Elimination of impulsive disturbances from archive audio files – comparison of three noise pulse detection schemes

Maciej Niedźwiecki and Marcin Ciołek
Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk, Poland
e-mails: maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl

Summary
The problem of elimination of impulsive disturbances (such as clicks, pops, ticks, crackles, and record scratches) from archive audio recordings is considered and solved using autoregressive modeling. Three classical noise pulse detection schemes are examined and compared: the approach based on open-loop multi-step-ahead signal prediction, the approach based on decision-feedback signal prediction, and the double threshold approach, based on analysis of residual errors. It is shown that the accuracy of the classical schemes can be significantly improved by means of combining the results of forward time and backward time signal processing.

PACS no. 3.60.Mn, 43.60.Uv

1. Introduction
Archived audio recordings are often degraded by impulsive disturbances and wideband noise [1], [2]. Clicks, pops, ticks, crackles, and record scratches are caused by aging and/or mishandling of the surface of gramophone records (shellac or vinyl), specks of dust and dirt, faults in the record stamping process (e.g. gas bubbles), and slight imperfections in the record playing surface due to the use of coarse grain filters in the record composition. In the case of magnetic tape recordings, impulsive disturbances can be usually attributed to transmission or equipment artifacts (e.g. electric or magnetic pulses).

Wideband background noise, such as the so-called surface noise of magnetic tapes and phonograph records, is an inherent component of all analog recordings.

Elimination of both types of disturbances from archive audio documents is an important element of saving our cultural heritage. The Polish Radio Archives and the Polish National Library Archives alone contain more than one million archive audio documents with different content (historic speeches, interviews, concerts, studio music recordings etc.), saved on different media, such as piano rolls, phonograph and gramophone records, magnetic tapes etc.

The British Library Sound Archive (which is among the largest collections of recorded sound in the world) holds over three million recordings, including over a million of disks and 200,000 tapes. Digitization of these documents is an ongoing process (in Poland carried out, among others, by the Polish National Digital Archives), which will be very soon followed by the next, obvious step – audio restoration. This makes research on audio restoration technology both practically useful and timely.

The majority of known approaches to elimination of impulsive disturbances from archive audio signals are based on adaptive prediction – the autoregressive (AR) or autoregressive moving average (ARMA) model of the analyzed signal is continuously updated and used to predict consecutive signal samples [3]–[10]. Whenever the absolute value of the one-step-ahead prediction error becomes too large, namely when it exceeds a prescribed multiple of its estimated standard deviation, a “detection alarm” is raised, and the predicted sample is scheduled for reconstruction. The test is then extended to multiple-step-ahead prediction errors – detection alarm is terminated when a given number of samples in a row remain sufficiently close to the predicted signal trajectory (or when the length of detection alarm reaches its maximum allowable value). Finally, once the pulse is localized, the corrupted samples are interpolated (using the same signal model which served for detection purposes) based on the uncorrupted neighboring samples.
The paper aims at comparing three classical AR-model based pulse detection schemes with their extended versions obtained by means of combining results of forward time and backward time signal processing – the technique proposed in [11]. It will be shown that such bidirectional processing noticeably improves detection accuracy, allowing one to increase quality of the restored sound.

2. Problem statement

We will assume that the sampled audio signal $y(t)$ has the form

$$y(t) = s(t) + \delta(t)$$

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) discrete time, $s(t)$ denotes the undistorted (clean) audio signal, and $\delta(t)$ is the sequence of noise pulses.

Let $d(t)$ be the pulse location function

$$d(t) = \begin{cases} 1 & \text{if } \delta(t) \neq 0 \\ 0 & \text{if } \delta(t) = 0 \end{cases} .$$

The problem of elimination of impulsive disturbances is usually solved in two steps. First, noise pulses are localized. The resulting estimated pulse location function has the form

$$\hat{d}(t) = \begin{cases} 1 & \text{if the sample is classified as an outlier} \\ 0 & \text{otherwise} \end{cases} .$$

Then, at the second stage of processing, all samples regarded as outliers $Y_d = \{y(t) : \hat{d}(t) = 1\}$ are interpolated based on the approved samples $Y_s = \{y(t) : \hat{d}(t) = 0\}$.

We will assume that the noiseless audio signal $s(t)$ obeys the following autoregressive model of order $r$

$$s(t) = \sum_{i=1}^{r} a_i s(t-i) + n(t)$$

where $a_i, i = 1, \ldots, r$ denote known autoregressive coefficients and $n(t)$ denotes zero-mean white driving noise with variance $\sigma_n^2$. The minimum-variance one-step-ahead prediction of $s(t)$ is given by

$$\hat{s}(t+1) = \sum_{i=1}^{r} a_i s(t-i) + 1$$

and the $k$-step-ahead prediction $\hat{s}(t+k)$ can be obtained by concatenating $k$ one-step-ahead predictions

$$\hat{s}(t+j) = \sum_{i=1}^{r} a_i \hat{s}(t+j-i), \quad j = 1, \ldots, k$$

where $\hat{s}(t+j) = s(t+j)$ for $j \leq 0$.

