RENOVATION OF ARCHIVE AUDIO RECORDINGS USING SPARSE AUTOREGRESSIVE MODELING AND BIDIRECTIONAL PROCESSING

Maciej Niedźwiecki and Marcin Ciołek

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

ABSTRACT

The paper presents a new approach to elimination of broadband noise and impulsive disturbances from archive audio recordings. The proposed adaptive Kalman-like algorithm, based on a sparse autoregressive model of the audio signal, simultaneously detects noise pulses, interpolates the irrevocably distorted samples and performs signal smoothing. It is shown that bidirectional (forward-backward) processing of the archive signal improves smoothing efficiency and allows one to localize noise pulses more accurately, leading to noticeable performance improvements compared to unidirectional processing.

***Index Terms*—** Restoration of audio signals, adaptive signal processing, sparse autoregressive models, bidirectional processing

1. INTRODUCTION

Archived audio recordings are often degraded by broadband noise and impulsive disturbances. Broadband noise, such as surface noise of magnetic tapes and phonograph records, is an inherent part of all analog recordings. Impulsive disturbances, such as clicks and pops, are usually caused by aging and/or mishandling of the recording medium, as well as by transmission or equipment artifacts – for more details see [1]–[7]. Elimination of both kinds of disturbances from archive audio documents is an important element of saving our cultural heritage.

We will assume that the sampled audio signal $y(t)$, corrupted by a mixture of a broadband noise $v(t)$ and impulsive disturbances $z(t)$, has the form

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) discrete time and $s(t)$ denotes the undistorted (clean) audio signal. We will regard $\{v(t)\}$ as a normally distributed white noise sequence $v(t) \sim \mathcal{N}(0, \sigma_v^2)$. We will not use any probabilistic model of the sequence of noise pulses. The following coarse "save or reject" model will be adopted instead

$$z(t) \sim \mathcal{N}(0, \sigma_z^2(t)), \;\; \sigma_z^2(t) = \begin{cases} 0 & \text{if} \quad d(t) = 0 \\ \infty & \text{if} \quad d(t) = 1 \end{cases}$$

where $d(t)$ denotes the pulse location function

$$d(t) = \begin{cases} 0 & \text{noise pulse absent} \\ 1 & \text{noise pulse present} \end{cases} .$$

Note that when $\sigma_z^2(t) = \infty$, the measurement $y(t)$ contains no useful information about the signal $s(t)$, i.e., it should be regarded as a missing sample.

Unlike [1]–[7], where restoration is based on an autoregressive (AR) or an autoregressive moving average (ARMA) signal representation, the following sparse autoregressive (SAR) model of the audio signal will be used for restoration purposes

$$s(t+1) = \sum_{i=1}^{r} \alpha_i(t)s(t-i+1) + \beta(t)s(t-T[t]+1) + n(t) \tag{2}$$

where $\alpha_1(t), \ldots, \alpha_r(t)$ and $\beta(t)$ denote the short-term and long-term autoregressive coefficients, respectively, $T[t] \gg r$ denotes the instantaneous fundamental period of the signal (e.g. in the case of speech signals the period of pitch excitation, if present), and $n(t) \sim \mathcal{N}(0, \sigma_n^2(t))$ denotes white driving noise, uncorrelated with the measurement noise. Even though formally of order $T[t]$, such a model is sparse as it contains only $r + 1 \ll T[t]$ nonzero coefficients. SAR models capture both short-term correlations, taken care of by the first component on the right-hand side of (2), and long-term correlations, taken care of by the second component on the right hand side of (2) of the analyzed time series. So far SAR models have been used mainly for the purpose of speech processing – see e.g. [8]–[11].

Assuming that the archived recording is processed off-line,i.e., that the entire measurement history $Y(N) = \{y(1), \ldots, y(N)\}$ is available, we will work out the estimate of the clean audio signal. This general problem requires solution of several subproblems, such as detection of noise pulses, interpolation of irrevocably distorted samples and signal smoothing.

