Sparse vector autoregressive modeling of audio signals and its application to the elimination of impulsive disturbances

Maciej Niedźwiecki * Marcin Ciołek *

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

Abstract: Archive audio files are often corrupted by impulsive disturbances, such as clicks, pops and record scratches. This paper presents a new method for elimination of impulsive disturbances from stereo audio signals. The proposed approach is based on a sparse vector autoregressive signal model, made up of two components: one taking care of short-term signal correlations, and the other one taking care of long-term correlations. The method is evaluated on a set of clean audio signals contaminated with real click waveforms extracted from silent parts of old gramophone recordings.

Keywords: elimination of impulsive disturbances, sparse vector autoregressive modeling

1. INTRODUCTION

Archive audio files, such as old gramophone or magnetic tape recordings, are often corrupted by impulsive disturbances such as clicks, pops and record scratches. Noise pulses, of different geometry and length, are usually caused by local degradation of the recording medium due to its ageing and/or mishandling (e.g. local groove damages on vinyl LP’s), or due to equipment errors during recording and transmission of the audio material (e.g. clicks due to electrical discharges, occasionally found in old tape recordings) – see e.g. Godsill and Rayner (1998), Vaseghi (2008).

Most of the existing schemes for elimination of impulsive disturbances, developed for monophonic audio signals, are based on autoregressive (AR) signal modeling and model-based adaptive prediction: an on-line identification of the AR model is carried out and its results are used to predict new samples from the old ones. If the magnitude of the prediction error is too large, e.g. if it exceeds three standard deviations of its nominal value, the sample is classified as an outlier and scheduled for reconstruction. At the second stage of processing all irreversibly distorted samples, localized by the outlier detector, are interpolated from the neighboring (undistorted) samples, using the AR-model based interpolation formula – see e.g. Vaseghi and Rayner (1990), Vaseghi and Frayling-Cork (1992), Godsill (1995a), Godsill and Rayner (1995b), Niedźwiecki and Cisowski (1993b), Niedźwiecki and Cisowski (1996), Niedźwiecki (1997).

Restoration results can be further improved if sparse autoregressive (SAR) modeling technique is used instead of the AR one – see Niedźwiecki and Ciołek (2013). Since SAR models capture both short-term and long-term correlations in the analyzed signal, they have considerably better descriptive capabilities. Sparse models are particularly useful for processing signals that are formed by exciting the acoustical waveguide with a periodic train of air pulses (voiced speech sounds, audio signals with strong vocal components, purely instrumental music with a contribution from some wind instruments) – see Wolfe, Garnier and Smith (2009).

Unlike outlier detectors based on conventional AR models, detectors that incorporate SAR models usually do not confuse excitation (pitch) pulses with noise pulses. Additionally, the quality of signal reconstruction based on the SAR model is usually better than that based on the AR model, especially in the presence of strong pitch activity.

Restoration of stereo audio signals is usually performed by splitting the tracks and processing them separately. In Niedźwiecki and Ciołek (2014b) we have shown that restoration results can be improved if both audio tracks are modeled jointly using the vector autoregressive (VAR) modeling approach. The benefits of VAR modeling can be observed both at the outlier detection stage (more accurate localization of noise pulses) and at the signal interpolation stage (undistorted material in one track can be used to “repair” the corrupted fragment in the other track).

In this paper we will focus on extension of the sparse autoregressive technique to stereo audio signals. The corresponding models will be further referred to as sparse vector autoregressive (SVAR).

2. SPARSE VECTOR AUTOREGRESSIVE MODELS OF AUDIO SIGNAL

2.1 Generic SVAR model

The stereo audio signal will be denoted by \( y(t) = [y_1(t), y_2(t)]^T \), where \( t = \ldots, -1, 0, 1 \ldots \) denotes normalized (dimensionless) discrete time, while \( y_1(t) \) and \( y_2(t) \) denote the left and right audio track, respectively. The SVAR model of \( y(t) \) can be written down in the form

\[ y(t) = \sum_{k=1}^{K} \beta_k y(t-k) + \epsilon(t) \]
\[ y(t) = \sum_{i=1}^{r} A_i y(t-i) + \sum_{i=r+1}^{\tau+p} A_i y(t-i) + n(t) \tag{1} \]

where

\[
A_i = \begin{bmatrix} a_{11,i} \\ a_{21,i} \\ a_{12,i} \\ a_{22,i} \end{bmatrix}, \quad i = 1, \ldots, r
\]

are the 2 x 2 matrices of AR coefficients and \( \{n(t)\} \), \( n(t) = [n_1(t), n_2(t)]^T \), denotes zero-mean white noise with covariance matrix \( \rho \).

