ELIMINATION OF CLIKS AND BACKGROUND NOISE FROM ARCHIVE GRAMOPHONE RECORDINGS USING THE "TWO TRACK MONO" APPROACH

Maciej NIEDŹWIECKI
Faculty of Electronics
Department of Automatic Control, Technical University of Gdańsk
ul. Narutowicza 11/12, Gdańsk , Poland
Tel: + 48 58 472519; fax +48 58 415821
e-mail: maciekn@sunrise.pg.gda.pl

Abstract

Old gramophone recordings are corrupted with a wideband noise (granulation noise) and impulsive disturbances (cliks, pops, record scratches) - both caused by aging and/or mishandling of the vinyl material. The paper presents an improved method of gramophone noise reduction which makes use of two signals obtained when a mono record is played back using the stereo equipment.

1 Problem statement

When old mono gramophone records are played back using the stereo equipment one obtains two signals which can be used for sound renovation. Except for large scratches which are observed in both "channels", small cliks usually occur in the left and right track at different locations allowing for application of more efficient (compared to the traditional "single track" approach) outlier detection/elimination schemes. Similarly, due to the lack of spatial correlation between the recording medium degradation on both sides of the groove, the background noise corrupting the left track is practically uncorelated with that affecting the right one. Combining both tracks one can therefore increase the signal to noise ratio which results in improved restoration quality.

Following the lines of [3] and [4], the audio signal $s(t)$ is described by the time-varying autoregressive (AR) model of order p given in the state space form

$$\varphi(t+1) = A[\theta(t)]\varphi(t) + be(t) \tag{1}$$

where $\varphi(t) = [s(t-1), \ldots, s(t-q)]^T$, $q \geq p$, denotes the regression vector, $\theta(t) = [a_1(t), \ldots, a_p(t)]^T$ is the vector of autoregressive coefficients, $\{e(t)\}$ is the white gaussian noise and

$$A[\theta(t)] =$$

$$= \begin{bmatrix} a_1(t) & \cdots & a_{p-1}(t) & a_p(t) & 0 & \cdots & 0 & 0 \\ 1 & & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$b^T = [\ 1 \ \cdots \ 0 \ 0 \ 0 \ \cdots \ 0 \ 0\ ]$$

Note that for $q > p$, (1) is a nonminimal state-space realization of the $AR(p)$ process. The "superfluous" components of the regression vector $s(t-p-1), \ldots, s(t-q)$ are introduced to enable for two-sided reconstruction of blocks of irrevocably distorted samples (up to $q-p-1$ samples in a row).

It is assumed that the evolution of time-varying process coefficients can be (locally) described by the random walk model

$$\theta(t+1) = \theta(t) + w(t) \tag{2}$$

where $\{w(t)\}$ is another gaussian white noise sequence, independent of $\{e(t)\}$.

Finally, it is assumed that the original audio signal is corrupted by the mixture of a broadband noise $(z_L(t), z_R(t))$ and impulsive disturbances $(v_L(t), v_R(t))$, i.e., the measurements corresponding to the left/right channel take the form

$$\begin{aligned} y_L(t) &= b^T\varphi(t) + z_L(t) + v_L(t) \\ y_R(t) &= b^T\varphi(t) + z_R(t) + v_R(t) \end{aligned} \tag{3}$$

or equivalently

$$y(t) = B^T\varphi(t) + z(t) + v(t)$$

where $y(t)=[y_L(t),y_R(t)]^T$, $z(t)=[z_L(t),z_R(t)]^T$, $v(t)=[v_L(t),v_R(t)]^T$ and $B=[b|b]$.
We will regard $\{z_L(t)\}$ and $\{z_R(t)\}$ as two normal, mutually uncorrelated white noise sequences

$$z_{L/R}(t)\sim\mathcal{N}(0,\sigma_z^2),\quad z_L(t)\perp z_R(t)$$

We will not attempt to use a detailed probabilistic model of impulsive disturbances, i.e., a model taking into account geometry and/or intensity of cliks specific for a particular archive recording. The following coarse "save or reject" model will be adopted instead

$$\begin{aligned} v_{L/R}(t) &\sim \mathcal{N}(0,\sigma_{L/R}^2(t)) \\ \sigma_{L/R}^2(t) &= \begin{cases} 0 & \text{if} \quad d_{L/R}(t)=0 \\ \infty & \text{if} \quad d_{L/R}(t)=1 \end{cases} \end{aligned}$$

where $d_L(t)$ and $d_R(t)$ denote the noise pulse indicators for the left and right channel, respectively

$$d_{L/R}(t)=\begin{cases} 0 & \text{noise impulse absent} \\ 1 & \text{noise impulse present} \end{cases}$$