2. UNIDIRECTIONAL PROCESSING

The algorithm proposed in this paper is made up of several simultaneously operated sub-algorithms used for signal estimation, signal identification and outlier detection, respectively. These sub-algorithms are appropriately coupled, e.g. the identification routines use the signal "cleaned up" by the estimation routine and the estimation routine uses current estimates of process coefficients, provided by the identification routines.

2.1. Signal interpolation/smoothing

The best, in the mean-square sense, estimates of the signal $s(t)$ governed by (1) – (2) can be obtained using the appropriately designed Kalman smoother [12]. However, since (2) is an autoregressive model of order $T[t]$, its equivalent state space description, which serves as a basis for designing Kalman algorithm, involves $T[t]$ space variables – much too large considering that the typical

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

values of $T[t]$ are well in excess of 100. For this reason some simplifications are needed to reduce computational complexity of the restoration algorithm.

Suppose that the maximum length of detection alarms is set to $k_{\max}$, which means that up to $k_{\max}$ samples in a row can be questioned and interpolated. Let $q = r + k_{\max} + 1$. We will assume that $q < T_{\min}$, where $T_{\min}$ denotes the minimum allowable value of the fundamental period $T[t]$. The following pseudo-state-space representation of the SAR model (2) will serve as a basis for designing the filtering/interpolation/smoothing algorithm

$$\begin{aligned}\boldsymbol{\psi}(t+1) &= \mathbf{A}(t)\boldsymbol{\psi}(t) + \mathbf{b}(t)u(t) + \mathbf{c}n(t)\\ y(t) &= \mathbf{c}^{\mathrm{T}}\boldsymbol{\psi}(t) + \zeta(t)\end{aligned} \tag{3}$$

where $\zeta(t) = v(t) + z(t)$, $\boldsymbol{\psi}(t) = [s(t), \dots, s(t-q+1]^{\mathrm{T}}$ denotes the "state" vector, $u(t)$ is an estimate of $s(t - T[t] + 1)$, treated as an exogenous (measurable) input signal, and

$$\mathbf{A}(t) = \begin{bmatrix} \widehat{\alpha}_1(t) & \cdots & \widehat{\alpha}_r(t) & 0 & \cdots & 0 & 0 \\ 1 & & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$\mathbf{b}(t) = [\widehat{\beta}(t), 0, \dots, 0]^{\mathrm{T}}, \ \ \mathbf{c} = [1, 0, \dots, 0]^{\mathrm{T}}.$$

According to (3), the SAR model (2) is treated as an ARX model (autoregressive with exogenous input), i.e., the quantity $s(t - T[t] + 1)$ appearing in the long-correlation term in (2) is regarded as an external excitation, rather as an internal (state) variable.

Designing the Kalman filter based on such an improper state space model, one arrives at the following reduced-complexity algorithm which preserves important features of the genuine Kalman interpolator/smoother

$$\begin{aligned}\widehat{\boldsymbol{\psi}}(t|t-1) &= \mathbf{A}(t-1)\widehat{\boldsymbol{\psi}}(t-1|t-1) + \mathbf{b}(t-1)u(t)\\ \boldsymbol{\Sigma}_{\psi}(t|t-1) &= \mathbf{A}(t-1)\boldsymbol{\Sigma}_{\psi}(t-1|t-1)\mathbf{A}^{\mathrm{T}}(t-1) + \mathbf{c}\mathbf{c}^{\mathrm{T}}\\ e(t) &= y(t) - \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\psi}}(t|t-1)\\ \mathbf{L}_{\psi}(t) &= \begin{cases} \frac{\boldsymbol{\Sigma}_{\psi}(t|t-1)\mathbf{c}}{\mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{\psi}(t|t-1)\mathbf{c} + \kappa(t-1)} & \text{if} \quad \widehat{d}(t) = 0 \\ 0 & \text{if} \quad \widehat{d}(t) = 1 \end{cases}\\ \widehat{\boldsymbol{\psi}}(t|t) &= \widehat{\boldsymbol{\psi}}(t|t-1) + \mathbf{L}_{\psi}(t)e(t)\\ \boldsymbol{\Sigma}_{\psi}(t|t) &= [\mathbf{I} - \mathbf{L}_{\psi}(t)\mathbf{c}^{\mathrm{T}}]\boldsymbol{\Sigma}_{\psi}(t|t-1)\end{aligned} \tag{4}$$