The quantities \( \tau \gg r \) and \( p \) are chosen so that \( \tau + 1 \leq T \leq \tau + p \), where \( T \) is the fundamental period of \( y(t) \) (if present). In the simplest case, for \( p = 1 \), the second sum on the right hand side of (1) reduces to a single component of the form \( A_T y(t-T) \).

Similarly as their SAR counterparts, SVAR models capture both short-term [the first sum on the right hand side of (1)] and long-term [the second sum on the right hand side of (1)] correlation structure of the analyzed time series. Even though formally (1) must be regarded as a VAR model of order \( \tau + p \), due to its special structure it parameterizes \( y(t) \) in terms of only \( \tau + p \ll \tau + p \) matrices of autoregressive coefficients, i.e., it is a sparse signal description.

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 (1) is asymptotically stable iff all zeros, \( z_i, i = 1, \ldots, 2(\tau + p), \) of the characteristic polynomial \( D(q^{-1}) = \det[A(q^{-1})] \), lie inside the unit circle in the complex plane: \( |z_i| < 1, i = 1, \ldots, 2(\tau + p) \).

When true model parameters \( A_i \) are replaced with their estimates \( \hat{A}_i(t) \), model stability monitoring and enforcement becomes a difficult problem. Since, under 44.1 kHz sampling, typical values of fundamental period are well in excess of 100, stability checks based on analysis of the characteristic polynomial \( D(q^{-1}) \) are hardly practical. The only feasible solution would be to apply such identification routine which can guarantee model stability for all possible data sequences. Unfortunately, even though such stability-preserving identification algorithms are available for the classical VAR model, the Whittle-Wiggins-Robinson scheme (generalization of the Levinson-Durbin algorithm) being one of the examples [Söderström and Stoica (1988)], no such algorithm seems to exist for the generic SVAR model. For this reason, in our further considerations the generic model (1) will be replaced with the factorized one, for which stability monitoring/enforcement can be easily performed.

### 2.2 Factorized SVAR model

Factorized SVAR model is defined as a cascade of two simpler models: one taking care of long-term signal correlations, and the other one taking care of short-term correlations. In the simplest case, which will be considered here, such a factorized description can be written down in the form

\[ y(t) = \sum_{i=1}^{r} B_i y(t-i) + x(t) \tag{2} \]

\[ x(t) = C x(t-T) + n(t) \tag{3} \]

which is asymptotically equivalent to (1) under the substitution

\[ A(q^{-1}) = B(q^{-1})C(q^{-1}) \tag{4} \]

where

\[ B(q^{-1}) = 1 - \sum_{i=1}^{\tau} B_i q^{-i}, \quad C(q^{-1}) = I - C q^{-T}. \]

According to (4), it holds that

\[ A_i = B_i, \quad i = 1, \ldots, r, \quad A_i = O, \quad r < i < T \]

\[ A_T = C, \quad A_{T+i} = -CB_i, \quad i = 1, \ldots, r \tag{5} \]

where \( O \) denotes the 2 x 2 null matrix.

