DETECTION OF IMPULSIVE DISTURBANCES IN ARCHIVE AUDIO SIGNALS

Marcin Ciołek[†] and Maciej Niedźwiecki[†]

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

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

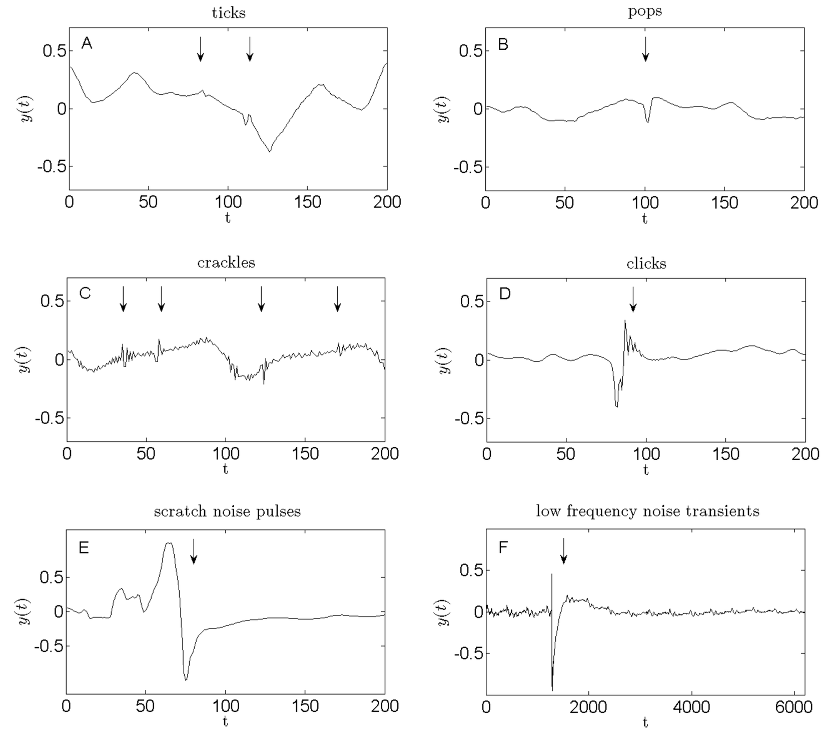
1. Background

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 ticks, pops, clicks and record scratches are usually caused by aging and mishandling of the recording medium, as well as by transmission or equipment artifacts.

Elimination of both kinds of disturbances from archive audio documents is an important element of saving our cultural heritage.

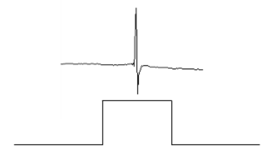


t denotes normalized (dimensionless) discrete time

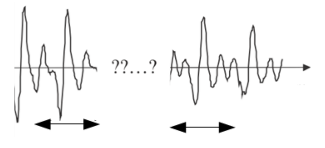
$y(t)$ denotes sampled (frequency rate 48 kSa/s) audio signal

Problem of elimination of impulsive disturbances can be divided into two subproblems:

1. Localization of noise pulses



2. Reconstruction of the corrupted fragments



2. Signal AR modeling

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

2) We are looking for a good estimate $\widehat{d}(t)$ of the pulse location function

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

3) We assume that the noiseless audio signal $s(t)$ obeys the following autoregressive (AR) signal model

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

where $a_1, \ldots, a_n$ denote known autoregressive coefficients and $\eta(t)$ denotes zero-mean white driving noise with variance ρ.

3. Signal reconstruction

Once the noise pulse was localized the corrupted fragment of the signal was reconstructed using the least squares interpolation method.

If one sample is missing at instant t, the interpolation formula can be derived in the form

$$\widetilde{s}(t) = \sum_{i=1}^{n} c_i[s(t-i) + s(t+i)]$$

$$c_i = [a_i - \sum_{j=1}^{n-1} a_j a_{j+1}]/[1 + \sum_{j=1}^{n} a_j^2]$$

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

$$e_*(t) = s(t) - \widetilde{s}(t), \quad \mathrm{var}[e_*(t)] = \rho_*$$

is related to the variance ρ by the equation

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

4. Model identification

The adaptive detection/reconstruction formula can be obtained by replacing known coefficients of the AR model with their estimates, $\widehat{a}_1(t), \ldots, \widehat{a}_n(t)$ and $\widehat{\rho}(t)$, yielded by the finite memory signal identification/ tracking algorithms, such as EWLS, SWLS, LMS or the Kalman filter.

