LOCALIZATION OF IMPULSIVE DISTURBANCES IN ARCHIVE AUDIO SIGNALS USING PREDICTIVE MATCHED FILTERING

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 problem of elimination of impulsive disturbances from archive audio signals is considered and its new solution, called predictive matched filtering, is proposed. The new approach is based on the observation that a large percentage of noise pulses corrupting archive audio recordings have highly repetitive shapes that match several typical "patterns", called click templates. To localize noise pulses, click templates can be correlated with the sequence of multi-step-ahead prediction errors yielded by the model-based signal predictor. It is shown that predictive matched filtering is an efficient and computationally affordable disturbance localization technique – when combined with the classical detection method based on autoregressive modeling, it can significantly improve restoration results.

Index Terms— Restoration of audio signals, outlier detection and elimination, adaptive signal processing.

1. INTRODUCTION

Archived audio recordings are often degraded by impulsive disturbances [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), specs 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). Elimination of noise pulses from archive audio documents is an important element of saving our cultural heritage.

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 sampled audio 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 undistorted (clean) audio signal, and $\delta(t)$ is the sequence of noise pulses.

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

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

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

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

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

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]–[9]. Whenever the absolute values of the multi-step-ahead prediction errors become too large, namely when they exceed a prescribed multiple of their estimated standard deviation, a "detection alarm" is raised, and the predicted samples are scheduled for reconstruction. Recently some nonlinear restoration techniques were also proposed [10].

The classical approach, mentioned above, is a general purpose outlier elimination scheme which does not rely on any information about the size and shape of noise pulses – even if such a prior knowledge is available. To the best of our knowledge, apart from [4], which focuses on very long disturbances such as record scratches, the only approach proposed so far, which incorporates prior knowledge about noise pulses into pulse detection/elimination procedure, is that described in the recent paper of Ávila and Biscainho [11]. The Bayesian pattern matching procedure proposed there is based on the idea of Gibbs sampling. It results in a numerical procedure which – unlike the procedure described below – is computationally very demanding.

2. CREATING CLICK TEMPLATES

While some noise pulses encountered in archive audio recordings have unique (and sometimes rather complicated) shapes, the majority of them form repeatable patterns which can be grouped in a relatively small number of classes represented by click templates. Typical shapes and duration of noise pulses may strongly depend on the recording medium (shellac, vinyl, magnetic tape), the way it was handled in the past (storage conditions, degree of wear), played back (pre-amplifier mode, turntable speed, type of stylus or tape deck), and digitized (sampling rate). Hence, the important feature of the proposed approach is its source adaptivity.

Exemplary noise pulses can be extracted from the silent parts of archive recordings preceding and/or succeeding the actual soundtracks. Extraction can be performed using any general purpose outlier detection scheme, e.g. by means of adaptive signal thresholding based on the 3-sigma rule.

Similar waveforms will be grouped, normalized, time-aligned and averaged, forming click templates. As a tool for shape similarity

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

analysis, we will use the quantity known in statistics as correlation (normalized covariance) coefficient.

Denote by $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_N\}$ the set consisting of N extracted noise pulses where $\mathcal{P}_i = \{p_i(1), \ldots, p_i(k_i)\}$ is the sequence of samples, of length k_i, forming the i-th pulse. Denote by $\widetilde{\mathcal{P}}_i = \{\widetilde{p}_i(1), \ldots, \widetilde{p}_i(k_i)\}$ the sequence of normalized pulse samples where

$$\widetilde{p}_i(k) = \frac{p_i(k) - \bar{p}_i}{\sqrt{\sum_{k=1}^{k_i}[p_i(k) - \bar{p}_i]^2}}\ , \quad \bar{p}_i = \frac{1}{k_i}\sum_{k=1}^{k_i} p_i(k)\ . \tag{2}$$