The factorized model (2)-(3) is stable provided that its both components are stable. Unlike the generic SVAR model, the stability of the short-term model governed by (2) can be easily ensured by the proper choice of the identification algorithm, and stability of the long-term model governed by (3) is easy to check and, if necessary, enforced. To examine stability of the model (3), one can use the diagonalization technique. Denote by \( Q \) a nonsingular matrix that converts \( C \) into a diagonal form

\[ Q C Q^{-1} = \Lambda \]

where \( \Lambda = \text{diag}\{\lambda_1, \lambda_2\} \) is made up of the eigenvalues of \( C \). Let \( \tilde{x}(t) = Q x(t) = [\tilde{x}_1(t), \tilde{x}_2(t)]^T \) and \( \tilde{n}(t) = Q n(t) = [\tilde{n}_1(t), \tilde{n}_2(t)]^T \). Note that the model equation (3) can be rewritten in an equivalent, decoupled form

\[ \tilde{x}_i(t) = \lambda_i \tilde{x}_i(t-T) + \tilde{n}_i(t), \quad i = 1, 2 \]

and therefore the model stability condition is a simple extension of that developed in the univariate case [see Ramachandran and Kabal (1987)]:

\[ |\lambda_i| < 1, \quad i = 1, 2. \]

### 3. IDENTIFICATION OF THE SVAR MODEL

To guarantee stability of the factorized model (2)-(3), its identification will be carried out in two steps. First, assuming that (2) is a VAR process of order \( r \), the estimates of the short-term model parameters \( B_1, \ldots, B_r \) will be obtained and used to generate the residual signal \( x(t) \). In the second step, parameters of the long-term model (3) will be estimated.

#### 3.1 Identification of the short-term model

For the purpose of identification of the short-term model (2) we will assume that the input process \( x(t) \), \( \text{cov}[x(t)] = \Sigma \), is white noise. Under this assumption (2) becomes a classical VAR model. Note that, according to (3), it holds that

\[ x(t) = n(t) + \sum_{i=1}^{\infty} C' n(t-iT). \]

Hence, for a white noise excitation \( n(t) \), one obtains the following steady state values of the short-term autocorrelation matrices of \( x(t) \)

\[ R_k = E[x(t)x^T(t-k)] = O, \quad k = 1, \ldots, T-1. \tag{6} \]

Hence, when \( r \ll T \) and identification is carried out for a short-term model, the whiteness assumption imposed on \( x(t) \) seems to be justifiable.
Our identification procedure will be based on solving the set of Yule-Walker equations that link parameters of the VAR model (2) with the covariance structure of the VAR signal. For a stationary VAR process it holds that

\[ \mathbf{L} - \mathbf{B}_1 \ldots \mathbf{B}_L \mathbf{R}^y = [\Sigma, \mathbf{O}, \ldots, \mathbf{O}] \]

where \( \mathbf{R}^y > 0 \) is a block-Toeplitz matrix

\[ \mathbf{R}^y = \begin{bmatrix} \mathbf{R}^y_0 & \mathbf{R}^y_1 & \ldots & \mathbf{R}^y_L \end{bmatrix} \]

made up of covariance matrices of the VAR process \( \mathbf{R}^y = \text{E}[y(t)y^T(t-k)] \), \( k = 0, \ldots, T \). Since in the case considered parameters of the VAR model (2) vary slowly with time, their estimates \( \hat{\mathbf{B}}_i(t), \hat{\mathbf{C}}_i(t), \hat{\Sigma}(t) \) will be obtained by means of solving the system of Yule-Walker equations analogous to (7)

\[ \hat{\mathbf{T}}^*_0(t) = \arg \min_{T \in \Delta_0} \mathbf{J}^*_t(T, M_0, m_0) \] (18)

\[ \hat{\mathbf{T}}^*_t(t) = \arg \min_{T \in \Delta^*_t} \mathbf{J}^*_t(T, M, m) \] (19)

In order to estimate \( T \) we will look for a sequence of residual noise samples, drawn from the signal past, that is the best match for the sequence of \( L \) most recently observed samples \( F_L(t) = \{x(t), \ldots, x(t-L+1)\} \), further called the reference frame. Let

\[ J_t(T, L) = \sum_{i=0}^{L-1} \| x(t-i) - x(t-T-i) \|^2 \]

