ELIMINATION OF IMPULSIVE DISTURBANCES FROM STEREO AUDIO RECORDINGS

Maciej Niedźwiecki and Marcin Ciołek

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control
Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland
maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl

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 stability-preserving Whittle-Wiggins-Robinson algorithm with exponential data weighting. 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 silent parts of old gramophone recordings.

Index Terms— Elimination of impulsive disturbances

1. 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, foults in the record stamping process etc. Elimination of such disturbances from archive audio documents is an important element of saving our cultural heritage. Although two tracks of a stereophonic audio signal can be split and processed separately (see e.g. [1] – [4] and the references therein), 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.

2. SIGNAL IDENTIFICATION

The measured stereo audio signal will be denoted by $\mathbf{y}(t) = [y_1(t), y_2(t)]^{\mathrm{T}}$, where $t = \dots, -1, 0, 1, \dots$, 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:

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

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

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. 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).

2.1. Stationary Case, Known Covariance Structure

The clean audio signal will be modeled as a vector (two-dimensional) autoregressive (VAR) process of order r

$$\mathbf{s}(t) = \sum_{i=1}^{r} \mathbf{A}_i \mathbf{s}(t-i) + \mathbf{n}(t), \quad \mathrm{cov}[\mathbf{n}(t)] = \boldsymbol{\rho} \tag{2}$$

where

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

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

Denote by $\mathbf{0}$ the 2×1 null vector, and by $\mathbf{O}$ and $\mathbf{I}$ – the 2×2 null and identity matrices, respectively.

In the majority of audio applications, including the adaptive detection/reconstruction problem considered in this paper, the stability of the signal model must be guaranteed to make the model-based analysis (such as the multiple-step-ahead signal prediction and its interpolation) well-posed. The classical stability-preserving identification method, which can be applied in the stationary case, is based on Yule-Walker equations that link the parameters of the model (2) with the covariance structure of the VAR signal

$$[\mathbf{I}, -\mathbf{A}_1, \dots, -\mathbf{A}_r]\, \boldsymbol{\mathcal{R}} = [\boldsymbol{\rho}, \mathbf{O}, \dots, \mathbf{O}] \tag{3}$$

where $\boldsymbol{\mathcal{R}} > 0$ is a block-Toeplitz matrix made up of covariance matrices of the VAR process $\mathbf{R}_k = \mathrm{E}[\mathbf{s}(t)\mathbf{s}^{\mathrm{T}}(t-k)]$, $k = 0, \dots, r$:

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \dots & \mathbf{R}_r \\ \mathbf{R}_1^{\mathrm{T}} & \mathbf{R}_0 & & \vdots \\ \vdots & & & \mathbf{R}_1 \\ \mathbf{R}_r^{\mathrm{T}} & \dots & & \mathbf{R}_0 \end{bmatrix}. \tag{4}$$

The order-recursive solution of (3) can be obtained using the well-known 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 [6]. An important feature of the method described above is that under $\boldsymbol{\mathcal{R}} > 0$ it always yields a stable VAR model.

2.2. Nonstationary Case, Unknown Covariance Structure

When parameters of the VAR model (2) vary slowly with time their estimates $\widehat{\mathbf{A}}_1(t), \ldots, \widehat{\mathbf{A}}_r(t), \widehat{\boldsymbol{\rho}}(t)$, based on the available observation history $\mathcal{Y}(t) = \{\mathbf{y}(1), \ldots, \mathbf{y}(t)\}$, can be obtained by solving the system of Yule-Walker equations analogous to (3)

$$[\mathbf{I}, -\widehat{\mathbf{A}}_1(t), \ldots, -\widehat{\mathbf{A}}_r(t)]\, \widehat{\boldsymbol{\mathcal{R}}}(t) = [\widehat{\boldsymbol{\rho}}(t), \mathbf{O}, \ldots, \mathbf{O}] \quad (5)$$