When comparing two waveforms, say $\mathcal{P}_i$ and $\mathcal{P}_j$, one should account for their (possibly) different length and lack of alignment. To find the best alignment, we will compute correlation-based similarity scores between $\widetilde{\mathcal{P}}_i$ and the sequence $\widetilde{\mathcal{P}}_j$ shifted by τ samples. Assuming that samples preceding $\widetilde{p}_j(1)$ and succeeding $\widetilde{p}_j(k_j)$ have zero values, the similarity score for the time shift τ can be expressed in the form

$$\rho_{ij}(\tau) = \sum_{k=\max(1,1-\tau)}^{\min(k_i,k_j-\tau)} \widetilde{p}_i(k)\widetilde{p}_j(k+\tau) \tag{3}$$

where $\tau \in \mathcal{T}_{ij} = [1-k_i, k_j - 1]$. Note that the summation range in (3) accounts for differences in the length of the compared sequences. The entire set of correlation coefficients $\rho_{ij}(\tau), \tau \in \mathcal{T}_{ij}$ can be efficiently computed using the FFT-based convolution algorithm.

Denote by $\tau_{ij} = \arg\max_{\tau\in\mathcal{T}_{ij}} \rho_{ij}(\tau)$ the time shift maximizing the similarity score, i.e., the one that guarantees the best alignment of $\widetilde{\mathcal{P}}_i$ and $\widetilde{\mathcal{P}}_j$. To measure the degree of similarity between $\mathcal{P}_i$ and $\mathcal{P}_j$, we will use the maximum correlation coefficient: $r_{ij} = \max_{\tau\in\mathcal{T}_{ij}} \rho_{ij}(\tau) = \rho_{ij}(\tau_{ij})$.

Based on the set of correlation coefficients $\{r_{ij}, i, j = 1, \ldots, N\}$, one can build an undirected similarity graph G showing an internal similarity structure of the analyzed set of noise pulse extracts. This graph has N vertices corresponding to different click waveforms $\mathcal{P}_1, \ldots, \mathcal{P}_N$. If the degree of similarity between $\mathcal{P}_i$ and $\mathcal{P}_j$ is sufficiently high, namely, if $r_{ij} \geq \gamma$, where γ is a threshold close to 1, e.g. $\gamma = 0.95$, the vertices associated with $\mathcal{P}_i$ and $\mathcal{P}_j$ $(i \neq j)$ are connected by an edge. Hence, the adjacency matrix of G has the form

$$\mathbf{L} = [l_{ij}]_{N\times N}, \;\; l_{ij} = \begin{cases} 1 & \text{if } r_{ij} \geq \gamma \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Click templates can be obtained by averaging click waveforms corresponding to maximum cliques of G [13],[14],[15], i.e., its maximum complete subgraphs [every two vertices of a complete (sub)graph must be connected by an edge; the maximum subgraph is the one with the largest number of vertices]. The proposed procedure is recursive and can be summarized as follows:

Initialize: $\mathbf{L}_1 \leftarrow \mathbf{L}, G_1 \leftarrow G, i \leftarrow 1$.

Step 1: Search for the maximum clique Q_i of the graph G_i defined by $\mathbf{L}_i$. If there are several maximum cliques with the same number of vertices n_i, choose the one for which the sum of similarity scores r_{ij} takes the largest value (summation being carried over all edges of Q_i). Alternatively, use a computationally more involved algorithm for finding the weighted maximum clique. If the size of the clique Q_i is sufficiently large, e.g. if $n_i \geq 10$, continue to **Step 2** – otherwise **Stop**.

Step 2: Remove from G_i all vertices and edges of Q_i, forming a new graph G_{i+1} with adjacency matrix $\mathbf{L}_{i+1}$ ($\mathbf{L}_{i+1}$ can be obtained by zeroing the corresponding rows and columns of $\mathbf{L}_i$). Set $i \leftarrow i+1$ and return to **Step 1**.

Once all cliques of sufficient size are found, their "centers" are localized. Denote by $\mathcal{S}_i = \{\widetilde{\mathcal{P}}_j, j \in J_i\}$ the set of normalized pulse waveforms associated with the clique Q_i (J_i is the set indicating which vertices of G belong to Q_i). The central element of $\mathcal{S}_i$, denoted by $\widetilde{\mathcal{P}}_{j_i}$, is the one for which the sum of outgoing edge weights (similarity scores) is maximized

$$j_i = \arg\max_{j\in J_i} \sum_{\substack{l\in J_i \\ l\neq j}} r_{jl}\ .$$

Such element can be interpreted as the one that is "most similar" to the remaining elements of $\mathcal{S}_i$.