where $\kappa(t-1) = \widehat{\sigma}_v^2/\widehat{\sigma}_n^2(t-1)$, and $\widehat{d}(t)$ denotes decision of the outlier detector. Since $\widehat{\boldsymbol{\psi}}(t|t)$ is the estimate of $\boldsymbol{\psi}(t) = [s(t), \dots, s(t-q+1)]^{\mathrm{T}}$ based on the measurements $Y(t)$, the last element of $\widehat{\boldsymbol{\psi}}(t|t)$ can be regarded as a smoothed estimate of $s(t-q+1)$

$$\widehat{s}(t-q+1|t) = \widehat{\boldsymbol{\psi}}^{\mathrm{T}}(t|t)\mathbf{g}, \ \ \mathbf{g} = [0, \dots, 0, 1]^{\mathrm{T}}. \tag{5}$$

Finally, as the input signal $u(t)$, one can use the estimate

$$u(t) = \widehat{s}(t - \widehat{T}[t] + 1|t - \widehat{T}[t] + q). \tag{6}$$

2.2. Estimation of autoregressive coefficients

Using the shorthand notation $\boldsymbol{\eta}(t) = [s(t), \dots, s(t-r+1)]^{\mathrm{T}}$, $\boldsymbol{\varphi}(t) = [\boldsymbol{\eta}^{\mathrm{T}}(t), s(t - T[t] + 1)]^{\mathrm{T}}$, $\boldsymbol{\alpha}(t) = [\alpha_1(t), \dots, \alpha_r(t)]^{\mathrm{T}}$ and $\boldsymbol{\theta}(t) = [\boldsymbol{\alpha}^{\mathrm{T}}(t), \beta(t)]^{\mathrm{T}}$, the SAR model can be written down in the form

$$s(t+1) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + n(t).$$

The following exponentially weighted least squares (EWLS) algorithm can be used to track the time-varying parameters of the SAR model [14]

$$\begin{aligned}\widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\theta}}(t-1) + \mathbf{L}_{\theta}(t)\varepsilon(t)\\ \varepsilon(t) &= \widehat{s}(t|t) - \widehat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\widehat{\boldsymbol{\theta}}(t-1)\\ \mathbf{L}_{\theta}(t) &= \frac{\boldsymbol{\Sigma}_{\theta}(t-1)\widehat{\boldsymbol{\varphi}}(t-1)}{\lambda + \widehat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\boldsymbol{\Sigma}_{\theta}(t-1)\widehat{\boldsymbol{\varphi}}(t-1)}\\ \boldsymbol{\Sigma}_{\theta}(t) &= \frac{1}{\lambda}\left[\mathbf{I} - \mathbf{L}_{\theta}(t)\widehat{\boldsymbol{\varphi}}^{\mathrm{T}}(t-1)\right]\boldsymbol{\Sigma}_{\theta}(t-1)\end{aligned} \tag{7}$$

where λ, $0 < \lambda < 1$ denotes the forgetting constant and $\widehat{s}(t|t) = \widehat{\boldsymbol{\psi}}^{\mathrm{T}}(t|t)\mathbf{c}$. The quantity $\widehat{\boldsymbol{\varphi}}(t)$ denotes the estimate of the regression vector provided by the signal estimation sub-algorithm $\widehat{\boldsymbol{\varphi}}(t) = [\widehat{\boldsymbol{\eta}}^{\mathrm{T}}(t), u(t)]^{\mathrm{T}}$, where $\widehat{\boldsymbol{\eta}}(t)$ is made up of the first r components of $\widehat{\boldsymbol{\psi}}(t|t)$ and $u(t)$ is given by (6).

