This is a post-peer-review, pre-copyedit version of an article published in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). The final authenticated version is available online at: https://doi.org/10.1109/ICASSP.2017.7952240

DETECTION OF IMPULSIVE DISTURBANCES IN ARCHIVE AUDIO SIGNALS

Marcin Ciołek and Maciej Niedźwiecki

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control
Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk , Poland
marcin.ciolek@pg.gda.pl, maciekn@eti.pg.gda.pl

ABSTRACT

In this paper the problem of detection of impulsive disturbances in archive audio signals is considered. It is shown that semi-causal/noncausal solutions based on joint evaluation of signal prediction errors and leave-one-out signal interpolation errors, allow one to noticeably improve detection results compared to the prediction-only based solutions. The proposed approaches are evaluated on a set of clean audio signals contaminated with real click waveforms extracted from silent parts of old gramophone recordings.

Index Terms— Audio signal restoration, outlier detection

1. INTRODUCTION

Impulsive disturbances, such as clicks, crackles and record scratches are usually caused by aging and/or mishandling of the surface of gramophone records, specs of dust and dirt etc. [1],[2]. Elimination of noise pulses from archive audio documents is an important element of saving our cultural heritage.

Most approaches to audio restoration are based on an autoregressive (AR) or an autoregressive moving average signal representation [3]–[13]. When stereo audio recordings are restored, a vector autoregressive signal representation can be alternatively used [14], [15]. In recent studies also the time-frequency approaches were successfully applied [16], [17].

For the sake of simplicity, in this paper we will deal only with the problem of elimination of impulsive disturbances, i.e., we will assume that the measured signal $y(t)$ has the form

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) discrete time, $s(t)$ denotes the noiseless signal and $\delta(t)$ is the noise pulse sample. It is assumed that the noise pulse sequence is statistically independent of $s(t)$, and that the pulses, varying in length and shape, are sparsely distributed in time. Let $s(t)$ be represented by the AR model of order n

$$s(t) = \sum_{i=1}^{n} a_i s(t-i) + \eta(t), \quad \mathrm{var}[\eta(t)] = \rho \tag{2}$$

where a_i is the i-th autoregressive coefficient and $\{\eta(t)\}$ denotes white noise sequence with variance ρ. Denote by $d(t)$ the noise pulse location function: $d(t) = 1$ will further mean that the sample $y(t)$ is corrupted with a noise pulse, otherwise $d(t) = 0$. Our goal will be to precisely localize noise pulses, i.e., to obtain a good estimate $\widehat{d}(t)$ of the function $d(t)$.

Classical approaches [3]–[5], based on examination of residual errors, suffer from some negative effects known as outlier masking and outlier smearing, and may fail to detect small noise pulses. More sophisticated approaches [6]–[11] work satisfactory when signal characteristics change in a smooth manner but in the presence of abrupt changes they usually generate false alarms. When the entire history of the analyzed signal is available this drawback can be eliminated using noncausal approaches [12], [13].

The paper presents a new pulse detection rule which is based on evaluation of both signal prediction errors and leave-one-out signal interpolation errors. It will be shown that such a strengthened decision rule significantly improves properties of both causal and noncausal detection schemes.

2. SIGNAL RECONSTRUCTION

According to [18], the AR-model based reconstruction of corrupted samples can be carried out independently, without any information loss, for each local analysis frame starting and ending with n undistorted samples $y(t) = s(t)$.

In the simplest case, if one sample is missing at instant t, the interpolation formula can be derived in the form

$$\begin{aligned} \widetilde{s}(t) &= \sum_{i=1}^{n} c_i [s(t-i) + s(t+i)] \\ c_i &= [a_i - \sum_{j=1}^{n-i} a_j a_{j+i}]/[1 + \sum_{j=1}^{n} a_j^2] \end{aligned} \tag{3}$$

where c_i is the i-th interpolation coefficient. The variance ρ_*, $\rho_* < \rho$, of the leave-one-out signal interpolation error $e_*(t)$

$$e_*(t) = s(t) - \widetilde{s}(t) \qquad \mathrm{var}[e_*(t)] = \rho_* \tag{4}$$

is related to the variance ρ by the equation

$$\rho_* = \rho/[1 + \sum_{j=1}^{n} a_j^2]. \tag{5}$$

*Calculations were carried out at the Academic Computer Centre in Gdańsk.