All waveforms grouped in $\mathcal{S}_i$ are extended with zeros on both sides, aligned with respect to the central waveform $\widetilde{\mathcal{P}}_{j_i}$, and averaged. Note that the optimal alignment shifts $\tau_{j_i l}, l \in J_i$ were already computed at the pre-processing stage. Since averaging shows tendency to create long tails (small but non-zero values preceding and succeeding the main pulse activity), and since such tails have a marginal impact on the subsequent shape similarity analysis, click templates are obtained by trimming the averaged waveforms, namely by removing from their beginning and end all samples with absolute values smaller than 5% of the peak value. When the length of a noise pulse is too short, shape matching becomes an ill-posed problem. For this reason templates covering less than 4 samples are eliminated.

3. SELECTION OF FEASIBLE CLICK TEMPLATES

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

$$s(t) = \sum_{i=1}^{r} a_i s(t-i) + n(t) \tag{4}$$

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

$$\widehat{s}(t+j|t) = \sum_{i=1}^{r} a_i \widehat{s}(t+j-i|t), \quad j = 1, \ldots, q \tag{5}$$

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

The adaptive prediction formula can be obtained by replacing known coefficients of the AR model, appearing in (5), with their estimates $\widehat{a}_1(t), \ldots, \widehat{a}_r(t)$ 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 [16], [17]. The order of autoregression can be fixed or chosen adaptively using the generalized Akaike's criterion [18].

The preliminary detection procedure is started each time the outlier alarm is raised, i.e., when the magnitude of the one-step-ahead prediction error $\varepsilon(t+1|t) = y(t+1) - \widehat{s}(t+1|t)$ exceeds μ times its estimated standard deviation $\widehat{\sigma}_\varepsilon(t+1|t) = \widehat{\sigma}_n(t)$

$$|\varepsilon(t+1|t)| > \mu\widehat{\sigma}_\varepsilon(t+1|t) \tag{6}$$

where μ is the detection threshold multiplier determined experimentally (usually the best results are obtained for $\mu \in [3, 5]$; $\mu = 3$ corresponds to the so-called "3-sigma" rule, well-known in statistics).

Denote by $\mathcal{C}_i = \{\widetilde{c}_i(1), \ldots, \widetilde{c}_i(m_i)\}, \quad 1 \leq i \leq L$ the i-th template (the average, normalized click waveform) and by $M = \max_{1\leq i\leq L} m_i$ – the length of the longest template. We will check the sequence of prediction errors $\{\varepsilon(t+1|t), \ldots, \varepsilon(t+q|t)\}$, $q > M$, for the presence of different click templates by evaluating the corresponding similarity scores

$$g_i(\tau) = \sum_{k=1}^{m_i} \widetilde{c}_i(k)\widetilde{\varepsilon}(t+k+\tau|t), \quad \tau = 0, \ldots, \tau_{\max} \qquad (7)$$

where τ denotes the alignment shift, $\tau_{\max}$ obeys $\tau_{\max} \leq q - M$, and $\{\widetilde{\varepsilon}(t+1+\tau|t), \ldots, \widetilde{\varepsilon}(t+m_i+\tau|t)\}$ is the sequence of normalized prediction errors [note that normalization, governed by (2), must be performed independently for each value of τ]. Time alignment is necessary to account for uncertainties embedded in triggering the primary detection alarm (determining the moment at which predictive analysis should start), and in knowing the exact width of detected pulses. When the shape similarity test is run only for $\tau = 0$ the results deteriorate.

If, for any value of $\tau \in [0, \tau_{\max}]$ it holds that $g_i(\tau) \geq \gamma_0$, where $\gamma_0 < \gamma$ is the similarity threshold (to account for the bias introduced by prediction errors, γ_0 is set to a smaller value than γ), the template $\mathcal{C}_i$ is regarded as feasible and scheduled for final verification. At any time, more than one feasible template may be found.