5. Causal detection

The popular noise pulse detection scheme is based on monitoring signal prediction errors:

1) Detection alarm is raised at the instant $t_0 + 1$ if the prediction error statistic $\alpha(t_0+1|t_0)$ exceeds detection threshold $\mu_\alpha^2, \quad \mu_\alpha \in [3, 4.5]$,

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

where

$$\alpha(t_0+1|t_0) = e^2(t_0+1|t_0)/\widehat{\rho}(t_0)$$

$$e(t_0+1|t_0) = y(t_0+1) - \sum_{i=1}^{n} \widehat{a}_i(t_0) y(t_0 - i + 1)$$

2) The test is extended, for $t > t_0$, to multi-step-ahead predictions using the open-loop or decision-feedback scheme

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

3) Detection alarm is terminated at the instant

$$t = t_0 + k_0 + 1 \quad \text{if} \quad \alpha(t-1|t_0) > \mu_\alpha^2$$

and

$$\alpha(t|t_0) \leq \mu_\alpha^2$$
$$t = t_0 + k_0 + 1, \ldots, t_0 + k_0 + n$$

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

Causal detection

The state space description of the AR signal

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{C}\eta(t+1)$$
$$y(t) = \mathbf{C}^{\mathrm{T}}\mathbf{x}(t) + \delta(t)$$

where

$$\mathbf{A} = \left[\frac{\theta^{\mathrm{T}}}{\mathbf{I}_{n-1} \quad \mathbf{0}_{n-1}}\right] \qquad \mathbf{C} = \left[\frac{1}{\mathbf{0}_{n-1}}\right]$$

$\mathbf{x}(t) = [s(t), \ldots, s(t-n+1)]^{\mathrm{T}}$ - state vector

$\theta^{\mathrm{T}} = [a_1, \ldots, a_n]$ - the row vector of AR coefficients

Detection algorithms based on the Kalman filter

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

I. open-loop variant

$$\widehat{\mathbf{x}}(t|t) = \widehat{\mathbf{x}}(t|t-1)$$
$$\mathbf{Q}(t|t) = \mathbf{Q}(t|t-1)$$

II. decision-feedback variant

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

$$\mathbf{L}(t) = \mathbf{Q}(t|t-1)\mathbf{C}/\sigma^2(t|t-1)$$
$$\widehat{\mathbf{x}}(t|t) = \widehat{\mathbf{x}}(t|t-1) + \mathbf{L}(t)e(t|t-1)$$
$$\mathbf{Q}(t|t) = \mathbf{Q}(t|t-1) - \mathbf{L}(t)\sigma^2(t|t-1)\mathbf{L}^{\mathrm{T}}(t)$$

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

$$\widehat{\mathbf{x}}(t|t) = \widehat{\mathbf{x}}(t|t-1)$$
$$\mathbf{Q}(t|t) = \mathbf{Q}(t|t-1)$$

Initial conditions:

$$\mathbf{Q}(\mathrm{t}_0|t_0) = \mathbf{O}_n$$
$$\widehat{\mathbf{x}}(t_0|t_0) = [y(t_0), \ldots, y(t_0-n+1)]^{\mathrm{T}}$$
$$\widehat{\theta}^{\mathrm{T}}(t_0) = [\widehat{a}_1(t_0), \ldots, \widehat{a}_n(t_0)]$$

Termination condition:

$$\alpha(t|t_0) = e^2(t|t-1)/\sigma^2(t|t-1) \leq \mu_\alpha^2$$
$$t = t_0 + k_0 + 1, \ldots, t_0 + k_0 + n$$

6. Semi-causal detection

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

1) New strengthened detection rule

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

where $T_\beta(t_0) = [t_0 + 2 - n, t_0 + 1]$

The interpolation error based statistic is used to confirm or cancel the prediction error alarm

$$\beta(t|t_0) = [e_*(t|t_0)]^2/\widehat{\rho}_*(t_0)$$
$$e_*(t|t_0) = y(t) - \widetilde{y}(t|t_0)$$

where

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

$$\widetilde{y}(t|t_0) = \sum_{i=1}^{n} \widehat{c}_i(t_0)[y(t-i) + y(t+i)]$$

$$\widehat{c}_i(t_0) = [\widehat{a}_i(t_0) - \sum_{j=1}^{n-i} \widehat{a}_j(t_0)\widehat{a}_{j+i}(t_0)]/[1 + \sum_{j=1}^{n} \widehat{a}_j(t_0)^2]$$