3. ESTIMATION OF SIGNAL PARAMETERS

When the investigated process is nonstationary, but its characteristics vary slowly with time, parameters of the AR model can be obtained using local estimation technique, e.g. by means of processing a fixed-length data segment $\{y(t-2k),\dots,y(t)\}$ of length $M = 2k+1$. The order of autoregression can be fixed or chosen adaptively using the generalized Akaike's criteria [19]. The estimates of AR parameters, $\widehat{a}_1(t),\dots,\widehat{a}_n(t)$ and $\widehat{\rho}(t)$, can be obtained by solving the following set of Yule-Walker (YW) equations

$$[1, -\widehat{a}_1(t),\dots,-\widehat{a}_n(t)]\,\mathbf{R}(t) = [\widehat{\rho}(t), 0, \dots, 0] \tag{6}$$

$$\mathbf{R}(t) = \begin{bmatrix} r_0(t) & \dots & r_n(t) \\ \vdots & \ddots & \vdots \\ r_n(t) & \dots & r_0(t) \end{bmatrix}$$

where

$$r_i(t) = \frac{1}{L} p_i(t), \quad i = 0,\dots,n \tag{7}$$

can be interpreted as a local estimate of the i-th autocorrelation coefficient of $y(t)$. The quantity $p_i(t)$ is given in the form

$$p_i(t) = \sum_{l=-k+i}^{k} w(l)w(l-i)h_i(t-k+l) \tag{8}$$

where $h_i(t) = y(t)y(t-i)$ and $w(l)$, $w(l) > 0$ for $l \in [-k,k]$, is a bell-shaped tapering function taking its largest value in the center and smoothly decaying to 0 at the edges. Finally, the normalizing constant in (7) takes the form

$$L = \sum_{l=-k}^{k} w^2(l). \tag{9}$$

Since $\mathbf{R}(t)$ is, by construction, positive definite and Toeplitz, the YW estimates (6) can be obtained using the Levinson-Durbin algorithm which guarantees model stability [20].

Since the unweighted YW estimates, ($w(l) = 1$, $\forall l$), are identical with the least squares estimates obtained for the original data sequence extended with n zero samples at the segment beginning and at its end, data tapering allows one to smooth out signal discontinuities introduced by such a modification, and hence to reduce the associated estimation bias [21]. We advice to use the cosinusoidal window, given by

$$w(l) = \cos\frac{\pi l}{2(k+1)}, \quad l \in [-k,k], \quad L = k+1. \tag{10}$$

We note that $w^2(l)$ is identical with the Hann (raised cosine) window. The cosinusoidal window offers good bias-variance tradeoff and allows for recursive computation of (8)

$$\begin{aligned} p_i(t) &= \frac{1}{2} f_i(t)\cos\tfrac{\pi i}{2(k+1)} + \frac{1}{2}\mathrm{Re}[g_i(t)e^{-\frac{j\pi i}{2(k+1)}}] \\ g_i(t+1) &= e^{-\frac{j\pi}{k+1}} g_i(t) + e^{\frac{j\pi i}{k+1}} h_i(t-2k+i) + e^{\frac{j\pi k}{k+1}} h_i(t+1) \\ f_i(t+1) &= f_i(t) - h_i(t-2k+i) + h_i(t+1) \end{aligned}$$

where

$$g_i(t) = \sum_{l=-k+i}^{k} h_i(t-k+l)e^{\frac{j\pi l}{k+1}},\; f_i(t) = \sum_{l=-k+i}^{k} h_i(t-k+l).$$

Computation of $p_i(t)$ requires 12 real multiply-add operations per time update for each lag, and does not depend on M.

To protect the identification algorithm against outliers, parameter estimation should be suspended at the beginning of each detection alarm and resumed after signal reconstruction.

4. ADAPTIVE DETECTION

4.1. Causal detection

The popular noise pulse detection scheme is based on monitoring signal prediction errors: detection alarm is raised at the instant t_0+1 if the prediction error statistic $\alpha(t_0+1|t_0)$ exceeds detection threshold μ_α^2:

$$\widehat{d}(t_0+1) = \begin{cases} 1 & \text{if} \quad \alpha(t_0+1|t_0) > \mu_\alpha^2 \\ 0 & \qquad \text{elsewhere} \end{cases} \tag{11}$$

where

$$\begin{aligned} \alpha(t_0+1|t_0) &= e^2(t_0+1|t_0)/\widehat{\rho}(t_0) \\ \mathrm{e}(t_0+1|t_0) &= y(t_0+1) - \sum_{i=1}^{n}\widehat{a}_i(t_0)y(t_0-i+1) \end{aligned} \tag{12}$$

and the multiplier μ_α is chosen so as to guarantee that

$$P\left(\alpha(t_0+1|t_0) > \mu_\alpha^2 \mid d(t_0+1) = 0\right) = \epsilon \tag{13}$$

and ϵ, $0 < \epsilon \ll 1$ denotes the significance level. Under Gaussian assumptions, for $\epsilon = 0.003$ one obtains $\mu_\alpha = 3$, which corresponds to the well-known "3-sigma" rule used to detect outliers in Gaussian signals. Since real signals are usually non-Gaussian, $\mu_\alpha \in [3, 4.5]$ is a typical choice.