Note that when the $r$ past signal measurements are not corrupted by noise pulses, which is equivalent to $d(t-r+1) = \ldots = d(t) = 0$, it holds that $\hat{y}(t+j) = \hat{s}(t+j)$, i.e., the predicted values of the signal $y(t)$ can be obtained using the formula analogous to (3)

$$\hat{y}(t+j) = \sum_{i=1}^{r} a_i \hat{y}(t+j-i), \quad j = 1, \ldots, k$$

where $\hat{y}(t+j) = y(t+j)$ for $j \leq 0$.

3. Detection based on open-loop prediction

The detection alarm is started at the instant $t_0 + 1$: $\hat{d}(t_0 + 1)$, if the magnitude of the one-step-ahead prediction error $\varepsilon(t_0 + 1|t_0) = y(t_0 + 1) - \hat{y}(t_0 + 1|t_0)$ exceeds $\mu$ times its estimated standard deviation $\sigma_{\varepsilon}(t_0 + 1|t_0) = \sigma_n$

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

where $\mu$ is the detection threshold multiplier determined experimentally (usually the best results are obtained for $\mu \in [3, 5]$; $\mu = 3$ corresponds to the well-known “3-sigma” rule used for detection of outliers in Gaussian signals).

The detection process is continued for multi-step-ahead predictions, i.e., the absolute values of the one-step-ahead prediction errors $\varepsilon(t_0 + k|t_0) = y(t_0 + k) - \hat{y}(t_0 + k|t_0)$, $k = 2, 3, \ldots$, are checked against the corresponding thresholds $\mu \sigma_{\varepsilon}(t_0 + k|t_0)$. The alarm ends at the instant $t_0 + k_0 + 1$: $\hat{d}(t_0 + k_0 + 1) = 0$, if $r$ consecutive prediction errors are sufficiently small, namely

$$|\varepsilon(t_0 + k_0 + i|t_0)| \leq \mu \sigma_{\varepsilon}(t_0 + k_0 + i|t_0), \quad i = 1, \ldots, r$$

or if the length of the detection alarm $k$ reaches the prescribed value $k_{\text{max}}$.

The final detection decision has the form

$$\hat{d}(t_0 + 1) = \ldots = \hat{d}(t_0 + k_0) = 1$$

$$\hat{d}(t_0 + k_0 + 1) = \ldots = \hat{d}(t_0 + k_0 + r) = 0 .$$

The variance of the multi-step-ahead prediction errors can be evaluated recursively using the following algorithm proposed by Stoica [12]

$$\sigma_{\varepsilon}^2(t_0+k_0|t_0) = \sigma_{\varepsilon}^2(t_0+k-1|t_0) + \sigma_n^2 f_{k-1}$$

$$f_{k-1} = g_k^0$$

$$g_k^i = g_{k-1}^{i+1} + a_{i+1} f_{k-1}$$

$$k = 2, \ldots, k_0$$

with initial conditions: $\sigma_{\varepsilon}^2(t_0+1|t_0) = \sigma_n^2$, $f_0 = 1$ and $g_1^i = a_{i+1}$, $i = 0, \ldots, r-1$. 
The adaptive prediction formula can be obtained by replacing known coefficients of the AR model, appearing in (3)-(6), with their estimates \( \hat{a}_1(t_0), \ldots, \hat{a}_p(t_0) \) and \( \hat{\sigma}_2^2(t_0) \) yielded by the finite-memory signal identification/tracking algorithm, such as the well-known exponentially weighted least squares (EWLS) algorithm, or the least mean square (LMS) algorithm [13], [14]. The order of autoregression can be fixed or chosen adaptively using the generalized Akaike’s criterion [15].

4. Detection based on decision-feedback prediction

The simple detection scheme described above, based on the open-loop multiple-step-ahead signal prediction, can be replaced with a more sophisticated scheme based on decision-feedback prediction. In this case prediction errors and the corresponding standard deviations are evaluated on-line by the Kalman filtering algorithm which takes into account its earlier accept/reject decisions.

