

ADAPTIVE SCHEME FOR ELIMINATION OF BACKGROUND NOISE AND IMPULSIVE DISTURBANCES FROM AUDIO SIGNALS

Maciej NIEDŹWIECKI
Krzysztof CISOWSKI

Faculty of Electronics
Department of Automatic Control
Technical University of Gdańsk
ul. Narutowicza 11/12
Gdańsk, Poland

RÉSUMÉ

On considère le problème de l'élimination du bruit de fond et de signaux acoustiques. On a montré que la tâche de la détection/alignement/restitution peut être formulée comme un problème de filtrage non linéaire et résolu par l'usage de la théorie du filtre de Kalman extensif. Les tests numériques exécutés pour les signaux acoustique réels confus par des perturbations réels et produits artificiellement confirment très bonnes propriétés de l'algorithme proposé et suggèrent, entre autres, être utile pour la restitution de vieux disques à gramophone.

1 Introduction

Even though the three operations comprising the task of adaptive restoration, namely : signal filtering/reconstruction, process identification and outlier detection were treated separately by many authors [1] - [6] no unified approach combining all three elements was presented so far.

Due to the lack of consistent theoretical analysis all restoration packages developed so far (see e.g. [7]) are basically "tool-boxes" consisting of a number of algorithms designed for different purposes and combined in a more or less mechanical way. Obviously, the subtle interplay which ties together different aspects of adaptive restoration is in this case ignored.

The method described in this paper overcomes the above-mentioned limitations and seems to be the first systematic treatment of the general restoration problem (combined background noise and outlier elimination in a time-varying environment). It is shown that the task of simultaneous detection/tracking/restoration can be stated as a nonlinear filtering problem and solved using the theory of extended Kalman filter (EKF).

The proposed algorithm can be viewed as a combination of two Kalman filters coupled in a nonlinear fashion. The first filter is designed to track slow parameter variations (a standard random walk model of parameter changes is adopted) and the second one is used for the purpose of on-line recovering of the regression vector (made up of past signal measurements). The outlier detector is based on monitoring of prediction errors and its decisions are carried out by means of the proper covariance assignment. At each time instant the decision threshold of the detector is determined in accordance with some accuracy measures updated by the EKF algorithm.

ABSTRACT

The problem of elimination of background noise and impulsive disturbances from audio signals is considered. It is shown that the task of simultaneous detection/tracking/restoration can be stated as a nonlinear filtering problem and solved using the theory of extended Kalman filter. Numerical tests carried out for real audio signals corrupted with both real and artificially generated disturbances confirm very good properties of the proposed algorithm and suggest, among others, its usefulness for the purpose of restoration of old gramophone recordings.

2 The proposed algorithm

2.1 Adaptive filtering

Consider the problem of filtering of the discrete time signal $s(t)$, $t = 1, 2, \dots$ based on the measurements corrupted with the white background noise $v(t)$

$$y(t) = s(t) + v(t) \quad (1)$$

Suppose that the signal $s(t)$ can be modelled as an autoregressive (AR) process of order p

$$s(t+1) = \sum_{i=1}^p a_i(t)s(t-i+1) + e(t) \quad (2)$$

driven by white noise sequence $\{e(t)\}$. Assume that both noise processes are independent and Gaussian

$$\begin{aligned} v(t) &\sim \mathcal{N}(0, \sigma_v^2) \\ e(t) &\sim \mathcal{N}(0, \sigma_e^2) \\ v(t) &\perp e(t) \end{aligned}$$

Finally, suppose that unknown process parameters $a_1(t), \dots, a_p(t)$ are subject to slow variation and that their evolution can be locally described by the following random walk model

$$\begin{aligned} a_i(t+1) &= a_i(t) + w_i(t) \\ i &= 1, \dots, p \end{aligned} \quad (3)$$

where $\{w(t)\}$, $w(t) = [w_1(t), \dots, w_p(t)]^T$ is the white Gaussian noise sequence, independent of $\{e(t)\}$ and $\{v(t)\}$

$$w(t) \sim \mathcal{N}(0, \sigma_w^2 I)$$



We note that the diagonal form of the adopted covariance matrix of $\{w(t)\}$ implies that signal parameters vary independently of each other which is a pretty standard simplifying assumption - see [2] for further details. Note also that the quantity σ_w^2 can be interpreted as the mean square rate of parameter variation.