Once the detection alarm is triggered, the test is extended to multi-step-ahead predictions using the open-loop or decision-feedback scheme, derived for the state space description of the AR signal

$$\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}\mathbf{x}(t) + \mathbf{C}\eta(t+1) \\ y(t) &= \mathbf{C}^{\mathrm{T}}\mathbf{x}(t) + \delta(t) \end{aligned} \tag{14}$$

where

$$\mathbf{A} = \left[\begin{array}{c} \boldsymbol{\theta}^{\mathrm{T}} \\ \hline \mathbf{I}_{n-1}\;\; \mathbf{0}_{n-1} \end{array}\right], \quad \mathbf{C} = \left[\begin{array}{c} 1 \\ \hline \mathbf{0}_{n-1} \end{array}\right],$$

and $\mathbf{x}(t) = [s(t),\dots,s(t-n+1)]^{\mathrm{T}}$ denotes the state vector, $\boldsymbol{\theta}^{\mathrm{T}} = [a_1,\dots,a_n]$ denotes the row vector of AR coefficients, and $\mathbf{I}_{n-1}$ denotes the $(n-1)\times(n-1)$ identity matrix. Detection algorithms, which should be started at the instant t_0+1 and continued for $t \geq t_0+1$, can be summarized as follows

$$\begin{aligned} \widehat{\mathbf{x}}(t|t-1) &= \widehat{\mathbf{A}}(t_0)\widehat{\mathbf{x}}(t-1|t-1) \\ \mathbf{Q}(t|t-1) &= \widehat{\mathbf{A}}(t_0)\mathbf{Q}(t-1|t-1)\widehat{\mathbf{A}}^{\mathrm{T}}(t_0) + \widehat{\rho}(t_0)\mathbf{C}\mathbf{C}^{\mathrm{T}} \\ \sigma^2(t|t-1) &= \mathbf{C}^{\mathrm{T}}\mathbf{Q}(t|t-1)\mathbf{C} \\ \mathrm{e}(t|t-1) &= y(t) - \mathbf{C}^{\mathrm{T}}\widehat{\mathbf{x}}(t|t-1) \\ \alpha(t|t_0) &= \mathrm{e}^2(t|t-1)/\sigma^2(t|t-1) \end{aligned} \tag{15a}$$

open-loop variant

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

decision-feedback variant

if $\alpha(t|t_0) \le \mu_\alpha^2$:

$$\begin{aligned}\mathbf{L}(t) &= \mathbf{Q}(t|t-1)\mathbf{C}/\sigma^2(t|t-1)\\ \widehat{\mathbf{x}}(t|t) &= \widehat{\mathbf{x}}(t|t-1) + \mathbf{L}(t)\mathrm{e}(t|t-1)\\ \mathbf{Q}(t|t) &= \mathbf{Q}(t|t-1) - \mathbf{L}(t)\sigma^2(t|t-1)\mathbf{L}^{\mathrm{T}}(t)\end{aligned} \tag{15c}$$

if $\alpha(t|t_0) > \mu_\alpha^2$:

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

where initial conditions should be set to $\mathbf{Q}(t_0|t_0) = \mathbf{O}_n$, $\widehat{\mathbf{x}}(t_0|t_0) = [y(t_0), \dots, y(t_0 - n + 1)]^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}}^{\mathrm{T}}(t_0) = [\widehat{a}_1(t_0), \dots, \widehat{a}_n(t_0)]$.

Detection alarm started at the instant t_0+1 is terminated at the instant t_0+k_0+1 if $\alpha(t_0+k_0|t_0) > \mu_\alpha^2$ and n consecutive prediction errors are sufficiently small

$$\begin{aligned}\alpha(t|t_0) = \mathrm{e}^2(t|t-1)/\sigma^2(t|t-1) \le \mu_\alpha^2\\ t = t_0 + k_0 + 1, \dots, t_0 + k_0 + n\end{aligned} \tag{16}$$

or if k_0 reaches its maximum allowable value $k_{\max}$ (which plays the role of a "safety valve").