4.1. Kalman predictor/filter/smoothere

We will start from solving a simpler problem of recovering an isolated block of \( k_0 \) 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 \( t_0 + k_0 \) (i.e., \( d(t_0 + 1) = \ldots = d(t_0 + k_0) = 1 \)), is preceded and succeeded by undistorted samples (i.e., \( d(t) = 0 \) for \( t < t_0 \) and \( t > t_0 + k_0 \)). We will assume that the location of the sequence of noise pulses is known exactly, i.e., \( \hat{d}(t) \equiv d(t) \). We will also assume that noise pulses \( \delta(t_0 + 1), \ldots, \delta(t_0 + k_0) \) can be modeled as a sequence of mutually uncorrelated Gaussian variables, independent of \( \{n(t)\} \), with known variances

\[
\sigma_\delta^2(t) = \text{var}[\delta(t)], \quad t_0 + 1 \leq t \leq t_0 + k_0.
\]

The solution, based on Kalman smoothing, 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 the Kalman filter/smooother we need a state space equivalent of the input-output description (1)-(2). Let \( q = 2r + k_0 \). Define the \( q \times 1 \) state vector \( x_q(t) = [s^T(t), \ldots, s^T(t - q + 1)]^T \) made up of the \( q \) most recent signal samples. Denote by \( \mathbf{O}_q \) the \( q \times 1 \) null vector, and by \( \mathbf{I}_q \) and \( \mathbf{I}_q - \mathbf{O}_q \) – the \( q \times q \) null and identity matrices, respectively. 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

\[
x_q(t + 1) = \mathbf{A}_q x_q(t) + c_q n(t + 1)
\]

\[
y(t) = c_q^T x_q(t) + \delta(t)
\]

where

\[
\mathbf{A}_q = \begin{bmatrix}
a_1 & a_2 & \ldots & a_r & 0 & \ldots & 0 & 0 \\
1 & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 \\
0 & 1 & \ldots & 0 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\]

is the \( q \times q \) state transition matrix and \( c_q = [1, 0, \ldots, 0]^T \) denotes the \( q \times 1 \) output vector.

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

\[
\hat{x}_q(t | t - 1) = \mathbf{A}_q \hat{x}_q(t - 1 | t - 1)
\]

\[
\mathbf{P}_q(t | t - 1) = \mathbf{A}_q \mathbf{P}_q(t - 1 | t - 1) \mathbf{A}_q^T + \sigma_\delta^2 c_q c_q^T
\]

\[
e(t) = y(t) - c_q^T \hat{x}_q(t | t - 1)
\]

\[
\beta(t) = c_q^T \mathbf{P}_q(t - 1 | t - 1)c_q + \sigma_\delta^2(t)
\]

\[
\mathbf{I}_q(t) = \mathbf{P}_q(t | t - 1) c_q / \beta(t)
\]

\[
\hat{x}_q(t) = \hat{x}_q(t | t - 1) + \mathbf{I}_q(t) e(t)
\]

\[
\mathbf{P}_q(t | t) = \mathbf{P}_q(t | t - 1) - \beta(t) \mathbf{I}_q(t) \mathbf{P}_q(t | t - 1) \mathbf{I}_q(t)^T.
\]

Since we have assumed that \( \delta(t) = 0 \) for \( t \leq t_0 \), the algorithm should be started at the instant \( t_0 + 1 \), with initial conditions \( \hat{x}_q(t_0 | t_0) = [y(t_0), \ldots, y(t_0 - q + 1)]^T \), \( \mathbf{P}_q(t_0 | t_0) = \mathbf{O}_q \), and stopped at the instant \( t_0 + k_0 + r \), after reading \( r \) undisturbed signal samples at the end of the corrupted fragment. The filtered state vector at the termination point \( t_0 + k_0 + r \) has the form

\[
\hat{x}_q(t_0 + k_0 + r | t_0 + k_0 + r) = [y(t_0 + k_0 + r), \ldots, y(t_0 + k_0 + 1), \hat{s}(t_0 + k_0), \ldots, \hat{s}(t_0 + 1), y(t_0), \ldots, y(t_0 - r + 1)]^T
\]

where \( \hat{s}(t_0 + 1), \ldots, \hat{s}(t_0 + k_0) \) 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 \( t_0 + k_0 + r + 1, t_0 + k_0 + r + 2, \ldots \), there is no point in continuing operation of the Kalman filter after reaching the point \( t_0 + k_0 + r \).

4.2. Detection of noise pulses

Similarly as in the open-loop case, the outlier detection alarm is triggered at the instant \( t_0 + 1 \) if the condition (5) is met. The test is then extended to consecutive prediction errors yielded by the Kalman filtering algorithm (10) initialized at the instant \( t_0 \). At each time instant \( t > t_0 \) the following variance scheduling rule is used

\[
\sigma_\delta^2(t) = \begin{cases}
0 & \text{if } \hat{d}_q(t) = 0 \\
\infty & \text{if } \hat{d}_q(t) = 1
\end{cases}
\]

\[
\text{(11)}
\]
where \( \hat{d}_0(t) \) denotes the preliminary detection decision

\[
\hat{d}_0(t) = \begin{cases} 
0 & \text{if } |e(t)| \leq \mu \sigma_e(t) \\
1 & \text{if } |e(t)| > \mu \sigma_e(t) 
\end{cases}
\]

(12)

and \( \sigma^2_e(t) = \text{var}[e(t)] = c^2_{\theta} P_{\theta}(t|t-1)c_0 \).