Quite obviously, by putting $\sigma_{L/R}^2(t)=\infty$ one indicates that the measurement taken at instant t bears no information about the recovered signal, i.e., the corresponding sample should be regarded as if it was simply missing.
Combining the regression vector $\varphi(t)$ and a parameter vector $\theta(t)$ in a $(q+p)$-dimensional "state" vector $x(t)=[\varphi^T(t),\theta^T(t)]^T$ one can rewrite equations (1) - (3) in the form

$$\begin{aligned} x(t+1) &= f[x(t)]+\omega(t) \\ y(t) &= C^T x(t)+\zeta(t) \end{aligned} \qquad (4)$$

where

$$f[x(t)]=\begin{bmatrix} A(t) & 0 \\ 0 & I_p \end{bmatrix} x(t),\quad \omega(t)=\begin{bmatrix} ce(t) \\ w(t) \end{bmatrix}$$

$$\zeta(t)=z(t)+v(t)$$

and $C=[c|c]$, $c^T=[b^T,0^T]$.
The problem of simultaneous identification of time-varying process characteristics (estimation of $\theta(t)$) and recovering the audio signal from noise (estimation of $\varphi(t)$) can be therefore regarded as a nonlinear filtering problem in the state space. In the next section we will describe a suboptimal algorithm for estimation of $x(t)$ based on the theory of extended Kalman filter (EKF).

2 The proposed solution

Denote by $F(t)$ the state transition matrix of the linearized system

$$F(t)=\nabla_x f[x]_{|x=\widehat{x}(t|t)}=\begin{bmatrix} A(t|t) & \begin{matrix}\widehat{\varphi}^T(t|t)\\ 0\end{matrix} \\ 0 & I_p \end{bmatrix}$$

where $A(t|t)=A[\widehat{\theta}(t|t)]$ and

$$\widehat{x}(t|t)=\begin{bmatrix} \widehat{\varphi}(t|t) \\ \widehat{\theta}(t|t) \end{bmatrix}$$

is the filtered state trajectory yielded by the EKF algorithm. Let $\Omega(t)=\mathrm{cov}[\omega(t)]/\sigma_e^2$ and $Z(t)=\mathrm{cov}[\zeta(t)]/\sigma_e^2$. The equations of the extended Kalman filter for the system governed by (4) take the form (cf. [1])

State estimation

$$\begin{aligned} \widehat{x}(t+1|t) &= f[\widehat{x}(t|t)] \\ \Sigma(t+1|t) &= F(t)\Sigma(t|t)F^T(t)+\Omega(t) \\ \widehat{x}(t|t) &= \widehat{x}(t|t-1)+L(t)\epsilon(t) \\ \Sigma(t|t) &= \Sigma(t|t-1)-L(t)C^T\Sigma(t|t-1) \end{aligned} \qquad (5)$$

where $L(t)$ is the Kalman gain

$$L(t)=\Sigma(t|t-1)C[C^T\Sigma(t|t-1)C+Z(t)]^{-1}$$

and $\epsilon(t)$ denotes the prediction error

$$\epsilon(t)=\begin{bmatrix} \epsilon_L(t) \\ \epsilon_R(t) \end{bmatrix}=\begin{bmatrix} y_L(t)-c^T\widehat{x}(t|t-1) \\ y_R(t)-c^T\widehat{x}(t|t-1) \end{bmatrix}$$

Signal renovation/reconstruction

Since $\widehat{\varphi}(t|t)=[E[s(t)|Y(t)],\ldots,E[s(t-q)|Y(t)]]^T$ the smoothed estimate of the audio signal can be obtained from

$$\widehat{s}(t-q|t)=d^T\widehat{\varphi}(t|t)=d_o^T\widehat{x}(t|t)$$

where $d^T=[0,\ldots,0,1]$ and $d_o^T=[d_o^T,0^T]$.
When $d(t-q)=1$ the quantity $\widehat{s}(t-q|t)$ can be interpreted as an optimal, in the mean square sense, reconstruction of the sample $s(t-q)$ based on all "past" and $q-1$ "future" measurements (except those identified as outliers). For an $AR(p)$ signal a block consisting of at least p outlier-free "future" samples is needed to guarantee good quality of reconstruction of a corrupted fragment [2]. Therefore, increasing q beyond $p+1$ one obtains an algorithm capable of reconstructing a series of outliers (up to $q-p-1$ in a row) arising when a record scratch is encountered.
Several fixes will help us turn (5) into a workable estimation scheme.