4. FINAL DETECTION OF TYPICAL NOISE PATTERNS

The accuracy of signal predictions $\widehat{s}(t+j|t)$ usually decreases with growing prediction horizon j and this fact should be taken account of when selecting and fitting the most appropriate (the most probable) click template. Below we will describe a procedure that fulfills this requirement, further referred to as predictive matched filtering.

Denote by $\mathbf{y}(t+q) = [y(t+q), \ldots, y(t+1)]^{\mathrm{T}}$, $\mathbf{s}(t+q) = [s(t+q), \ldots, s(t+1)]^{\mathrm{T}}$, $\widehat{\mathbf{s}}(t+q) = [\widehat{s}(t+q), \ldots, \widehat{s}(t+1)]^{\mathrm{T}}$, $\boldsymbol{\delta}(t+q) = [\delta(t+q), \ldots, \delta(t+1)]^{\mathrm{T}}$ the vectors of the corrupted audio, clean audio, predicted audio and noise pulse samples, respectively. According to (1), it holds that

$$\mathbf{y}(t+q) = \mathbf{s}(t+q) + \boldsymbol{\delta}(t+q) \ . \qquad (8)$$

Suppose that the noise pulse coincides with the appropriately scaled, time shifted and bias-corrected i-th click template $\mathcal{C}_i$, i.e., that the disturbance vector $\boldsymbol{\delta}(t+q)$ can be written down in the form

$$\boldsymbol{\delta}(t+q) = \alpha_i \mathbf{r}_i(\tau_i) + \beta_i \mathbf{h}_i(\tau_i) \qquad (9)$$

where α_i, β_i and τ_i denote the scale, bias and location coefficients, respectively, and: $\mathbf{r}_i(\tau_i) = [\mathbf{0}^{\mathrm{T}}_{q-m_i-\tau_i}, \mathbf{c}_i^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}}_{\tau_i}]^{\mathrm{T}}$ $\mathbf{h}_i(\tau_i) = [\mathbf{0}^{\mathrm{T}}_{q-m_i-\tau_i}, \mathbf{1}^{\mathrm{T}}_{m_i}, \mathbf{0}^{\mathrm{T}}_{\tau_i}]^{\mathrm{T}}$, $\mathbf{c}_i = [\widetilde{c}_i^{\mathrm{T}}(m_i), \ldots, \widetilde{c}_i^{\mathrm{T}}(1)]^{\mathrm{T}}$, $\mathbf{1}_{m_i} = [1, \ldots, 1]^{\mathrm{T}}$. The symbols $\mathbf{0}_i$ and $\mathbf{1}_i$ denote i-dimensional vectors of zeros and ones, respectively.

To put the template matching problem in the convenient statistical framework, we will assume that the driving noise $n(t)$ in the AR signal description (4), and hence also the signal $s(t)$ itself, are normally distributed. In such a case the vector of prediction errors $\boldsymbol{\eta}(t+q) = \mathbf{s}(t+q) - \widehat{\mathbf{s}}(t+q|t)$ is also normally distributed

$$\boldsymbol{\eta}(t+q) \sim \mathcal{N}(\mathbf{0}_q, \boldsymbol{\Sigma}(t+q|t))$$

where $\boldsymbol{\Sigma}(t+q|t)$ denotes the $q \times q$ covariance matrix of prediction errors which will be derived in the next subsection. Assuming that r signal samples preceding detection alarm are uncorrupted, i.e., $y(t-i) = s(t-i), i = 0, r-1$, one arrives at the following likelihood function

$$p(\mathbf{y}(t+q)|\mathbf{y}_0(t), \mathbf{c}_i, \alpha_i, \beta_i, \tau_i) = p(\boldsymbol{\eta}(t+q)|\mathbf{c}_i, \alpha_i, \beta_i, \tau_i)$$
$$= \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}(t+q|t)|}} \exp\left\{-\frac{1}{2}\,\mathbf{z}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\mathbf{z}_i(\tau_i)\right\}$$