Note that when \( \sigma^2_e(t) \to \infty \) the sample \( y(t) \) is regarded as corrupted with infinite-variance noise and - as such - rejected and interpolated based on the neighboring samples; when \( \sigma^2_e(t) = 0 \) the sample is provisionally accepted, i.e., regarded as disturbance-free.

It is straightforward to check that under (11) the corresponding values of \( 1/\beta(t) \) are given by

\[
\frac{1}{\beta(t)} = \begin{cases} 
1/\sigma^2_e(t) & \text{if } \hat{d}_0(t) = 0 \\
0 & \text{if } \hat{d}_0(t) = 1 .
\end{cases}
\]

(13)

The detection alarm is terminated at the instant \( t_0 + k_0 + 1: \hat{d}(t_0 + k_0 + 1) = 0 \), if \( r \) consecutive prediction errors are sufficiently small

\[
|e(t_0 + k_0 + i)| \leq \mu \sigma_e(t_0 + k_0 + i), \quad i = 1, \ldots, r.
\]

(14)

To avoid “accidental acceptances” of samples located in the middle of long-lasting artifacts, even if some samples \( y(t) \) from the interval \( [t_0 + 2, t_0 + k_0 - 1] \) were provisionally accepted (\( \hat{d}_0(t) = 0 \)), the final outlier detection decision has the form identical with (7), which guarantees that detection alarms always form solid blocks of “ones” preceded and succeeded by at least \( r \) “zeros”.

Remark

It should be noted that the stopping rule (14) is identical with (6) if and only if it holds that

\[
\hat{d}_0(t_0 + 1) = \ldots = \hat{d}_0(t_0 + k_0) = 1
\]

(15)
i.e., when all \( k_0 \) samples from the interval \( [t_0 + 1, t_0 + k_0] \) are regarded as trustworthy. In a case like this prediction errors \( e(t) \) yielded by the Kalman algorithm (10) are identical with the multi-step-ahead prediction errors that can be obtained using the open-loop formula (4), i.e.,

\[
e(t_0 + k) = \varepsilon(t_0 + k|t_0), \quad 1 \leq k \leq k_0.
\]

However, if for at least one \( t \in [t_0 + 1, t_0 + k_0] \), the prediction error \( e(t) \) takes the value smaller or equal to the detection threshold \( \mu \sigma_e(t) \) (which results in setting \( \hat{d}_0(t) = 0 \)), the corresponding signal predictions, given by \( c^2_{\theta} \tilde{X}_0(t|t-1) \), depend not only on the samples collected prior to the instant \( t_0 + 1 \), but also on the samples that were provisionally accepted afterwards – such predictions will be further called decision-feedback since they depend on the variance scheduling decisions made before \( t \). When the condition (15) is not met, the decision-feedback predictions differ from the open-loop predictions.

4.3. Variable-order algorithm

Taking advantage of the special structure of the matrices/vectors \( A_q \), \( c_q \) and \( P_q(t_0|t_0) \), 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 (\( q = q_{\text{max}} = 2r + k_{\text{max}} \)) version.

Let \( \theta_r = [\theta_1, \ldots, \theta_r]^T \) and denote by \( \theta_q = [\theta_q^T, 0^T_{q-r}]^T \), \( q > r \), the augmented parameter vector. Denote by \( X^{(1)} \) the vector made up of the first column of the matrix \( X \).

The variable-order algorithm which combines (10) with (12)-(13) can be summarized as follows:

**Initialization**

\[
\hat{X}_r(t_0|t_0) = [y^T(t_0), \ldots, y^T(t_0 - r + 1)]^T \\
P_r(t_0|t_0) = O_r
\]

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

\[
\hat{y}(t|t-1) = \theta_r^T \hat{X}_r(t_r+t_0-1|t-1) - 1 \quad t \quad e(t) = y(t) - \hat{y}(t|t-1) \\
\hat{X}_{r+t+t_0-1}(t|t-1) = \hat{X}_{r+t+t_0-1}(t|t-1) \\
\hat{h}_{r+t+t_0-1}(t|t-1) = \hat{P}_{r+t+t_0-1}(t|t-1) + \sigma^2_{t}(t) \\
P_{r+t+t_0-1}(t|t-1) = P_{r+t+t_0-1}(t|t-1) + \sigma^2_{t}(t)
\]