where $\widehat{\boldsymbol{\mathcal{R}}}(t)$ is the block-Toeplitz matrix obtained by means of replacing in (4) the true signal covariances $\mathbf{R}_k$ with their local sample estimates. In the absence of impulsive noise such local estimates can be obtained using the approach known as exponential weighting (or exponential forgetting). Perhaps the most straightforward application of this technique would be to replace each $\mathbf{R}_k$ with

$$\widehat{\mathbf{R}}_k(t) = \frac{1}{l_0(t)} \sum_{j=k+1}^{t} \lambda_0^{t-j} \mathbf{y}(j)\mathbf{y}^{\mathrm{T}}(j-k) \quad (6)$$

where λ_0, $0 < \lambda_0 < 1$, denotes the so-called forgetting constant and $l_0(t) = \sum_{j=1}^{t} \lambda_0^{t-j}$ denotes the effective width of the exponential window. The value of λ_0 should be chosen so as to trade off the bias and variance components of the mean-squared parameter tracking error [6]. However, since the matrix $\widehat{\boldsymbol{\mathcal{R}}}(t)$ made up of the blocks computed according to (6) is not guaranteed to be positive definite, such a solution is not feasible.

A simple way out of difficulty, proposed hereby, is to compute covariance estimates according to

$$\widehat{\mathbf{R}}_k(t) = \frac{\mathbf{Y}_k(t)}{l(t)}\ , \quad l(t) = \sum_{j=1}^{t} \lambda^{2(t-j)} = \frac{1-\lambda^{2t}}{1-\lambda^2}$$

$$\mathbf{Y}_k(t) = \sum_{j=k+1}^{t} \mathbf{y}_\lambda(j,t)\mathbf{y}_\lambda^{\mathrm{T}}(j-k,t)\ ,\ k = 0, \ldots, r \quad (7)$$

where λ, $0 < \lambda < 1$, denotes the forgetting constant and $\{\mathbf{y}_\lambda(1,t), \ldots, \mathbf{y}_\lambda(t,t)\}$ is the exponentially weighted data sequence

$$\mathbf{y}_\lambda(j,t) = \lambda^{t-j}\mathbf{y}(t), \ \ j \le t.$$

The quantities $l(t)$ and $\mathbf{Y}_k(t)$ can be computed recursively according to

$$l(t) = \lambda^2 l(t-1) + 1, \ \ t \ge 1$$
$$\mathbf{Y}_k(t) = \lambda^2 \mathbf{Y}_k(t-1) + \lambda^k \mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-k), \ \ t > k \quad (8)$$

with initial conditions set to $l(0) = 0$ and $\mathbf{Y}_k(k) = \mathbf{O}$.

When the forgetting constant λ is close to 1, the estimates based on exponential data weighting (7) yield similar results as exponentially weighted estimates (6) obtained for $\lambda_0 = \lambda^2$. The important difference is, however, that unlike the weighted estimation scheme, the data weighting technique guarantees positive definiteness of the matrix $\widehat{\boldsymbol{\mathcal{R}}}(t)$. In order to see this note that – when made up of the blocks (7) – the matrix $\widehat{\boldsymbol{\mathcal{R}}}(t)$can be written down in the form

$$\widehat{\boldsymbol{\mathcal{R}}}(t) = \frac{1}{l(t)} \sum_{j=1}^{t+r} \mathbf{z}(j)\mathbf{z}^{\mathrm{T}}(j)$$