While the open-loop scheme (15a,15b) detects an entire block of corrupted samples, the decision-feedback scheme (15a,15c) approves/rejects samples in a sequential way, one by one. In this case the multi-step-ahead signal prediction and the corresponding variance of prediction error at the instant t depend on earlier decisions of the outlier detector, i.e., on decisions made prior to t. However, to avoid negative effects of "accidental approvals", when detection alarm is terminated according to (16), both rejected and approved samples (if any) are scheduled for reconstruction, i.e., the final form of the detection alarm is $\widehat{d}(t_0 + 1) = \cdots = \widehat{d}(t_0 + k_0) = 1$.

4.2. Semi-causal detection

When detection threshold μ_α is low, causal detector is prone to raise many false alarms. We will introduce a new semi-causal detector which not only reduces significantly the number of false alarms, but also allows one to lower detection threshold making the detector more sensitive to noise pulses.

Consider a block of corrupted samples starting at instant $t_0 + 1$ and ending at instant $t_0 + k_0$. According to (3-5), the adaptive leave-one-out signal interpolation formula can be written down in the form

$$\begin{aligned}\widetilde{y}(t|t_0) &= \sum_{i=1}^{n} \widehat{c}_i(t_0)[y(t-i) + y(t+i)]\\ \widehat{c}_i(t_0) &= [\widehat{a}_i(t_0) - \sum_{j=1}^{n-i} \widehat{a}_j(t_0)\widehat{a}_{j+i}(t_0)]/[1 + \sum_{j=1}^{n} \widehat{a}_j(t_0)^2].\end{aligned} \tag{17}$$

The estimate of ρ_* (the interpolation error variance) can be obtained from

$$\widehat{\rho}_*(t_0) = \widehat{\rho}(t_0)/[1 + \sum_{j=1}^{n} \widehat{a}_j^2(t_0)]. \tag{18}$$

The proposed detection scheme is based on monitoring of both prediction error based $[\alpha(t|t_0)]$ and interpolation error based $[\beta(t|t_0)]$ statistics, where

$$\begin{aligned}\beta(t|t_0) &= [\mathrm{e}_*(t|t_0)]^2/\widehat{\rho}_*(t_0)\\ \mathrm{e}_*(t|t_0) &= y(t) - \widetilde{y}(t|t_0).\end{aligned} \tag{19}$$

In the presence of noise pulse starting at the instant $t_0 + 1$, the quantity $\beta(t|t_0)$, $t \in [t_0 + 2 - n, t_0 + k_0 + n + 1]$ takes usually much higher values than its prediction based counterpart [since it depends on both past and future values of $\delta(t)$]. This suggests that it may be worthwhile to use two different detection thresholds μ_β and μ_α, such that $\mu_\beta \ge \mu_\alpha$.

The following strengthened alarm triggering rule is proposed

$$\widehat{d}(t_0 + 1) = \begin{cases} 1 & \text{if } \alpha(t_0 + 1|t_0) > \mu_\alpha^2 \text{ and} \\ & \beta(t|t_0) > \mu_\beta^2 \text{ for at least one } t \in T_\beta(t_0) \\ 0 & \text{elsewhere} \end{cases} \tag{20}$$

where $T_\beta(t_0) = [t_0 + 2 - n, t_0 + 1]$. Note that according to (20) the prediction based detection alarm is confirmed if $\beta(t|t_0)$ exceeds the corresponding detection threshold at least once in the recent past – otherwise the prediction based detection alarm is canceled. Since noise pulses may form complex oscillatory patterns, the rule (20) usually performs better than the strictly synchronized alarm triggering rule: $\widehat{d}(t_0 + 1) = 1$ if $\alpha(t_0 + 1|t_0) > \mu_\alpha^2$ and $\beta(t_0 + 1|t_0) > \mu_\beta^2$, which may fail to correctly detect the beginning of the pulse when excessive prediction error $e(t_0+1|t_0)$ is "accidentally" accompanied by a small interpolation error $e_*(t_0 + 1|t_0)$.

Unlike prediction parameters, $\widehat{\rho}(t)$ and $\widehat{a}_i(t)$, which must be updated in a continuous manner, the interpolation parameters, $\widehat{\rho}_*(t)$ and $\widehat{c}_i(t)$, have to be computed only at the instants t where $\alpha(t + 1|t) > \mu_\alpha^2$ (to confirm or cancel the prediction based detection alarm). Since the detection scheme based on (20) incorporates only n "future" signal samples $y(t_0 + 2), \dots, y(t_0 + n + 1)$, it will be referred to as semi-causal. It is suitable for real-time processing.