**Outlier detection step**

\[
\hat{d}_0(t) = \begin{cases} 
0 & \text{if } |e(t)| \leq \mu \sigma_e(t) \\
1 & \text{if } |e(t)| > \mu \sigma_e(t) 
\end{cases}
\]

(16)

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

**Case 1**: if \( \hat{d}_0(t) = 0 \) or \( t \geq t_0 + k_{\text{max}} \) then

\[
\hat{X}_{r+t+t_0-1}(t|t) = \hat{X}_{r+t+t_0-1}(t|t-1) + \hat{h}_{r+t+t_0-1}(t|t) \\
P_{r+t+t_0-1}(t|t) = P_{r+t+t_0-1}(t|t-1)
\]

**Case 2**: if \( \hat{d}_0(t) = 1 \) then

\[
\hat{X}_{r+t+t_0-1}(t|t) = \hat{X}_{r+t+t_0-1}(t|t-1) \\
P_{r+t+t_0-1}(t|t) = P_{r+t+t_0-1}(t|t-1)
\]

The final detection decision has the form (7).

Similarly as in the open-loop prediction approach, the signal-adaptive version of the algorithm listed above can be obtained by means of replacing the quantities \( \theta_r \) and \( \sigma^2_{t}(t) \) with their estimates \( \hat{\theta}_r(t_0) \) and \( \hat{\sigma}^2_{t}(t_0) \), respectively.
5. Alarm extension technique

As argued in [11], typical geometry of local damages of the recording medium (e.g., groove damages) results in pulses that usually start and end in a "gentle" way – abrupt changes, which constitute the main "body" of a click, are preceded and succeeded by much smaller but systematic changes resulting in smooth pre-click and post-click signal distortions. This effect becomes more pronounced as the sampling rate grows. In the presence of soft edges, detection alarms raised by prediction-based algorithms (both open-loop and decision-feedback) are seldom triggered at the very beginning of noise pulses, which may result in small but audible distortions of the reconstructed audio material. The effect described above can be alleviated by decreasing the detection multiplier \( \mu \), i.e., by making the detector more sensitive to unpredictable signal changes. This, however, may dramatically increase the number and length of detection alarms, causing the overall degradation of the results.

A solution, proposed in [11] and recommended also here, is to shift back the front edge of detection alarm (once triggered) by a small, fixed number of samples \( \Delta \), i.e., from the instant \( t_0 + 1 \) to \( t_0 + 1 - \Delta \). For 44.1 and 48 kHz recordings \( \Delta = 2 \) is usually a good choice.

6. Detection based on double threshold approach

An interesting approach that allows one to avoid the oversensitivity problem mentioned in Section 5, based on double thresholding, was described in [16]. Unlike sequential prediction-based methods presented in Sections 3 and 4, the double threshold approach incorporates block processing. The signal is divided into blocks (possibly overlapping), each of which is analyzed separately.

Consider a block consisting of \( N \) samples \( Y(N) = \{y(1), \ldots, y(N)\} \). The detection procedure is two-step.

The aim of the first step is to find the abnormally large values of residual errors \( \eta(t) \). In most cases large residual errors can be attributed to the presence of noise pulses. First, based on \( Y(N) \) the coefficients of an autoregressive model of the signal are estimated \( \hat{\theta}(N) = [\hat{a}_1(N), \ldots, \hat{a}_r(N)]^T \), and the sequence of residual errors is evaluated

$$\eta(t) = y(t) - \sum_{i=1}^{r} \hat{a}_i(N)y(t-i), \quad t = 1, \ldots, N$$

where the initial conditions \( y(1-r), \ldots, y(0) \) are drawn from the previous block. Then, the preliminary pulse detection is carried out using the following decision rule

$$\hat{d}_0(t) = \begin{cases} 0 & \text{if } |\eta(t)| \leq \mu_0 \hat{\sigma}_n(N) \\ 1 & \text{if } |\eta(t)| > \mu_0 \hat{\sigma}_n(N) \end{cases}$$

where \( \mu_0 \) (detection multiplier) is set to a relatively large value (e.g., \( \mu_0 = 5 \)), and \( \hat{\sigma}_n(N) \) denotes the robust local estimate of the standard deviation of \( n(t) \) such as

$$\hat{\sigma}_n^2(N) = 1.4 \med \{\eta^2(t), t = 1, \ldots, N\}.$$ (18)

The function \( \med(\cdot) \) denotes median – the central value of the ordered sequence of numbers.