We will show that under the assumptions made the problem of recovering of the signal $s(t)$ based on the set of noisy measurements $Y(t) = \{y(t), \dots, y(1)\}$ can be stated as a nonlinear filtering problem in the state space and solved using the theory of extended Kalman filter (EKF). Actually, introducing the notation

$$\theta(t) = \begin{bmatrix} a_1(t) \\ \vdots \\ a_p(t) \end{bmatrix} \quad \varphi(t) = \begin{bmatrix} s(t) \\ \vdots \\ s(t-p+1) \end{bmatrix} \quad c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and

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

equations (1) - (3) can be rewritten in the form

$$\begin{aligned} \varphi(t+1) &= A(t)\varphi(t) + ce(t) \\ \theta(t+1) &= \theta(t) + w(t) \\ y(t) &= c^T \varphi(t) + v(t) \end{aligned} \quad (4)$$

or, using a more compact notation, in the form

$$\begin{aligned} x(t+1) &= f[x(t)] + \omega(t) \\ y(t) &= c_0^T x(t) + v(t) \end{aligned} \quad (5)$$

where

$$x(t) = \begin{bmatrix} \varphi(t) \\ \theta(t) \end{bmatrix} \quad \omega(t) = \begin{bmatrix} ce(t) \\ w(t) \end{bmatrix} \quad c_0 = \begin{bmatrix} c \\ 0 \end{bmatrix}$$

and

$$f[x(t)] = \begin{bmatrix} A(t) & 0 \\ 0 & I \end{bmatrix} x(t)$$

Since $f[\cdot]$ is a nonlinear function (note that $A(t)$ depends on $\theta(t)$ and hence on $x(t)$) equations (5) comprise a nonlinear filtering problem which can be solved using the EKF approach. Denote by $F(t)$ the state transition matrix of the linearized system

$$F(t) = \nabla_x f[x]_{|x=\hat{x}(t|t)} = \begin{bmatrix} A(t|t) & \hat{\varphi}^T(t|t) \\ 0 & I \end{bmatrix} \quad (6)$$

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

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

is the filtered state trajectory yielded by the EKF algorithm. Let

$$\Omega = \text{cov}[\omega(t)]/\sigma_e^2 = \begin{bmatrix} cc^T & 0 \\ 0 & \xi I \end{bmatrix}, \quad \xi = \frac{\sigma_w^2}{\sigma_e^2} \quad (7)$$

Then equations of the extended Kalman filter for the system governed by (5) take the form (cf. [1])

$$\begin{aligned} \hat{x}(t+1|t) &= \begin{bmatrix} A(t|t) & 0 \\ 0 & I \end{bmatrix} \hat{x}(t|t) \\ \hat{x}(t|t) &= \hat{x}(t|t-1) + \tilde{L}(t)\epsilon(t) \\ \epsilon(t) &= y(t) - c_0^T \hat{x}(t|t-1) \\ &= y(t) - \hat{\theta}^T(t-1|t-1)\hat{\varphi}(t-1|t-1) \\ \tilde{L}(t) &= \frac{\tilde{\Sigma}(t|t-1)c_0}{c_0^T \tilde{\Sigma}(t|t-1)c_0 + \kappa_0} \\ \tilde{\Sigma}(t|t) &= (I - \tilde{L}(t)c_0^T)\tilde{\Sigma}(t|t-1) \\ \tilde{\Sigma}(t+1|t) &= F(t)\tilde{\Sigma}(t|t)F^T(t) + \Omega \end{aligned} \quad (8)$$