Once detection alarm is triggered, both prediction error based and interpolation error based statistics are updated and combined to decide upon the termination point. The alarm started at the instant $t_0 + 1$ is terminated at the instant $t_* = t_0 + k_0^* + 1$ if one of two stop conditions is fulfilled

$$\begin{aligned}&\alpha(t_* - 1|t_0) > \mu_\alpha^2 \ \text{ and } \ \alpha(t_* - 1 + i|t_0) \le \mu_\alpha^2,\\ &\qquad \text{or}\\ &\beta(t_* - 1|t_0) > \mu_\beta^2 \ \text{ and } \ \beta(t_* - 1 + i|t_0) \le \mu_\beta^2,\\ &\qquad i = 1, \dots, n\end{aligned} \tag{21}$$

or if $k_0^* = k_{\max}$. According to (21), when the prediction alarm lasts longer than the interpolation one, the termination point t_* coincides with the end of the interpolation alarm.

4.3. Noncausal detection

If the entire history of the analyzed signal $y(1), \ldots, y(N)$ is available, detection of impulsive disturbances can be carried out more efficiently if the results of forward-time (causal or semi-causal) outlier detection are combined with the analogous results of backward-time detection. The set of local fusion rules, allowing one to combine forward/backward detection alarms, was experimentally established in [13].

5. SIMULATION RESULTS

To evaluate the detection algorithms we used $40 = 4 \times 10$ clean audio recordings representing 4 music categories (jazz, choir, opera, classical), lasting for about 22 seconds each, sampled at the rate of 48 kHz and contaminated with real click waveforms extracted from silent parts of old gramophone recording[1]. 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 (the same for all recordings) consisted of 2175 noise pulses, with width ranging between 4 and 31 samples, picked at random from the click database containing 710 waveforms. The total number of corrupted samples was equal to 23898, which constitutes 2.49% of all samples in the analyzed fragment.

All detectors incorporated AR models of order $n = 10$. AR parameters were obtained using the local estimation technique with data tapering ($M = 481$). Detection threshold μ_α was restricted to the range $[3, 4.5]$. For the values of μ_α higher than 4.5 the noise pulse detector overlooks small pulses and/or yields undersized detection alarms. For $\mu_\alpha < 3$ the noise pulse detector triggers a large number of false and/or oversized detection alarms. In the case of semi-causal detection, detection threshold μ_β was set to 4.5. The maximum length of detection alarm was set to $k_{\max} = 50$. Once the noise pulse was localized the corrupted fragment of the signal was reconstructed using the least squares interpolation method described in [18].

Our evaluation was performed using the Perceptual Evaluation of Audio Quality (PEAQ) tool – a specialized software which scores the restored audio (by comparing it with the original, noiseless recording) using several perceptual criteria [22], [23]. Even though PEAQ 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 [11]. PEAQ scores take negative values that range from -4 (very annoying distortions) to 0 (imperceptible distortions). In the impulsive noise removal context, improvement of the PEAQ score by 0.1 (or more) is usually audible, i.e., perceptually significant.

[1]All algorithms (the MATLAB code) and all recordings, along with the results of their processing, are available through the website: http://eti.pg.edu.pl/katedra-systemow-automatyki/ICASSP2017

μ_α	A	B	C	D	A^*	B^*	C^*	D^*
4.5	-3.32	-3.30	-1.00	-0.88	-1.40	-1.58	-0.53	-0.43
4	-3.31	-3.28	-1.24	-0.83	-1.38	-1.54	-0.69	-0.44
3.5	-3.31	-3.26	-2.23	-0.79	-1.42	-1.49	-1.22	-0.45
3	-3.33	-3.23	-3.41	-0.77	-1.65	-1.41	-2.80	-0.47

Table 1: Comparison of the average PEAQ scores obtained for 8 unidirectional/bidirectional detection algorithms described in the paper. All results were obtained for 40 artificially corrupted audio files. The average score of the input (corrupted) recordings was equal to -3.6 and the average "ground truth" score, obtained when interpolation of the corrupted samples was based on exact knowledge of pulse locations, was equal to -0.29. Interpretation of PEAQ scores: 0 = imperceptible (signal distortions), -1 = perceptible but not annoying, -2 = slightly annoying, -3 = annoying, -4 = very annoying.