The purpose of the second step is to precisely localize the beginning and the end of each detection alarm. Suppose that \( m \) such alarms were raised, i.e.,

$$\hat{d}_0(t) = \begin{cases} 1 & \text{for } t \in \cup_{i=1}^{m} D_i^0 \\ 0 & \text{elsewhere} \end{cases}$$

where

$$D_i^0 = [t_i^0, \bar{t}_i], \quad t_i^0 \leq \bar{t}_i, \quad i = 1, \ldots, m.$$ Each preliminary detection alarm \( D_i^0 \) is checked for a possible extension. The extended (widened) alarm has the form

$$D_i = [\ell_i, \bar{\ell}_i], \quad \ell_i \leq t_i^0, \quad \bar{\ell}_i \geq \bar{t}_i, \quad i = 1, \ldots, m$$

where \( \ell_i/\bar{\ell}_i \) denote the minimum/maximum values such that

$$|\eta(t)| > \mu \hat{\sigma}_n(N) \quad \forall t \in D_i$$

and \( \mu = \mu_0/2 \) (location multiplier) takes a relatively small value. Low decision threshold makes the outlier detector less tolerant to departures of the signal from its expected path, allowing one to localize the beginning and end points of each detection alarm in a more precise manner – see Fig 1.

To eliminate detection errors caused by “destructive interference” (occurring when some residual errors take values close to zero in the middle of long-lasting noise pulses), the authors of [16] proposed refinement of the alarm delimiting rule (19). According to the new rule, the condition (19) may be violated as long as this happens for not more than \( l \) consecutive residual errors, where \( l \) denotes a small number (e.g. \( l = 3 \)).
7. Limitations of the existing approaches

The common drawback of all three approaches discussed in Sections 3, 4, and 6, is the lack of precision in determining the end points of detection alarms. As the prediction horizon grows, the variance of the multi-step-ahead prediction errors tends to the estimated signal variance, which means that the outlier detector becomes increasingly tolerant to discrepancies between the observed and predicted signal values. A direct consequence of this is that the approach based on open-loop prediction shows tendency to raise too short detection alarms, i.e., alarms that end before the entire pulse waveform is complete. The results improve if the decision-feedback approach is used, since samples provisionally accepted in the middle of detection alarms may significantly decrease the prediction error variance, which increases sensitivity of the outlier detector. Therefore, unless the condition (15) is met, detection alarms raised by the scheme based on decision-feedback predictions are usually longer than those yielded by the open-loop scheme.

Unlike the prediction-based approaches, the double threshold approach shows tendency to produce overly long detection alarms. It is not difficult to explain this effect. Suppose that the pulse waveform starts at the instant $t_0 + 1$ and ends at the instant $t_0 + k_0 + 1$. Note that, even though the sample $y(t_0 + k_0 + 1)$ is outlier-free, the corresponding value of the residual error $\eta(t_0 + k_0 + 1)$ usually remains large as it is evaluated based on $r$ preceding signal samples, at least some of which are contaminated by outliers. It is not until the sample $y(t_0 + k_0 + r + 1)$ is reached, that residual errors are entirely unaffected by the detected noise pulse. As a result, when the adopted order of autoregression is large ($r \geq 20$ is a recommended choice under 44.1 and 48 kHz sampling), the corresponding detection alarms are usually much longer than the “ground truth” ones.

Another important limitation of the existing schemes is due to the fact that they all are based on the results of unidirectional signal analysis. A more detailed comment on this feature will be presented in the next section.

8. Bidirectional processing

So far we have assumed that the archive audio signal is analyzed sequentially, forward in time. In such a case a sample is regarded as an outlier if it is “inconsistent” with the signal past, which is indicated by excessive values of prediction or residual errors. When signal characteristics change abruptly, e.g., when an entirely new sound starts to build up, all causal detection schemes are prone to generate false detection alarms, calling in question uncorrupted signal samples simply because they do not match the signal past. Since such samples are consistent with the signal “future”, rather than its “past”, the number of false alarms can be significantly reduced if results of forward-time detection are combined with the analogous results of backward-time detection. The latter can be obtained by means of processing audio signal backward in time (provided, of course, that the entire recording is available). In addition to reducing the number and length of false alarms, bidirectional processing allows one to carve detection alarms more carefully (smaller number of overlooked noise pulses, better front/end matching of noise pulses).