where $\mathbf{y}_0(t) = [y(t), \ldots, y(t-r+1)]^{\mathrm{T}}$ denotes the vector of initial conditions, $|\mathbf{A}|$ denotes determinant of a square matrix $\mathbf{A}$, and

$$\mathbf{z}_i(\tau_i) = \boldsymbol{\varepsilon}(t+q) - \alpha_i \mathbf{r}_i(\tau_i) - \beta_i \mathbf{h}_i(\tau_i)$$

where $\boldsymbol{\varepsilon}(t+q) = [\varepsilon(t+q|t), \ldots, \varepsilon(t+1|t)]^{\mathrm{T}}$.

The maximum likelihood (ML) estimates of the coefficients α_i, β_i and τ_i can be obtained by maximizing the likelihood function specified above, i.e., by minimizing the quadratic cost function

$$J(\alpha_i, \beta_i, \tau_i) = \mathbf{z}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\mathbf{z}_i(\tau_i). \qquad (10)$$

This can be achieved in two steps as follows:

Step 1: Determine the best fitting values of α_i and β_i for consecutive values of $\tau_i = 0, \ldots, q - m_i$

$$\{\widehat{\alpha}_i(\tau_i), \widehat{\beta}_i(\tau_i)\} = \arg\min_{\alpha_i, \beta_i} J(\alpha_i, \beta_i, \tau_i) \ .$$

It can be easily shown that

$$\widehat{\alpha}_i(\tau_i) = \frac{f_1 f_5 - f_3 f_4}{f_1 f_2 - f_3^2} \ , \quad \widehat{\beta}_i(\tau_i) = \frac{f_2 f_4 - f_3 f_5}{f_1 f_2 - f_3^2}$$

where $f_1, \ldots, f_5$ are scalar quantities given by

$$f_1(\tau_i) = \mathbf{h}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\mathbf{h}_i(\tau_i)$$
$$f_2(\tau_i) = \mathbf{r}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\mathbf{r}_i(\tau_i)$$
$$f_3(\tau_i) = \mathbf{h}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\mathbf{r}_i(\tau_i)$$
$$f_4(\tau_i) = \mathbf{h}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\boldsymbol{\varepsilon}(t+q)$$
$$f_5(\tau_i) = \mathbf{r}_i^{\mathrm{T}}(\tau_i)\boldsymbol{\Sigma}^{-1}(t+q|t)\boldsymbol{\varepsilon}(t+q).$$

Step 2: Determine the best value of the location parameter τ_i

$$\widehat{\tau}_i = \arg\min_{\tau_i} J(\tau_i, \widehat{\alpha}_i(\tau_i), \widehat{\beta}_i(\tau_i)).$$

The best-matching click template $\mathcal{C}_{i_0}$, among all feasible ones, can be found using the following rule

$$i_0 = \arg\min_i J(\widehat{\tau}_i, \widehat{\alpha}_i(\widehat{\tau}_i), \widehat{\beta}_i(\widehat{\tau}_i)).$$

The covariance matrix of prediction errors $\boldsymbol{\Sigma}(t+q|t)$ can be easily evaluated using the apprriopiately designed Kalman predictor. The diagonal elements of the matrix $\boldsymbol{\Sigma}(t+q|t)$ – starting from its top left corner and ending at its bottom right corner – can be identified as multi-step-ahead prediction error variances $\sigma_\varepsilon^2(t+q|t), \ldots, \sigma_\varepsilon^2(t+1|t)$; these variances can be also evaluated using the scalar recursive algorithm proposed by Stoica [20].

5. LOCALIZATION AND INTERPOLATION OF CORRUPTED SIGNAL SAMPLES

When the noise pulse detected at the instant t does not match any of the templates, the classical prediction-based approach is used, i.e., detection alarm, started at the instant $t+1$, is terminated at the instant $t+n+1$ if r consecutive prediction errors are sufficiently small

$$|\varepsilon(t+n+j|t)| \leq \mu\widehat{\sigma}_\varepsilon(t+n+j|t), \quad j = 1, \ldots, r$$

or if the length n of the detection alarm reaches its maximum allowable value denoted by $n_{\max}$. Hence, the corresponding detection alarm forms a solid block of "ones": $\widehat{d}(k) = 1$ for $k \in \widehat{D}_t = [\widehat{t}_{\mathrm{B}}, \widehat{t}_{\mathrm{E}}]$, $\widehat{t}_{\mathrm{B}} = t+1$, $\widehat{t}_{\mathrm{E}} = t+n$.