where $\mathbf{z}(j) = [\widetilde{\mathbf{y}}_\lambda^{\mathrm{T}}(j,t), \ldots, \widetilde{\mathbf{y}}_\lambda^{\mathrm{T}}(j-r,t)]^{\mathrm{T}}$ and $\{\widetilde{\mathbf{y}}_\lambda(j,t), j = 1-r, \ldots, t+r\}$ denotes the sequence of weighted samples preceded and succeeded by r zero samples: $\widetilde{\mathbf{y}}_\lambda(j,t) = \mathbf{y}_\lambda(j,t)$ for $j \in [1,t]$ and $\widetilde{\mathbf{y}}_\lambda(j,t) = \mathbf{0}$ for $j \in [1-r, 0] \cup [t+1, t+r]$. The (almost sure) positive definiteness of $\widehat{\boldsymbol{\mathcal{R}}}(t)$, which guarantees model stability, follows from two facts: 1) the quadratic form associated with $\widehat{\boldsymbol{\mathcal{R}}}(t)$ obeys $\mathbf{w}^{\mathrm{T}}\widehat{\boldsymbol{\mathcal{R}}}(t)\mathbf{w} = [1/l(t)] \sum_{j=1}^{t+r} [\mathbf{w}^{\mathrm{T}}\mathbf{z}(j)]^2 \ge 0$, and 2) the VAR process is not (linearly) deterministic.

A single step of the proposed recursive stability-preserving identification algorithm can be summarized as follows:
1) Update covariance estimates $\widehat{\mathbf{R}}_k(t), k = 0, \ldots, r$, using (7)-(8); 2) Use the WWR algorithm to compute parameter estimates $\widehat{\mathbf{A}}_1(t), \ldots, \widehat{\mathbf{A}}_r(t)$ and $\widehat{\boldsymbol{\rho}}(t)$.

The WWR algorithm with exponential data weighting is a new type of lattice filter which does not seem to have been exploited before. For an overview of lattice estimation techniques see e.g. [7]. Note that, unlike the univariate case, in the multivariate case only a few existing lattice algorithms guarantee model stability.

2.3. Estimation in the Presence of Outliers

The estimates (7) 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 signal covariances (7) 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).

3. DETECTION OF NOISE PULSES AND SIGNAL INTERPOLATION

3.1. 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

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

The solution, based on Kalman smoothing [8], 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/smoother 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. Denote by $\mathbf{0}_q$ the $2q \times 1$ null vector, and by $\mathbf{O}_q$ and $\mathbf{I}_q$ – the $2q \times 2q$ null and identity matrices, respectively. According to our earlier notation: $\mathbf{0}_1 = \mathbf{0}$, $\mathbf{O}_1 = \mathbf{O}$ and $\mathbf{I}_1 = \mathbf{I}$. 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].

$$\begin{aligned} \mathbf{x}_q(t+1) &= \boldsymbol{\mathcal{A}}_q \mathbf{x}_q(t) + \boldsymbol{\mathcal{C}}_q \mathbf{n}(t+1) \\ \mathbf{y}(t) &= \boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \mathbf{x}_q(t) + \boldsymbol{\delta}(t) \end{aligned} \tag{9}$$

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.

Based on (9) and on the available prior knowledge, the Kalman filter/predictor recursions can be written down as follows

$$\begin{aligned} \widehat{\mathbf{x}}_q(t|t-1) &= \boldsymbol{\mathcal{A}}_q \widehat{\mathbf{x}}_q(t-1|t-1) \\ \mathbf{P}_q(t|t-1) &= \boldsymbol{\mathcal{A}}_q \mathbf{P}_q(t-1|t-1) \boldsymbol{\mathcal{A}}_q^{\mathrm{T}} + \boldsymbol{\mathcal{C}}_q \boldsymbol{\rho} \boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \\ \mathbf{e}(t) &= \mathbf{y}(t) - \boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \widehat{\mathbf{x}}_q(t|t-1) \\ \mathbf{S}(t) &= \boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \mathbf{P}_q(t|t-1) \boldsymbol{\mathcal{C}}_q + \boldsymbol{\Delta}(t) \\ \mathbf{L}_q(t) &= \mathbf{P}_q(t|t-1) \boldsymbol{\mathcal{C}}_q \mathbf{S}^{-1}(t) \\ \widehat{\mathbf{x}}_q(t|t) &= \widehat{\mathbf{x}}_q(t|t-1) + \mathbf{L}_q(t) \mathbf{e}(t) \\ \mathbf{P}_q(t|t) &= \mathbf{P}_q(t|t-1) - \mathbf{L}_q(t) \mathbf{S}(t) \mathbf{L}_q^{\mathrm{T}}(t). \end{aligned} \tag{10}$$

Since we have assumed that $\boldsymbol{\delta}(t) = \mathbf{0}$ for $t \le t_0$, 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 $t_0 + m + r$.