The set of local, case-dependent fusion rules that can be used to combine forward and backward detection alarms, denoted respectively by $\hat{D}_{fb}^f(t)$ and $\hat{D}_{fb}^b(t)$, was proposed and experimentally verified in [11]. Alarms are combined in a way that depends on their mutual configuration called a detection pattern. For example, when forward and backward detection alarms form blocks that at least partially overlap

$$\hat{D}_{fb}^f(t) = 1 \text{ for } t \in [t_{fb}^f, \overline{t_{fb}^f}] = D_{ij}^f,$$

$$\hat{D}_{fb}^b(t) = 1 \text{ for } t \in [t_{fb}^b, \overline{t_{fb}^b}] = D_{ij}^b,$$

$$D_{ij}^f \cap D_{ij}^b \neq \emptyset$$

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

$$\hat{D}_{fb}^b(t) = 1 \text{ for } t \in [t_{fb}^b, \overline{t_{fb}^b}] = D_{ij}^b.$$

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

9. Experimental results

The efficiency and accuracy of different pulse detection techniques were evaluated using artificially corrupted audio files. Our repository of clicks was made up of 500 click waveforms extracted from silent parts of old gramophone recordings (under 48 kHz sampling) forming the set $\mathcal{P}$. The audio test base consisted of 5 clean recordings contaminated with click waveforms randomly drawn from the set $\mathcal{P}$. Clean audio recordings contained from 23 to 29 seconds of classical music (Bach, Mozart, Vivaldi, Smetana) sampled at the rate of 48 kHz. The audio material was 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. Clicks were picked at random from the repository and added every 300 signal samples. Such a regular spacing between consecutive
noise pulses was a deliberate choice as regularly occurring signal distortions imperfections are more audible than those appearing in irregular time constellations.

Performance evaluation was made for 6 approaches: the open-loop prediction based approach (A), the decision-feedback prediction based approach (B), the double-threshold based approach (C), and the bidirectional versions of the approaches listed above, denoted by $A^*$, $B^*$, and $C^*$, respectively. The decision-feedback detection algorithm was made up of the EWLS parameter tracker and the Kalman filter based signal predictor.

All compared detection/reconstruction algorithms incorporated AR models of order $r = 20$. In approaches A and B signal identification was carried out using the EWLS algorithm equipped with a forgetting factor $\lambda = 0.995$. The detection multiplier was set to $\mu = 4.5$ and the alarm extension parameter was set to $\Delta = 2$. For the double-threshold approach, the default values of internal parameters recommended in [16] were adopted. In all cases reconstruction of samples called in question by the outlier detector was carried out using the AR-model based interpolator incorporating $r$ samples preceding and $r$ samples succeeding the block of corrupted samples – see [17] for more details.

Prior to comparing detection efficiency of different approaches, the adaptive AR-model based interpolation algorithm, supported with information about the exact location of inserted clicks $[d(t) \equiv \hat{d}(t)]$ was run on each of 5 test recordings and the results were evaluated via listening tests. The purpose of this “ground truth” experiment was to check how much signal interpolation alone (carried out in the presence of perfect detection of inserted noise pulses) affects the final reconstruction results. Since in all 5 cases listening tests reported no audible difference between the original audio material and the reconstructed one, it was clear that all audible distortions (if any) observed later, when adaptive interpolation was combined with adaptive detection, must have been caused by detection errors, such as missing detections, inaccurate detections, and false detections.

To evaluate performance of different detection/reconstruction algorithms, we used the perceptual evaluation of audio quality (PEAQ) tool [18], [19]. PEAQ scores take negative values that range from -4 (very annoying distortions) to 0 (imperceptible distortions). The PEAQ standard uses a number of psychoacoustical evaluation techniques which are combined to give a measure of the quality difference between the original audio signal and its processed version. Even though it was introduced as an objective method to measure the quality of perceptual coders, without any reference to audio restoration, we have found it useful for our purposes as it gives scores that are well correlated with the results of time consuming listening tests. 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.

In addition to PEAQ-based evaluation, three other objective measures of fit were used to quantify the obtained results — the pulse energy coverage statistic, the degree of underdetection, and the degree of overdetection. All measures, defined below, are indirect as they quantify accurateness of the detection process.

The pulse energy coverage statistic, proposed in [11], measures the percentage of the overall energy of noise pulses captured by the detector

$$c = \frac{\sum_{t \in T^d} \delta^2(t)}{\sum_{t \in T_o} \delta^2(t)} \%$$

where $T^d = \{ t : \hat{d}(t) = 1 \land d(t) = 1 \}$, $T_o = \{ t : d(t) = 1 \}$.