When a particular noise template $\mathcal{C}_{i_0}$ is detected, the classical outlier detection procedure is not pursued and the detection alarm has the form: $\widehat{d}(k) = 1$ for $k \in \widehat{D}_t = [\widehat{t}_\mathrm{B}, \widehat{t}_\mathrm{E}]$, $\widehat{t}_\mathrm{B} = t + \widehat{\tau}_{i0} + 1$, $\widehat{t}_\mathrm{E} = t + \widehat{\tau}_{i0} + m_{i_0}$.

The corrupted signal samples $y(\widehat{t}_\mathrm{B}), \dots, y(\widehat{t}_\mathrm{E})$ are interpolated based on r samples preceding and r samples succeeding the reconstructed fragment – the details can be found e.g. in [12] and [21].

6. ALARM EXTENSION TECHNIQUE

In the presence of "soft" pulse edges, detection alarms 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 μ, 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. The alternative solution, proposed in [12] and recommended also here, is to shift back the front edge of detection alarm (once triggered) by a small, fixed number of samples Δ_1. This means that if detection alarm is raised at the instant $t + 1$, the template matching procedure is initialized at the instant $t - \Delta_1 + 1$ instead of $t + 1$. For the same reason (remember that click templates are created by trimming the average pulse waveforms), once detection alarm based on template matching is determined, it is beneficiary to widen it prior to interpolation by moving back its front edge, and moving forward its back edge by a small, fixed number of samples Δ_2: $\widehat{t}_\mathrm{B} \leftarrow (\widehat{t}_\mathrm{B} - \Delta_2)$, $\widehat{t}_\mathrm{E} \leftarrow (\widehat{t}_\mathrm{E} + \Delta_2)$.

7. EXPERIMENTAL RESULTS

Our repository of clicks was made up of $N = 500$ click waveforms extracted from an old gramophone record. The bidirectional processing algorithm described in [12], which is very precise in determining both the beginning and end points of each noise pulse, was used for this purpose. Based on this training set, 14 click templates, shown in Fig. 1, were established ($\gamma = 0.95$). The information about the size of the corresponding clique (n_i) and the length of click template (m_i) is displayed beneath each plot depicted in Fig. 1. Note that $\sum_{i=1}^{14} m_i = 226$, which means that almost 50% of all extracted noise waveforms were found to be "typical" and utilized in the process of formation of click templates.

The classical outlier detection was based on the AR model of order $r = 10$. Signal identification was carried out using the EWLS algorithm equipped with a forgetting factor $\lambda = 0.99$. The detection multiplier was set to $\mu = 3.5$, the pre-detection similarity threshold was set to $\gamma_0 = 0.8$, and the alarm extension parameters were set to $\Delta_1 = \Delta_2 = \Delta = 2$.

Our test (see Tab. 1) was performed on 5 real archive gramophone recordings, sampled at 22.05 kHz. Two approaches were compared: the AR-model based approach and the one incorporating predictive matched filtering. Since, in the case considered, the reference (clean) audio files were not available, the evaluation had to rely on listening tests. The blind multiple-choice ordering test was carried out, during which each of 20 test persons was asked to indicate the best recording among the two evaluated ones. If both recordings sounded similarly, the listener could mark both of them as the "best" ones. All auditions were made using the same audio set equipped with high-quality headphones designed for critical audio monitoring. The compared recordings, or their selected fragments, could be played back as many times as needed to reach the final conclusion. The advantages of using the proposed technique are clear.