2) The test is extended, for $t > t_0$, to multi-step-ahead predictions using the open-loop or decision-feedback scheme

$$\widehat{d}(t) = \begin{cases} 1 & \text{if} \ \alpha(t|t_0) > \mu_\alpha^2 \text{ and} \\ & \beta(t|t_0) > \mu_\beta^2 \\ 0 & \text{elsewhere} \end{cases}$$

3) Detection alarm is terminated at the instant

$$t_* = t_0 + k_0^* + 1$$

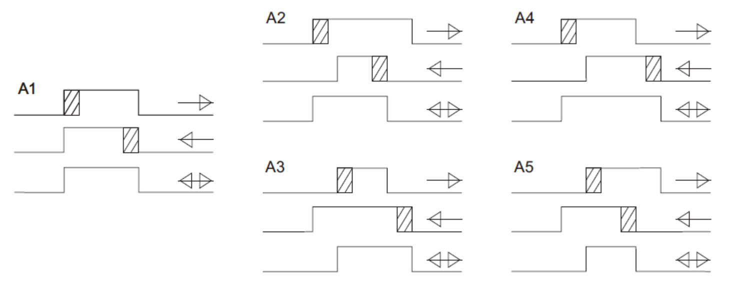
if one of two stop conditions is fulfilled

$$\alpha(t_*-1|t_0) > \mu_\alpha^2 \text{ and } \alpha(t_*-1+i|t_0) \leq \mu_\alpha^2,$$
$$\text{or}$$
$$\beta(t_*-1|t_0) > \mu_\beta^2 \text{ and } \beta(t_*-1+i|t_0) \leq \mu_\beta^2,$$
$$i = 1, \ldots, n$$

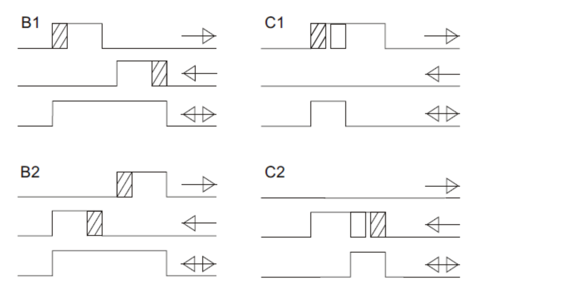
7. Noncausal detection

The bidirectional (forward-backward) detection of noise pulses

Class A



Class B and C



Atomic fusion rules selected by experts and confirmed by objective quality measure PEAQ tool. 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.

8. Experimental results

Artificially corrupted signals

40 = 4 × 10 clean audio recordings representing 4 music categories (jazz, choir, opera, classical), lasting for about 22 seconds each, sampled at the rate of 48 kSa/s and contaminated with real click waveforms extracted from silent parts of old gramophone recording.

Objective sound quality measure

The Perceptual Evaluation of Audio Quality (PEAQ) tool – a specialized software which scores the restored audio (by comparing it with the original, noiseless recording) using several perceptual criteria.

Table 1. Comparison of the average PEAQ scores obtained for 8 unidirectional/ bidirectional detection algorithms. All results were obtained for 40 artificially corrupted audio files. The average score of the input (corrupted) recordings was equal to −3.6 and the average "ground truth" score, obtained when interpolation of the corrupted samples was based on exact knowledge of pulse locations, was equal to −0.29.

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

PEAQ

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

Unidirectional algorithms:

A — causal, equipped with open-loop detection scheme,
B — semi-causal, equipped with open-loop detection scheme,
C — causal, equipped with decision-feedback detection scheme,
D — semi-causal, equipped with decision-feedback detection scheme.

Bidirectional algorithms, i.e. noncausal extensions of algorithms A, B, C, D denoted by A*, B*, C*, D*, respectively.

The average PEAQ score obtained for the combined least squares autoregressive+sinusoid (LSAR+SIN) method was −0.88.

Informal listening tests, performed on real archive gramophone recordings, support the above findings.

9. Conclusions

1. New pulse detection rules combines analysis of one-step-ahead signal prediction errors with critical evaluation of leave-one-out signal interpolation errors.

2. The new detectors have increased ability to reliably cancel false alarms.

3. Perceptual scores, obtained using the PEAQ tool, confirm that the proposed detection rules yield better results than the classical ones, based on evaluation of signal prediction errors only.

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