2.1 Explicit form of the *a posteriori* updates

Denote by $\gamma(t)=\Sigma(t|t-1)c$ the first column of the *a priori* covariance matrix $\Sigma(t|t-1)$ and by $\beta(t)=c^T\Sigma(t|t-1)c$ the element placed in its upper left corner. It is straightforward to check that under the assumptions made

$$L(t)=\frac{\gamma(t)}{\alpha(t)}[1-d_L(t)|1-d_R(t)]$$

where

$$\alpha(t)=\begin{cases}2\beta(t)+\kappa & \text{if } d_{L/R}(t)=0\\ \beta(t)+\kappa & \text{otherwise}\end{cases}$$

and $\kappa=\sigma_z^2/\sigma_e^2$.
Consequently, the last two recursions of the EKF filter can be put down in the form

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

$$+\begin{cases}\frac{\gamma(t)}{\beta(t)+\frac{\kappa}{2}}\bar{\epsilon}(t) & \text{if} \quad d_L(t)=0, d_R(t)=0\\ \frac{\gamma(t)}{\beta(t)+\kappa}\epsilon_L(t) & \text{if} \quad d_L(t)=0, d_R(t)=1\\ \frac{\gamma(t)}{\beta(t)+\kappa}\epsilon_R(t) & \text{if} \quad d_L(t)=1, d_R(t)=0\\ 0 & \text{if} \quad d_L(t)=1, d_R(t)=1\end{cases} \tag{6}$$

$$\Sigma(t|t)=\Sigma(t|t-1)+$$

$$-\begin{cases}\frac{\gamma(t)\gamma^T(t)}{\beta(t)+\frac{\kappa}{2}} & \text{if} \quad d_L(t)=0, d_R(t)=0\\ \frac{\gamma(t)\gamma^T(t)}{\beta(t)+\kappa} & \text{if} \quad d_L(t)=0, d_R(t)=1\\ \frac{\gamma(t)\gamma^T(t)}{\beta(t)+\kappa} & \text{if} \quad d_L(t)=1, d_R(t)=0\\ 0 & \text{if} \quad d_L(t)=1, d_R(t)=1\end{cases} \tag{7}$$

where

$$\bar{\epsilon}(t)=\frac{\epsilon_L(t)+\epsilon_R(t)}{2}$$

denotes the prediction error averaged over two tracks.

2.2 Parameter tracking

Assuming that the changes of different process parameters are mutually independent and occur at the same average rate, i.e., cov[w(t)] $=\sigma_w^2 I_p$, one gets

$$\Omega=\begin{bmatrix} bb^T & 0\\ 0 & \xi I_p\end{bmatrix}$$

where $\xi=\sigma_w^2/\sigma_e^2$.
When such simplified model of parameter variation is adopted one can influence the EKF algorithm by means of adjusting two scalar coefficients : κ - deciding upon the degree of signal smoothing and ξ - controlling the parameter adaptation rate.

2.3 Detection of outliers

Since localization of impulsive disturbances is not known *a priori* the quantities $d_L(t)$ and $d_R(t)$ in (6) and (7) should be replaced with the corresponding estimates $\widehat{d}_L(t)$ and $\widehat{d}_R(t)$, respectively.
Owing to the properties of the Kalman filter

$$p(\epsilon_{L/R}(t)|Y(t-1), d_{L/R}(t)=0)\cong\mathcal{N}(0,\sigma_\epsilon^2(t))$$

where $\sigma_\epsilon^2(t))=\eta(t)\sigma_e^2$ and $\eta(t)=c^T\Sigma(t|t-1)c+\kappa$ $=\beta(t)+\kappa$.
A reasonable outlier detection rule can be therefore defined in the form

$$\widehat{d}_{L/R}(t)=\begin{cases}0 & \text{if} \quad |\epsilon_{L/R}(t)|\le\mu\widehat{\sigma}_\epsilon(t)\\ 1 & \text{if} \quad |\epsilon_{L/R}(t)|>\mu\widehat{\sigma}_\epsilon(t)\end{cases}$$

where $\widehat{\sigma}_\epsilon^2(t)=\eta(t)\widehat{\sigma}_e^2(t-1)$, μ is a user-dependent detection threshold (in most cases the best results are obtained for $\mu\in[3,5]$) and $\widehat{\sigma}_e^2(t)$ is the local (exponentially weighted) maximum likelihood estimate of the input noise variance

$$\widehat{\sigma}_e^2(t)=$$

$$\begin{cases}\lambda\widehat{\sigma}_e^2(t-1)+(1-\lambda)\frac{\epsilon_L^2(t)+\epsilon_R^2(t)}{2\eta(t)} & \text{if } \widehat{d}_{L/R}(t)=0\\ \widehat{\sigma}_\epsilon^2(t-1) & \text{otherwise}\end{cases}$$