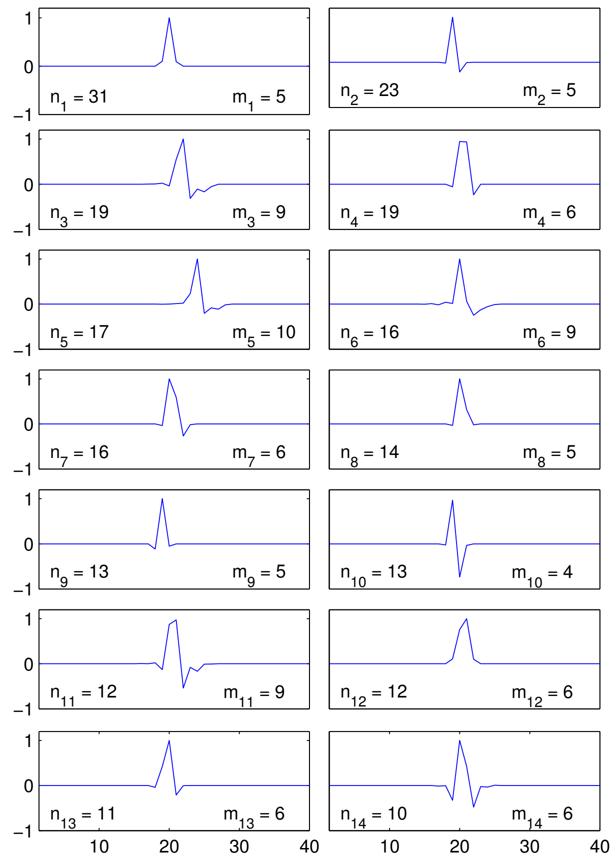

Fig. 1. A collection of 14 click templates obtained for the set of 500 pulse waveforms extracted from a gramophone record. To preserve the original look of noise pulses all waveforms were amplitude-normalized but not debiased.

Table 1. Comparison of the results of declicking based on the proposed predictive matched filtering approach (PMF) and the AR-model based approach (AR).

Recording	Advantage PMF	Advantage AR	Deuce
1	17	3	0
2	18	0	2
3	16	1	3
4	16	0	4
5	18	1	1

8. RELATION TO PRIOR WORK

Even though the term "matched filtering" appeared in early publications on elimination of impulsive disturbances [3], [4], the technique used there is entirely different from the one proposed in this paper. The authors of the abovementioned papers analyzed an impact that an *idealized* (Kronecker-type) noise pulse has on the output of the AR-model based inverse filter. They suggested that in order to localize such pulses in the input (corrupted audio) signal, one could convolve the sequence of one-step-ahead signal prediction errors, yielded by the inverse filter, with the sequence made up of autoregressive coefficients (put in reverse order), and threshold the obtained results. Quite clearly, this approach does not incorporate any knowledge of typical noise patterns.

9. REFERENCES

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

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

[3] S.V. Vaseghi and P.J.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, pp. 791–801, 1992.

[5] 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.

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

[7] S.J. 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, 1995.

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

[9] 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. 18, pp. 1107-1115, 2010.

[10] P.A.A. Esquef, L.W.P. Biscainho and V. V alim aki, "An efficient algorithm for the restoration of audio signals corrupted with low-frequency pulses," *J. Audio Eng. Soc.*, vol. 51, pp. 502-517, 2013.

[11] F.R. Ávila and L.W.P. 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.

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

[13] F. Cazals and C. Karande, "A note on the problem of reporting maximal cliques," *Theor. Comput. Sci.*, vol. 407, pp. 564-568, 2008.

[14] C. Bron and J. Kerbosch, "Algorithm 457: Finding all cliques of an undirected graph," *Comm. ACM*, vol. 16, pp. 575-577, 1973.

[15] J. Konc and D. Janezic, "An improved branch and bound algorithm for the maximum clique problem," *MATCH Commun. Math. Comput. Chem.*, vol 58, pp. 569-590.

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

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

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

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

[20] P. Stoica, "Multistep prediction of autoregressive signals," *Electronics Letters*, vol. 29, pp. 554–555, 1993.

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

[22] J.R. Deller Jr., J.G. Proakis, and J.H.L. Proakis, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.