2.3. Estimation of the fundamental period

The usefulness of the SAR model (2) critically depends on precise knowledge of the fundamental period $T[t]$. In many cases detection/interpolation performance may drop significantly even if the estimated value of $T[t]$ differs from its true value by only one or two samples. Denote by $T_{\min}/T_{\max}$ the smallest/largest values of the fundamental period that will be considered. At each time instant t four competitive estimates of the fundamental period are considered: $\widehat{T}_i(t), i = 1, \dots, 4$. These estimates are obtained by means of minimizing, over $T \in [T_{\min}, T_{\max}]$, the sums of squared differences between the selected fragments of the analyzed audio signal. In the first two cases ($i = 1, 2$), one finds the best match between the m most recent samples $\widehat{s}(\cdot)$ (in our experiments m was set to 50) of the already processed signal and the analogous sequence derived either from the past of the signal $\widehat{s}(\cdot)$ ($i = 1$) or from the past of the original (unprocessed) signal $y(\cdot)$ ($i = 2$). The second variant helps one to avoid 'serial' detection errors which occur when a single incorrect decision of an outlier detector starts a chain of related 'derivative' decision errors. In the last two cases ($i = 3, 4$), the reference frame is made up of the $m - 10$ most recent samples and 10 'future' samples (not yet processed). This often allows one to obtain more precise estimates of the fundamental period – unless the 'future' samples are severely distorted. The final estimate $\widehat{T}[t]$ is the one that provides the best match.

2.4. Estimation of noise variances

The local (exponentially weighted) estimate of $\sigma_n^2(t)$ takes the form

$$\widehat{\sigma}_n^2(t) = \begin{cases} \lambda\widehat{\sigma}_n^2(t-1) + (1-\lambda)\varepsilon^2(t) & \text{if} \quad \widehat{d}(t) = 0 \\ \widehat{\sigma}_n^2(t-1) & \text{if} \quad \widehat{d}(t) = 1 \end{cases}.$$

The intensity of the measurement noise $v(t)$ is usually constant for a given recording and can be estimated by means of processing the "silent" fragments [$s(t) = 0$] of the analyzed recording.

2.5. Detection of noise pulses

Similar to [4]–[5], the outlier detection is based on monitoring prediction errors. The detection alarm starts [$\widehat{d}(t) = 1$] when the model-based one-step-ahead prediction error exceeds μ times its estimated standard deviation (typically $\mu \in [3, 5]$)

$$|e(t)| > \mu\widehat{\sigma}_e(t)$$

and it is continued [$\widehat{d}(t) = \ldots = \widehat{d}(t + k_0 - 1) = 1$] until r consecutive prediction errors are sufficiently small, namely: $|e(t + k_0 + i)| \leq \mu\widehat{\sigma}_e(t + k_0 + i)$, $i = 0, \ldots, r - 1$, or until the length of the detection alarm $D = [t, \ldots, t + k_0 - 1]$, equal to k_0, reaches its maximum allowable value $k_{\max}$. The estimates of the prediction error variance can be obtained from $\widehat{\sigma}^2_{e(t)} = \gamma(t)\widehat{\sigma}^2_n(t-1)$, where $\gamma(t) = \mathbf{c}^{\mathrm{T}}\mathbf{\Sigma}_\psi(t|t-1)\mathbf{c} + \kappa(t-1)$.