Additionally, denote by \( T_{\text{min}}/T_{\text{max}} \) the smallest/largest value of the fundamental period that will be considered under a given sampling frequency. The proposed estimation procedure is two-step. First, a crude estimate of \( T \), denoted by \( \hat{T}_0(t) \), is found using global search for a long reference frame of width \( M_0 \)

\[ \hat{T}_0(t) = \arg \min_{T \in \Delta_0} J_t(T, M_0) \] (14)

where \( \Delta_0 = [T_{\text{min}}, T_{\text{max}}] \). Then a precise estimate, denoted by \( \hat{T}(t) \), is found by means of local search around \( \hat{T}_0(t) \), performed for a short reference frame of width \( M < M_0 \)

\[ \hat{T}(t) = \arg \min_{T \in \Delta(t)} J_t(T, M) \] (15)

where \( \Delta(t) = [\max\{T_{\text{min}}, \hat{T}_0(t) - k_0\}, \min\{T_{\text{max}}, \hat{T}_0(t) + k_0\}] \), and \( k_0 \) is a small integer number.

Combining (14) and (15), one arrives at the following estimate of the long-term coefficient matrix

\[ \hat{\mathbf{C}}(t) = \hat{\mathbf{R}}^x_{\hat{T}(t)}(t) \left[ \hat{\mathbf{R}}^x_0(t) \right]^{-1} \]

The resulting long-term model can be written down in the form

\[ x(t) = \hat{\mathbf{C}}(t)x[t - \hat{T}(t)] + \mathbf{n}(t) \] (17)

**Remark:** To guarantee stability of the long-term model (17), one should check, at each time instant \( t \), the eigenvalues of the matrix \( \hat{\mathbf{C}}(t) \). If the condition \( |\lambda_{\text{max}}(t)| = \max\{|\lambda_1(\hat{\mathbf{C}}(t))|, |\lambda_2(\hat{\mathbf{C}}(t))|\} < 1 \) is met, no intervention is needed. Otherwise, model stability can be easily enforced by replacing the matrix \( \hat{\mathbf{C}}(t) \) with its scaled version \( \lambda_0/|\lambda_{\text{max}}(t)| \hat{\mathbf{C}}(t) \), where \( \lambda_0 = 1 - \epsilon \), and \( \epsilon > 0 \) denotes a small stability margin (e.g. \( \epsilon = 0.01 \)).

**B) Noncausal estimation of \( T \)**

As already mentioned, predictive capabilities of the SVAR model strongly depend on precision of determining the fundamental period \( T \). Whenever fundamental period undergoes substantial changes, causal estimates \( \hat{T}(t) \), entirely based on the signal past, may fail to follow \( T(t) \) with sufficient accuracy. Much better estimates are usually obtained when, in addition to the most recently observed samples, the reference frame contains a certain number of “future” (i.e., not yet processed) samples.

The corresponding noncausal estimate \( \hat{T}^*(t) \) can be obtained using the two-step procedure analogous to (14)-(15)

\[ \hat{T}_0^*(t) = \arg \min_{T \in \Delta_0} J^*_t(T, M_0, m_0) \] (18)

\[ \hat{T}^*(t) = \arg \min_{T \in \Delta^*_t} J^*_t(T, M, m) \] (19)
where
\[ J^*_t(T, L, l) = \sum_{i=0}^{L-1} \| x(t+l-i|t) - x(t+l-T-i|t) \|^2 \]
\[ \Delta^*(t) = [\max\{T_{\min}, \hat{T}^*_0(t) - k_0\}, \min\{T_{\max}, \hat{T}^*_0(t) + k_0\}] \]

An alternative to the Kalman filter is the outlier detector based on step-ahead predictions.

\[ \hat{J}(t) = \frac{1}{T} \sum_{i=0}^{T-1} \| x(t+i|t) - x(t+i+1|t) \|^2 \]

where \( \hat{J}(t) \) is the sum of squared differences between the predicted and actual values of the signal at time \( t \). The detector compares \( \hat{J}(t) \) with a threshold \( \sigma^2 \) to decide whether an alarm is raised. If \( \hat{J}(t) > \sigma^2 \), an alarm is triggered.

\[ \sigma^2 = \frac{1}{T} \sum_{i=0}^{T-1} \| x(t+i|t) - x(t+i+1|t) \|^2 \]