3.2. Signal Prediction and Detection of Noise Pulses

Similar to [2], our pulse detection scheme will be based on monitoring signal prediction errors. Denote by $\boldsymbol{\varepsilon}(t|t-1) = \mathbf{y}(t) - \sum_{i=1}^{r} \widehat{\mathbf{A}}_i(t)\mathbf{y}(t-i) = [\varepsilon_1(t|t-1), \varepsilon_2(t|t-1)]^{\mathrm{T}}$ the one-step-ahead signal prediction error. The stereo detection alarm is started at the instant $t_0 + 1$ if in at least one of the channels the magnitude of the prediction error $\varepsilon_j(t|t-1)$ exceeds μ times its standard deviation

$$|\varepsilon_j(t_0 + 1|t_0)| > \mu\sigma_j(t_0 + 1|t_0), \;\; j = 1 \text{ and/or } 2 \tag{11}$$

where $\sigma_j^2(t_0 + 1|t_0) = [\widehat{\boldsymbol{\rho}}(t_0)]_{jj}$, 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 $\mathbf{e}(t) = [e_1(t), e_2(t)]^{\mathrm{T}}$ yielded by the Kalman filtering algorithm initialized at the instant t_0. Detection alarm is stopped at the instant $t = t_0 + m$ if r consecutive prediction errors take for both channels sufficiently small values: $\widehat{d}_1(t) = \widehat{d}_2(t) = 0, t = t_0 + m + 1, \ldots, t_0 + m + r$ where

$$\widehat{d}_j(t) = \begin{cases} 0 & \text{if} \;\; |e_j(t)| \le \mu\sigma_j(t) \\ 1 & \text{if} \;\; |e_j(t)| > \mu\sigma_j(t) \end{cases}, \;\; j = 1, 2 \tag{12}$$

$\sigma_j^2(t) = [\boldsymbol{\Sigma}(t)]_{jj}$, $j = 1, 2$, and $\boldsymbol{\Sigma}(t) = \mathrm{cov}[\mathbf{e}(t)] = \boldsymbol{\mathcal{C}}_q^{\mathrm{T}} \mathbf{P}_q(t|t-1) \boldsymbol{\mathcal{C}}_q$ denotes the corresponding covariance matrix. Alternatively, detection alarm is terminated if $t - t_0$ reaches its maximum allowable value equal to $m_{\max}$.

Signal reconstruction is governed by the following noise covariance scheduling based on (12):

[1] When μ is set to 3, condition (11) is usually referred to as "3-sigma" outlier detection rule.

$$\boldsymbol{\Delta}(t) = \begin{cases} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 0 \\ \begin{bmatrix} 0 & 0 \\ 0 & \gamma \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 0 \wedge \widehat{d}_2(t) = 1 \\ \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = 1 \wedge \widehat{d}_2(t) = 0 \\ \begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix} & \text{if} \quad \widehat{d}_1(t) = \widehat{d}_2(t) = 1 \end{cases} \tag{13}$$

$$\gamma \to \infty \, .$$

Note that when $[\boldsymbol{\Delta}(t)]_{jj} \to \infty$, the sample $y_j(t)$ is regarded as corrupted with infinite-variance noise and – as such – rejected and scheduled for interpolation; when $[\boldsymbol{\Delta}(t)]_{jj} = 0$ the sample is preserved without changes.