3. BIDIRECTIONAL PROCESSING

All causal detection algorithms, such as the one described in the previous section, localize and schedule for interpolation fragments that are "unpredictable", i.e., inconsistent with the signal past. However, outlier detectors based on forward consistency checks have some obvious limitations – whenever characteristics of the proposed audio signals change abruptly, e.g. at the beginning of new sounds, they generate false detection alarms. Since many of the questioned fragments are consistent with the signal future, rather than its past, the number of false alarms can be reduced if detection is based on backward consistency checks, which is possible when the analyzed signal is processed backward in time. The best performance can be achieved if the results of forward-time and backward-time detection/interpolation are combined appropriately. From this point on, we will assume that two detection signals are available: $\widehat{d}_{\mathrm{f}}(t)$ and $\widehat{d}_{\mathrm{b}}(t)$, obtained by means of forward-time and backward-time processing, respectively. The backward-time algorithm is identical with the forward-time one but it processes time-reversed data.

3.1. Preprocessing

Unlike artificially generated noise pulses, real impulsive disturbances corrupting audio signals are rarely confined to isolated samples. Moreover, most of them have "soft" edges (the more so, the higher sampling rate) which stems from the typical geometry of local damages of the recording medium (e.g. groove damages). The straightforward consequence of this fact is that detection alarms are seldom triggered at the very beginning of noise pulses. This may lead to small but audible distortions of the reconstructed audio material. Although detection delays can be reduced, or even eliminated, by lowering the detection multiplier μ, i.e., by making the outlier detector more sensitive to "unpredictable" signal changes, the improvement comes at a price: low detection thresholds may dramatically increase the number and length of detection alarms, causing the overall degradation of the results. An alternative solution, which works pretty well in practice, is based on shifting back the beginning of each detection alarm (once determined) by a small fixed number of samples further denoted by n_0.

3.2. Bidirectional detection

The simplest approach to combining results of forward-time and backward-time detection is the one based on global decision rules, such as the intersection rule ($\cap$) or the union rule ($\cup$). In the first case detection alarm is raised only when the sample is questioned by both detectors, and in the second case – when it is questioned by at least one of the detectors. Preliminary tests have shown that neither of these rules works satisfactorily in practice. The intersection rule is too conservative – it tends to overlook many small noise pulses and produces underfitted (too short) detection alarms. The union rule is too liberal – it yields many overfitted (too long) detection alarms which, after interpolation, result in audible signal distortions.

To avoid problems mentioned above, different configurations of forward and backward detection alarms, further referred to as detection patterns, were divided into several classes and subclasses. Each class was analyzed separately in order to determine the best way of combining detection alarms. The final detection decision is a result of application of a certain number of local, case-dependent decision rules, called atomic fusion rules, rather than using a single global rule applicable to all cases.

Detection alarms are sorted out in consecutive analysis frames, defined as the *minimum-length* intervals that start and end with r no-alarm decisions and contain at least one forward or backward detection alarm. Situations where the analysis frame covers at most one forward detection alarm and at most one backward detection alarm are referred to as elementary detection patterns. Elementary patterns were divided into three classes: A – two overlapping detection alarms, B – two non-overlapping detection alarms separated by less than r samples, and C – one forward or backward detection alarm. The fourth class D was made up of all complex detection patterns, i.e., those which incorporate more than 2 forward/backward detection alarms that cannot be subdivided into elementary patterns.

In addition to the union rule and the intersection rule mentioned above, three "nonstandard" fusion rules were examined: the "front edge – front edge" rule (FF), the "first edge – last edge" rule (FL), and the "front edge" rule (F). In the FF case the aggregated detection alarm starts at the front edge of the forward alarm, and ends at the front edge of the backward alarm (which, after time reversal, becomes its back edge). The FF rule is practically motivated – it is known that the moment of triggering the detection alarm is usually determined more precisely than the moment of its termination. Under the FL (compactified union) rule the alarm starts at the first edge of the forward/backward alarms, and ends at their last edge. Finally, according to the F rule (applied to class-C patterns only), the front/back edge sample is "sandwiched" between n_0 preceding samples (added at the preprocessing stage) and n_0 succeeding samples.