### 4. ELIMINATION OF IMPULSIVE DISTURBANCES

Suppose that the measured stereo audio signal has the form
\[ y(t) = s(t) + \delta(t) \]
where \( s(t) = [s_1(t), s_2(t)]^T \) denotes clean audio signal and \( \delta(t) = [\delta_1(t), \delta_2(t)]^T \) is a signal made up of noise pulses sparsely distributed in time. Denote by \( d(t) = [d_1(t), d_2(t)]^T \) the pulse location function

\[ d_j(t) = \begin{cases} 0 & \text{if no noise pulse detected} \\ 1 & \text{if noise pulse detected} \end{cases}, \quad j = 1, 2. \]

Detection of noise pulses, which are often referred to as outliers, can be based on monitoring signal prediction errors. Denote by \( \hat{y}(t+1|t) \) the one-step-ahead signal prediction, and by \( \epsilon(t+1|t) = y(t+1|t) - \hat{y}(t+1|t) = [\epsilon_1(t+1|t), \epsilon_2(t+1|t)]^T \) the corresponding prediction error. Finally, denote by \( \hat{d}(t) = [\hat{d}_1(t), \hat{d}_2(t)]^T \) the output of the outlier detector

\[ \hat{d}_j(t) = \begin{cases} 0 & \text{if no noise pulse not detected} \\ 1 & \text{if noise pulse detected} \end{cases}, \quad j = 1, 2. \]

Following Niedźwiecki and Ciołek (2014b), detection alarm can be triggered in the channel \( j \) at the instant \( t_0 \)
\[ \hat{d}_j(t_0+1) = 1 \] if the magnitude of the corresponding one-step-ahead signal prediction error \( |\epsilon_j(t_0+1)| \) exceeds \( \mu \) times, \( \mu \in [3, 5] \), its estimated standard deviation

\[ |\epsilon_j(t_0+1)| > \mu \hat{\sigma}_j(t_0) \]

where \( \hat{\sigma}_j(t_0) = \hat{\Sigma}(t_0)_{jj} \). After applying the outlier detection rule (23) to prediction errors yielded by each of the competing models, one obtains three provisional detection results \( \hat{d}_{\text{VAR}}(t_0+1), \hat{d}_{\text{SVAR}}(t_0+1), \) and \( \hat{d}_{\text{SVAR}}(t_0+1) \), respectively. Detection alarm is raised if all three prediction errors can be regarded as excessive, namely if

\[ \hat{d}_{\text{VAR}}(t_0+1) \neq 0, \quad \hat{d}_{\text{SVAR}}(t_0+1) \neq 0 \]

and \( \hat{d}_{\text{SVAR}^*}(t_0+1) \neq 0 \). Once the detection alarm is triggered at the instant \( t_0 \) (after a sequence of at least \( r \) no-alarm decisions), the prediction error test similar to (22) can be carried out for multi-step-ahead predictions. As shown in Niedźwiecki and Ciołek (2014b), this task can be accomplished using the appropriately designed variable-order Kalman filter, initialized at the instant \( t_0 \). Detection alarm is terminated at the instant \( t_0 + m \) if \( r \) consecutive prediction errors take for both channels sufficiently small values:

\[ \hat{d}(t_0+m+i) = 0 \]

for \( i = 1, \ldots, r \). The Kalman filter algorithm is based on the model (VAR, SVAR or SVAR*) that minimizes the sum of squared one-step-ahead prediction errors observed in the recent past:

\[ \sum_{i=1}^r \| \epsilon(t_0 + 1 - i) \|^2. \]

If one of the sparse models wins the competition, Kalman filter is based on its pseudo-state-space representation – see Niedźwiecki and Ciołek (2014a) for more details.