3.3. Adaptive Detection and Interpolation

According to [5], 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. Suppose that condition (11) is met at instant $t_0 + 1$. Once this happens, the parameter tracking procedure is temporarily stopped, and the Kalman filter based detection procedure is started. We will introduce two important modifications. First, the true model parameters will be replaced with their most recent estimates $\widehat{\mathbf{A}}_1(t_0), \ldots, \widehat{\mathbf{A}}_r(t_0)$ and $\widehat{\boldsymbol{\rho}}(t_0)$. Second, the fixed-order Kalman filter, presented in Section 3.1, 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 (10). Taking advantage of this structure, one can show that the order of the Kalman filter (10) can be – without affecting estimation results – gradually increased, starting from $r+1$ at the instant $t_0 + 1$, 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.

3.4. Algorithm

Denote by $\widehat{\boldsymbol{\Theta}}_r(t) = [\widehat{\boldsymbol{\theta}}_1(t)|\widehat{\boldsymbol{\theta}}_2(t)]$, where $\widehat{\boldsymbol{\theta}}_j(t) = [\widehat{\boldsymbol{\alpha}}_{j1}^{\mathrm{T}}(t), \ldots, \widehat{\boldsymbol{\alpha}}_{jr}^{\mathrm{T}}(t)]^{\mathrm{T}}$ is the vector of coefficients characterizing the j-th channel, 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 (10) with (12)-(13) 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

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

3.5. Closing Detection Gaps

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

4. EXPERIMENTAL RESULTS

To evaluate the proposed approach we used 5 clean audio recordings, sampled at the rate of 48 kHz, 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 audio recordings contained from 25 to 33 seconds of classical music. The excerpts were chosen so as to cover different temporal and spectral features of audio signals. 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.

The results of comparison of the classical AR-model based approach with that based on the VAR model are shown in Table 1. Our evaluation was performed using the Perceptual Evaluation of Audio Quality (PEAQ) tool [9]. Note that in all cases considered, the PEAQ scores obtained using the VAR model are better (i.e., closer to 0) than those obtained using the AR model. Listening tests, performed on real archive audio files, support these findings.

Table 1: Comparison of the PEAQ scores obtained for the results of univariate (AR) and multivariate (VAR) processing of 5 artificially corrupted audio recordings: $\mathrm{PEAQ}_{\mathrm{in}}$ denotes the score of the input (corrupted) recording, $\mathrm{PEAQ}_{\mathrm{out}}$ denotes the score of the processed recording, and $\mathrm{PEAQ}_{\mathrm{ref}}$ denotes the score obtained when interpolation of the corrupted samples is based on exact knowledge of pulse locations. Interpretation of PEAQ scores: 0 = imperceptible (signal distortions), -1 = perceptible but not annoying, -2 = slightly annoying, -3 = annoying, -4 = very annoying.

Rec.	$\mathrm{PEAQ}_{\mathrm{in}}$	Model	$\mathrm{PEAQ}_{\mathrm{ref}}$	$\mathrm{PEAQ}_{\mathrm{out}}$
1	-3,666	AR	-0,006	-0,531
		VAR	-0,001	-0,493
2	-3,733	AR	-0,023	-1,159
		VAR	-0,030	-1,109
3	-3,458	AR	-0,350	-0,907
		VAR	-0,248	-0,772
4	-3,640	AR	-0,234	-1,255
		VAR	-0,205	-1,243
5	-3,645	AR	-0,026	-0,646
		VAR	-0,053	-0,629

5. CONCLUSION

The problem of elimination of impulsive disturbances from stereo audio recordings was solved using the vector autoregressive modeling technique. The proposed approach combines a new model-stability-preserving identification algorithm – the Whittle-Wiggins-Robins algorithm with exponential data weighting – with the 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.

6. REFERENCES

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

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

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

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

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

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

[7] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–867, 1982.

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

[9] ITU-R Recommendation BS.1387, *Method for Objective Measurement of Perceived Audio Quality*, 1998.