Fig. 1 summarizes experimental results obtained for all elementary patterns for 10 test recordings - clean audio signals (chosen so as to cover different temporal and spectral features of audio signals) contaminated with the sequence of noise pulses extracted from an archive gramophone recording. Such an "disturbance transplantation" technique allows one to check restoration algorithms under realistic conditions. Each test recording was obtained under the sampling rate of $f_{\mathrm{s}} = 22.05$ kHz and contained from 23 to 29 seconds of the audio material. To enable listening tests focused on a particular class of detection patterns, test recordings were prepared in a special way. For example, to compare 3 atomic fusion rules ($\cup, \cap$, FF) associated with the A_2 pattern (cf. Fig. 1), 3 variants of each test recording were created, confined to A_2 interventions only – all other analysis frames were filled with the undistorted audio material. Since such listening tests are very time-consuming we relied on the opinion of three experts in the field of sound restoration (experienced sound engineers). According to them, the best results were obtained using: the FF rule – for all A-class patterns and all D-class patterns (not shown in Fig. 1); the FL rule – for all B-class patterns, and the F rule – for all C-class patterns.

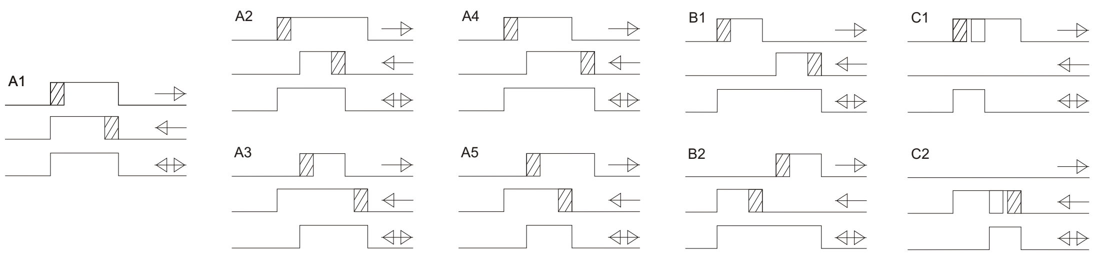

Fig. 1. Atomic fusion rules selected by experts. The plots show the results of forward detection (→), backward detection (←) and bidirectional detection (↔) for all elementary detection patterns. Shaded areas denote extensions added at the preprocessing stage

3.3. Bidirectional interpolation/smoothing

When the alarm fusion step is finished, i.e., when the forward detection signal $\widehat{d}_{\rm f}(t)$ is merged with the backward detection signal $\widehat{d}_{\rm b}(t)$, forming the bidirectional detection signal $\widehat{d}_{\rm fb}(t)$, the forward/backward estimation/identification algorithms summarized in sections 3.1 – 3.4 are run again, this time governed by the external detection signal $\widehat{d}_{\rm fb}(t)$. Denote the resulting signal estimates by $\widehat{s}_{\rm f}(t)$ and $\widehat{s}_{\rm b}(t)$. Following [13], the bidirectional signal estimate $\widehat{s}_{\rm fb}(t)$ can be obtained by means of computing convex combination of the results yielded by the forward-time and backward-time algorithms:

$$\widehat{s}_{\rm fb}(t) = w_{\rm f}(t)\widehat{s}_{\rm f}(t) + w_{\rm b}(t)\widehat{s}_{\rm b}(t),\ \ t = 1,\dots,N \tag{8}$$

where $w_{\rm f}(t) = \widehat{\sigma}^2_{n_{\rm f}}(t)/[\widehat{\sigma}^2_{n_{\rm f}}(t) + \widehat{\sigma}^2_{n_{\rm b}}(t)]$ and $w_{\rm b}(t) = \widehat{\sigma}^2_{n_{\rm b}}(t)/[\widehat{\sigma}^2_{n_{\rm f}}(t)+\widehat{\sigma}^2_{n_{\rm b}}(t)]$ are the weights that depend on the local predictive performance of both algorithms. Note that $w_{\rm f}(t) + w_{\rm b}(t) = 1, \forall t$.