Table 1 shows the average PEAQ scores obtained for 4 unidirectional algorithms: causal, equipped with open-loop detection scheme (A), semi-causal, equipped with open-loop detection scheme (B), causal, equipped with decision-feedback detection scheme (C), semi-causal, equipped with decision-feedback detection scheme (D), and 4 bidirectional algorithms, i.e. noncausal extensions of algorithms A, B, C, D denoted by A^*, B^*, C^*, D^*, respectively. The results show clearly advantages of bidirectional processing (A, B, C, D vs. A^*, B^*, C^*, D^*, respectively), advantages of the decision-feedback strategy (A vs. C, B vs. D, A^* vs. C^*, B^* vs. D^*), and advantages of applying semi-causal detection (A vs. B, C vs. D, C^* vs. D^*). Notice that semi-causal detection is robust to the choice of detection threshold μ_α which confirms its ability to reliably cancel false alarms.

Next, the approach D^*, with the best average PEAQ scores, was compared against the state-of-the-art restoration algorithm – the combined least squares autoregressive+sinusoid (LSAR+SIN) method, proposed in [2]. This algorithm is a well-known standard in professional audio restoration. We used the source code with default settings which was recently tested in [16], and is provided on the web page of Nuzman [24]. The average PEAQ score obtained for LSAR+SIN was -0.88 which shows that the approach D^* is superior to LSAR+SIN. Informal listening tests, performed on real archive gramophone recordings[1] corrupted with pops, clicks and crackles, support the above findings.

6. CONCLUSION

The problem of elimination of impulsive disturbances from archive audio signals was considered and new pulse detection rules, combining analysis of one-step-ahead signal prediction errors with critical evaluation of leave-one-out signal interpolation errors, were proposed. The new detectors have increased ability to reliably cancel false alarms. Perceptual scores, obtained using the PEAQ tool, confirm that the proposed detection rules yield better results than the classical ones, based on evaluation of signal prediction errors only.

7. REFERENCES

[1] S.V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 4th ed. John Wiley and Sons Ltd., 2008.

[2] J.S. Godsill and J.P.W. Rayner, *Digital Audio Restoration*. London, U.K.: Springer, 1998.

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

[4] S.V. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings," *J. Audio Eng. Soc.* vol. 40, no. 10, pp. 791–801, 1992.

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

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

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

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

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

[10] M. Niedźwiecki and M. Ciołek, "Elimination of clicks from archive speech signals using sparse autoregressive modeling," in *Proc. 20th Eur. Signal Process. Conf.*, Bucharest, Romania, pp. 2615–2619, 2012.

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

[12] S. Canazza, G. De Poli, and G. Mian, "Restoration of audio documents by means of extended Kalman filter,", *IEEE Trans. Audio, Speech Language Process.* vol. 18, no. 6, pp. 1107–1115, 2010.

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

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

[15] M. Niedźwiecki and M. Ciołek, "Sparse vector autoregressive modeling of audio signals and its application to the elimination of impulsive disturbances, in: *17th IFAC Symposium on System Identification*, Beijing, China, pp. 1202–1207, 2015.

[16] M. Ruhland, J. Bitzer, M. Brandt, and S. Goetze, "Reduction of Gaussian, supergaussian, and impulsive noise by interpolation of the binary mask residual," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1680 – 1691, 2015

[17] R. Tong, Y. Zhou, L. Zhang, G. Bao, and Z. Ye, "A Robust time-frequency decomposition model for suppression of mixed Gaussian-impulse noise in audio signals", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 69–79, 2015.

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

[19] M. Niedźwiecki, "On the localized estimators and generalized Akaike's criteria" IEEE Trans. Automat. Contr., vol. 29, pp. 970–983, 1981.

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

[21] M.S. Kay, *Modern Spectral Estimation: Theory and Application*, Englewoods Cliffs NJ: Prentice-Hall, 1988.

[22] ITU-R Recommendation BS.1387-1, "Method for Objective Measurements of Perceived Audio Quality," 2001.

[23] 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.

[24] J. Nuzman, "Audio restoration: An investigation of digital methods for click removal and hiss reduction", University of Maryland, Institute for Advanced Computer Studies, 2004 [Online]. Available: www.github.com/jnuzman/audio-restoration-2004, last seen in August 2016

[25] M. Niedźwiecki and M. Ciołek, "Adaptive detection of impulsive disturbances in multivariate signals," submitted to IEEE Transactions on Signal Processing.