Denote by $T_o = \{ t : \hat{d}(t) = 0 \land d(t) = 1 \}$ and $T_d = \{ t : d(t) = 1 \land d(t) = 0 \}$ the sets indicating positions of corrupted samples that were overlooked by the outlier detector (true negatives) and positions of uncorrupted samples that were questioned by the outlier detector (false positives), respectively. The degree of underdetection will be defined as

$$u = \frac{|T_o|}{|T_d|}$$

where $|S|$ denotes cardinality (number of elements) of the set $S$. Similarly, the degree of overdetection can be defined as

$$o = \frac{|T_d|}{|T_o|}.$$
Table I. Direct (PEAQ) and indirect (pulse energy coverage) objective performance measures for the compared detection schemes: the unidirectional open-loop prediction based approach (A), the unidirectional decision-feedback prediction based approach (B), the double threshold based approach (C), the bidirectional open-loop prediction based approach (A*), the bidirectional decision-feedback prediction based approach (B*), and the bidirectional double threshold based approach (C*). Additionally, the first table shows results obtained for the corrupted audio files (REF), and ground truth results (GT), obtained when the signal is reconstructed under the perfect knowledge of pulse location. The best scores within the groups of unidirectional methods (A, B, C) and bidirectional methods (A*, B*, C*) are shown in boldface.

Interpretation of PEAQ scores: 0 = imperceptible (signal distortions), −1 = perceptible but not annoying, −2 = slightly annoying, −3 = annoying, −4 = very annoying.

<table>
<thead>
<tr>
<th>Audio file</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PEAQ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REF</td>
<td>−3.68</td>
<td>−3.69</td>
<td>−3.70</td>
<td>−3.59</td>
<td>−3.77</td>
</tr>
<tr>
<td>A</td>
<td>−3.18</td>
<td>−3.26</td>
<td>−3.68</td>
<td>−3.37</td>
<td>−3.68</td>
</tr>
<tr>
<td>B</td>
<td>−0.21</td>
<td>−0.18</td>
<td>−0.59</td>
<td>−0.29</td>
<td>−0.33</td>
</tr>
<tr>
<td>C</td>
<td>−0.32</td>
<td>−0.28</td>
<td>−1.04</td>
<td>−0.55</td>
<td>−0.98</td>
</tr>
<tr>
<td>A*</td>
<td>−0.20</td>
<td>−0.15</td>
<td>−0.45</td>
<td>−0.21</td>
<td>−0.27</td>
</tr>
<tr>
<td>B*</td>
<td>−0.15</td>
<td>−0.12</td>
<td>−0.41</td>
<td>−0.20</td>
<td>−0.25</td>
</tr>
<tr>
<td>C*</td>
<td>−0.25</td>
<td>−0.23</td>
<td>−0.62</td>
<td>−0.35</td>
<td>−0.48</td>
</tr>
<tr>
<td>GT</td>
<td>−0.03</td>
<td>−0.07</td>
<td>−0.08</td>
<td>−0.01</td>
<td>−0.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pulse energy coverage [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>A*</td>
</tr>
<tr>
<td>B*</td>
</tr>
<tr>
<td>C*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Degree of underdetection</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>A*</td>
</tr>
<tr>
<td>B*</td>
</tr>
<tr>
<td>C*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Degree of overdetection</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>A*</td>
</tr>
<tr>
<td>B*</td>
</tr>
<tr>
<td>C*</td>
</tr>
</tbody>
</table>

In the group of bidirectional approaches (A*, B*, C*) the best PEAQ scores were again obtained for the decision-feedback scheme (B*), and the worst scores – for the double threshold based scheme (C*). Similarly as in the unidirectional case, the relatively poor performance of the worst scheme (this time C*) was caused by poor placement of detection alarms – very low degree of overdetection combined with high degree of underdetection indicate that most of detection alarms were overly short.

In all cases the bidirectional versions of the analyzed approaches yield better results than their unidirectional versions: for the approaches B and C the average rate of improvement is equal to 32%; for the approach A the improvement is very substantial and exceeds 90%.

Finally note that when the competing approaches are ranked based on comparison of their pulse energy coverage statistics, the conclusions are exactly the same as those reached by comparing the corresponding PEAQ scores.

Listening tests, performed on real archive audio files, support all findings summarized above.

10. CONCLUSIONS

The problem of elimination of impulsive disturbances from archive audio recordings was considered and three noise pulse detection schemes were examined: the approach based on open-loop multi-step-ahead signal prediction, the approach based on decision-feedback signal prediction (governed by the variable-order Kalman filter), and the double threshold approach, based on analysis of residual errors. The efficiency of the compared schemes was checked on a set of artificially corrupted audio files (clean audio signals contaminated with click waveforms extracted from old gramophone recordings) using the perceptual evaluation of audio quality (PEAQ) tool. For all test recordings the best PEAQ scores were obtained using the decision-feedback detection algorithm.

The analyzed detection schemes were next compared with their extended versions obtained by means of combining results of forward time and backward time signal processing. It was shown that, in all cases considered, such bidirectional technique noticeably improves detection accuracy, which results in better quality of the restored sound. Again, the best results can be obtained using the decision-feedback scheme.

Acknowledgement

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

References


