdańsk University of Technology.

Marcin Ciołek received the M.Sc. degree in automatic control from the Gdańsk University of Technology, Gdańsk, Poland, in 2010, where is currently pursuing the Ph.D. degree. Since 2011, he has been working as an Assistant Professor in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. His professional interests include speech, music and biomedical signal processing.

Krzysztof Cisowski received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the M.A. degree from the Academy of Musical Arts, Gdańsk, Poland, in 1983, 2001 and 1990, respectively. From 1983 to 1985 he played in the Baltic Opera Orchestra. In 1992, he joined the staff of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Technical University of Gdańsk, where he is currently Adjunct Professor. His research interests are in the area of signal processing.