$$0<1-\lambda\ll 1$$

3 Refinements

3.1 Track alignment

Even though theoretically the signals observed in both "tracks" of the mono recording should differ in their noise, i.e., high frequency components only, the actual differencies are more substantial. Due to the fact that the surface of the vinyl record is not ideally flat and/or the two playback channels are not perfectly balanced a low frequency drift between both tracks is usually observed and should be appropriately dealt with. It is important to realize that the presence of a nonnegligible offset between the left and right track can substantially reduce efficiency of the proposed scheme. First, due to increase in the mean square prediction error the sensitivity of the outlier detector can be noticeably decreased. Second, and more importantly, signal discontinuities may appear at instants where the algorithm switches from the "combined track" restoration ($\widehat{d}_L(t)=\widehat{d}_R(t)$=0) to the "single track" restoration ($\widehat{d}_L(t)=1$,$\widehat{d}_R(t)=0$ or $\widehat{d}_L(t)=0$,$\widehat{d}_R(t)=1$)

and *vice versa*. A simple adaptive preprocessing algorithm can be used to adjust the signals $y_L(t)$ and $y_R(t)$.
Denote by $\delta(t)$ the difference between the left and right channel

$$\delta(t) = y_L(t) - y_R(t)$$

and by $\widehat{d}_\delta(t)$ the output of the outlier detector similar to that described in the preceding section

$$\widehat{d}_\delta(t) = \begin{cases} 0 & \text{if} \quad |\delta(t)| \le 3\widehat{\sigma}_\delta(t-1) \\ 1 & \text{if} \quad |\delta(t)| > 3\widehat{\sigma}_\delta(t-1) \end{cases}$$

where $\widehat{\sigma}^2_\delta(t)$ is the local estimate of the noise variance

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

The following simple sliding-window algorithm can be used for the purpose of aligning the right track with the left one

$$\widetilde{y}_R(t) = y_R(t) + \bar{\delta}(t) \tag{8}$$

$$y^*_R(t) = g(t)\widetilde{y}_R(t) \tag{9}$$

where

$$\begin{aligned} \bar{\delta}(t) &= \frac{1}{m(t)} \sum_{i=t-M}^{t+M} \delta(i)(1 - \widehat{d}_\delta(i)) \\ g(t) &= \frac{1}{m(t)} \sum_{i=t-M}^{t+M} y_L(i)\widetilde{y}_R(i)(1 - \widehat{d}_\delta(i)) \\ m(t) &= \sum_{i=t-M}^{t+M} (1 - \widehat{d}_\delta(i)) \end{aligned}$$

and M determines the size of the local analysis window (the values from the range [20,50] are recommended).
Note that the debiasing filter (8) and the scaling filter (9) are linear phase and that both can be easily put in a recursive form.

3.2 Outlier detection revisited

Based on the results of many tests performed on real audio signals some heuristic modifications to the outlier detection scheme described in section 2 can be recommended.
First of all, it was found out that the renovation algorithm works more reliably if detection alarms are clustered, i.e., if they are forced to form solid detection blocks, each consisting of a sequence of "ones" (up to $q-p-1$ in a row) preceded and succeeded by at least p "zeroes". Alarm clustering allows one to avoid reconstruction errors caused by "accidental acceptancies" of samples localized in the middle of long-lasting artifacts such as record scratches.
The second modification we suggest is concerned with the way of combining detection alarms in both tracks.

Cautious detector

If detection blocks in the left and right channel partially overlap, i.e., if it holds

$$\begin{aligned} \widehat{d}_L(t) = 1 \quad &\text{for} \quad t \in T_L = [t_1, t_2] \\ \widehat{d}_R(t) = 1 \quad &\text{for} \quad t \in T_R = [t_3, t_4] \end{aligned}$$

$$T_L \cap T_R \ne O$$

then put

$$\widehat{d}_L(t) = \widehat{d}_R(t) = 1 \quad \text{for} \quad t \in T_L \cup T_R$$

According to this rule each time a disturbance is spotted in both channels (even if not at exactly the same time instants) a join, possibly enlarged, detection block is formed. Otherwise, that is if detection alarms in one channel are not accompanied by alarms in the other one the original detection rule is used, i.e., the restoration is based on the material taken from the uncorrupted track.

References

[1] B.D.O. Anderson and J.B. Moore, "Optimal filtering", Prentice-Hall, 1979.

[2] M. Niedźwiecki, "Statistical reconstruction of multivariate time series", IEEE Trans. Signal Processing, vol. 41, pp. 451-457, 1993.

[3] M. Niedźwiecki, "Recursive algorithm for elimination of measurement noise and impulsive disturbances from ARMA signals", 7th European Signal Processing Conference, pp. 1289-1292, Edinburgh, 1994.

[4] M. Niedźwiecki and K. Cisowski, "Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals", IEEE Trans. Signal Processing, vol. 44, March 1996.