4. EXPERIMENTAL RESULTS

Because of the lack of space, only the results of tests checking the declicking performance of the proposed algorithm will be reported here. Validation was based on two sets of recordings – 10 artificial and 10 authentic.

The artificially generated database was obtained by adding noise pulses to clean audio signals. The same set of audio recordings was used as that incorporated for selection of fusion rules, but the impulsive disturbances (606 noise pulses covering 4099 samples) were extracted from another archive recording. Hence, the performance of the proposed declicking procedure was checked on a different data set than that used earlier for training purposes. The proposed algorithm was run with default settings: $r = 6$, $k_{\max} = 125$, $T_{\min} = 130$, $T_{\max} = 600$, $\lambda = 0.99$, $\mu = 3.5$, $n_0 = 2$. Table 1 shows the results of comparison of 4 approaches to elimination of impulsive disturbances: the approach based on forward-time processing (traditional), the approach based on backward-time processing, the mixture approach of Canazza *et al.* [13], and the proposed bidirectional approach. The results of the ordering test show clearly superiority of the proposed method.

Table 2 shows the analogous results obtained when the proposed approach, based on local case-dependent alarm fusion rules, is compared with approaches based on two global case-independent rules: union rule and intersection rule. In this experiment the data base consisted of 10 real gramophone recordings, covering a wide range of musical styles, from classical music and opera, to pop and blues. It is clear from this comparison that the case-dependent rules yield better results than the case-independent ones.

Table 1. Comparison of declicking algorithms made by 20 test persons. The scores show the number of times where the evaluated algorithm yielded the best results within the analyzed group of recordings (more than one recording could be nominated).

Recording	Forward	Backward	Mixed	Proposed
1	0	0	0	20
2	0	1	0	20
3	0	0	0	20
4	0	0	0	20
5	1	0	0	20
6	0	1	0	19
7	0	1	0	19
8	0	0	0	20
9	0	0	1	19
10	0	0	0	20

Table 2. Comparison of the results of declicking based on the proposed local case-dependent alarm fusion rules with the analogous results obtained using the global case-independent rules: union and intersection (more than one recording could be nominated).

Recording	Intersection	Union	Proposed
1	0	8	17
2	0	2	18
3	0	2	20
4	0	1	20
5	0	12	15
6	0	8	19
7	0	7	17
8	0	14	17
9	0	4	19
10	0	5	17

5. RELATION TO PRIOR WORK

SAR models have so far been used only for processing (coding, declicking) speech signals [8] – [11]. The reduced-order Kalman-like restoration algorithm is the first SAR-based procedure capable of simultaneous reduction of impulsive disturbances and broadband noise corrupting archived audio signals. The second contribution of the paper is due to bidirectional processing. To the best of our knowledge, the idea of bidirectional processing was previously exploited only once, in the paper [13]. The method proposed there is based on combining restoration results obtained independently by means of forward-time and backward-time signal processing. Our approach is different. Based on the results of tests, performed on real audio signals, corrupted by real impulsive disturbances, we work out a set of local, case-dependent fusion rules that are further used to combine forward and backward detection alarms. This results in much better restored sound quality.

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 audio signals," *Proc. Quatrozieme Colloque GRETSI*, pp. 519–522, 1993.

[3] S.V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, Wiley, 1996.

[4] 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.

[5] M. Niedźwiecki,"Identification of time-varying processes in the presence of measurement noise and outliers," *SYSID.*, vol. 4, pp. 1765-1770, 1997.

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

[7] J.S. Godsill, and J.P.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] P.R. Ramachandran, and P. Kabal, "Stability and performance analysis of pitch filters in speech coders," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, pp. 937–946, 1987.

[9] P.R. Ramachandran, and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 467–478, 1989.

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

[11] M. Niedźwiecki and M. Ciołek, "Elimination of clicks from archive speech signals using sparse autoregressive modeling," Proc. 20th European Signal Processing Conference, Bucharest, Romania, 2012.

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

[13] 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. 41, pp. 1107-1115, 2010.

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