In addition to localizing the detected noise pulse, once the detection process is completed, Kalman filter provides the model-based interpolation of the corrupted fragment. If the sequence of detection decisions is modified at the post-processing stage, as suggested in Niedźwiecki and Ciołek (2013) (widening and/or clustering detection alarms), Kalman filter has to be re-run to obtain new signal interpolations. Alternatively, a non-recursive (batch)
5. EXPERIMENTAL RESULTS

To compare detection/reconstruction efficiency of different approaches, we used 6 artificially corrupted stereo audio files, sampled at the rate of 48 kHz, with a dominant contribution of wind instruments (saxofon, trumpet, horn, clarinet). The 20 second long impulsive noise template, added to each clean audio recording, was made up of 957 pairs of equally spaced noise pulses extracted from silent parts of archive audio recordings: 212 pulses corrupting the left channel only, 262 pulses corrupting the right channel only, and 483 pulses corrupting both channels.

Performance evaluation was made for 4 approaches based on scalar (AR, SAR) and vector (VAR, SVAR) modeling of audio signals. In the first case, two tracks of each stereo recording were split and processed separately. In all cases the order of the short-term AR/VAR model was the same and equal to \( r = 10 \).

To compare in a fair way 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 Niedźwiecki (2001). Since vector models require estimation of two times more coefficients than their scalar counterparts, to guarantee fair comparison the estimation window widths were set to \( N = 1000, L = 80 \) for scalar modeling and to \( N = 2000, L = 160 \) for vector modeling. The fundamental period estimation settings were equal to: \( M_0 = 160, M = 40, m_0 = 20, m = 5, T_{\text{min}} = 50, T_{\text{max}} = 1200, k_0 = 3 \). The adopted stability margin was \( \epsilon = 0.01 \). Finally, the detection multiplier was set to \( \mu = 4 \).

To evaluate performance of different detection/reconstruction algorithms, we used the perceptual evaluation of audio quality (PEAQ) tool – see ITU-R Recommendation (1998) and Kabal (2003). PEAQ scores take negative values that range from -4 (very annoying distortions) to 0 (imperceptible distortions). The PEAQ standard uses a number of psycho-acoustical 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. Some caution is still required in the interpretation of PEAQ scores. While in telecommunication applications signal distortions are more or less evenly spread over time, in our current context they affect only isolated fragments of the audio material. As a result, much higher (i.e., much closer to 0) values of the PEAQ score, which is a “per sample” distortion measure, are needed to guarantee high quality of the restored audio. 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.

Tab. 1 summarizes performance statistics (PEAQ) for the compared approaches. Additionally, the ground truth (GT) results are shown [obtained when the signal is reconstructed under the perfect knowledge of pulse location: \( d(t) \equiv d(t) \)], as well as the results obtained for the corrupted signals prior to reconstruction (REF). The results summarized in Tab. 1 show clearly advantages of vector modeling (VAR versus AR, SVAR versus SAR) and advantages of sparse modeling (SAR versus AR, SVAR versus VAR). The best results are obtained when both techniques are combined, i.e., when the SVAR approach is used.
Table 1. Comparison of the PEAQ scores obtained for the results of univariate (AR, SAR) and multivariate (VAR, SVAR) processing of 6 artificially corrupted audio recordings: REF denotes the score of the input (corrupted) recording, PEAQ denotes the score of the processed recording, and GT 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. The best scores are shown in boldface.

| File | REF | Model | GT   | PEAQ   |
|------|-----|-------|------|--------|
| 1    | -3.845 | AR    | -0.182 | -1.589 |
|      |       | SAR   | -0.082 | -1.342 |
|      |       | VAR   | -0.025 | -0.648 |
|      |       | SVAR  | -0.013 | -0.602 |
| 2    | -3.724 | AR    | -0.147 | -1.101 |
|      |       | SAR   | -0.062 | -0.826 |
|      |       | VAR   | -0.133 | -0.579 |
|      |       | SVAR  | -0.031 | -0.553 |
| 3    | -3.802 | AR    | -0.147 | -0.671 |
|      |       | SAR   | -0.085 | -0.567 |
|      |       | VAR   | -0.070 | -0.491 |
|      |       | SVAR  | -0.053 | -0.329 |
| 4    | -3.666 | AR    | -0.395 | -2.905 |
|      |       | SAR   | -0.420 | -1.830 |
|      |       | VAR   | -0.983 | -2.698 |
|      |       | SVAR  | -0.025 | -0.784 |
| 5    | -3.700 | AR    | -0.313 | -3.150 |
|      |       | SAR   | -0.167 | -1.826 |
|      |       | VAR   | -0.042 | -2.603 |
|      |       | SVAR  | -0.012 | -0.688 |
| 6    | -3.836 | AR    | -0.123 | -1.254 |
|      |       | SAR   | -0.101 | -1.010 |
|      |       | VAR   | -0.165 | -0.852 |
|      |       | SVAR  | -0.076 | -0.675 |

6. CONCLUSION

In this paper we presented a new method for elimination of impulsive disturbances from stereo audio recordings. The new approach is based on sparse vector autoregressive modeling of audio signal. To satisfy the stability requirement, the sparse vector autoregressive signal representation is sought in a factorized form, as a cascade of a short-term model, taking care of the short-term signal correlations, and a long-term model, taking care of its long-term correlations. The paper is illustrated by results obtained for artificially corrupted audio signals. Experimental results confirm good detection and reconstruction capabilities of the proposed approach, better than those offered by the classical autoregressive approach.

REFERENCES

Canazza S, De Poli G., & Mian G.A. (2010). Restoration of audio documents by means of extended Kalman filter. *IEEE Trans. Audio, Speech Language Process.*, 18, 1107-1115.

Godsill S.J., & Rayner P.J.W. (1995a). A Bayesian approach to the restoration of degraded audio signals. *IEEE Trans. Speech, Audio Process.*, 3, 267–278.

Godsill S.J., & Rayner P.J.W. (1995b) Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Trans. Speech, Audio Process.*, 6, 352–372.

Godsill S.J., & Rayner P.J.W. (1998). *Digital Audio Restoration*. Springer-Verlag.

ITU-R Recommendation BS.1387. (1998). Method for Objective Measurements of Perceived Audio Quality.

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

Niedźwiecki M. (1993a) Statistical reconstruction of multivariate time series. *IEEE Transactions on Signal Processing*, 41, 451–457.

Niedźwiecki M., & Cisowski K. (1993b). Adaptive scheme for elimination of broadband noise and impulsive disturbances from audio signals. *Proc. Quatrieme Colloque GRETSI*, Juan-les-Pins, France, 519–522.

Niedźwiecki M., & Cisowski K. (1996). Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals. *IEEE Transactions on Signal Processing*, 44, 528–537.

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

Niedźwiecki M. (2001). *Identification of Time-varying Processes*, Wiley.

Niedźwiecki M., & Ciolek M. (2013). Elimination of impulsive disturbances from archive audio signals using bidirectional processing. *IEEE Transactions on Audio, Speech and Language Processing*, 21, 1046–1059.

Niedźwiecki M., & Ciolek M. (2014a). Renovation of archive audio recordings using sparse autoregressive modeling and bidirectional processing. ICASSP 2014, Proc. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 5949 – 5953.

Niedźwiecki M., & Ciolek M. (2014b). Elimination of impulsive disturbances from stereo audio recordings. EUSIPCO 2014, Proc. 2014 European Signal Processing Conference, Lisbon, Portugal, 1 – 5.

Ramachandran P.R., & Kabal P. (1987). Stability and performance analysis of pitch filters in speech coders. *IEEE Trans. Acoust., Speech, Signal Process.*, 35, 937–946.

Söderström T. & Stoica P. (1988). *System Identification*. Prentice-Hall.

Vaseghi S.V., & Rayner P.J.W. (1990). Detection and suppression of impulsive noise in speech communication systems. *IEEE Proceedings*, 137, 38–46.

Vaseghi S.V., & Frayling-Cork R. (1992). Restoration of old gramophone recordings. *J. Audio Eng. Soc.*, 40, 791–801.

Vaseghi S.V. (2008). *Advanced Signal Processing and Digital Noise Reduction*, Wiley.

Wolfe J., Garnier M., & Smith J. (2009). Vocal tract resonances in speech, singing, and playing musical instruments. *HFSP J.*, 3, 6–23.