where

$$\kappa_0 = \frac{\sigma_v^2}{\sigma_e^2}$$

and the matrices $\tilde{\Sigma}(t|t)$ and $\tilde{\Sigma}(t+1|t)$ stand for *approximate* normalized (with respect to the input noise variance σ_e^2) *a posteriori* and *a priori* covariance matrices of the estimation error

$$\begin{aligned} \text{cov}[x(t)|Y(t)] &\cong \sigma_e^2 \tilde{\Sigma}(t|t) \\ \text{cov}[x(t+1)|Y(t)] &\cong \sigma_e^2 \tilde{\Sigma}(t+1|t) \end{aligned} \quad (9)$$

The algorithm (8) can be viewed as a combination of two Kalman filters coupled in a nonlinear fashion. The first filter is designed to remove the measurement noise and the second one is used for the purpose of parameter tracking. The tracking sub-filter uses the estimate of the regression vector $\varphi(t)$ yielded by the signal restoration sub-filter. Similarly, the restoration sub-filter relies on current estimates of process coefficients $\theta(t)$ worked out by the tracking sub-filter. The scalar measurement noise variance to driving noise variance ratio $\kappa_0 \geq 0$ and parameter change variance to driving noise variance ratio ξ , $1 \gg \xi \geq 0$, are two important user-dependent "knobs". Vaguely speaking, the value of the noise-to-signal ratio κ_0 decides upon the degree of signal smoothing introduced by the EKF algorithm (if κ_0 is set to zero the input signal remains unmodified). The quantity ξ controls the "effective memory" of the parameter tracking part of the algorithm and should be chosen in accordance with the rate of nonstationarity of the identified process [6] (if process coefficients are time-invariant ξ should be zero - otherwise it should be set to a small positive number).

2.2 Outlier detection/elimination

A simple outlier detector can be built using the "3-sigma" rule frequently utilized in statistics. Observe that

$$\hat{y}(t|t-1) = E[y(t)|Y(t-1)] = E[s(t)|Y(t-1)] = c_0^T \hat{x}(t|t-1)$$

and hence

$$\epsilon(t) = y(t) - \hat{y}(t|t-1) = c_0^T (x(t) - \hat{x}(t|t-1)) + v(t)$$

leading to (cf.(9))

$$p(\epsilon(t)) = \mathcal{N}(0, \sigma_e^2(t)) \quad (10)$$

where

$$\begin{aligned} \sigma_e^2(t) &= \eta(t|t-1)\sigma_e^2 \\ \eta(t|t-1) &= c_0^T \tilde{\Sigma}(t|t-1)c_0 + \kappa_0 \end{aligned} \quad (11)$$



A reasonable "consistency test" based on (10) becomes

$$d(t) = \begin{cases} 0 & \text{if } |\epsilon(t)| \leq 3\sigma_\epsilon(t) \text{ (noise impulse absent)} \\ 1 & \text{if } |\epsilon(t)| > 3\sigma_\epsilon(t) \text{ (noise impulse present)} \end{cases} \quad (12)$$

According to (12) the sample is classified as an outlier if the magnitude of the corresponding prediction error exceeds three times its standard deviation.

Denote by $\tilde{\Sigma}_\varphi(t|t)$, $\tilde{\Sigma}_{\varphi\theta}(t|t) = \tilde{\Sigma}_{\theta\varphi}^T(t|t)$ and $\tilde{\Sigma}_\theta(t|t)$ the $p \times p$ blocks of the *a posteriori* covariance matrix $\tilde{\Sigma}(t|t)$ associated with $\hat{x}(t|t)$

$$\tilde{\Sigma}(t|t) = \begin{bmatrix} \tilde{\Sigma}_\varphi(t|t) & \tilde{\Sigma}_{\varphi\theta}(t|t) \\ \tilde{\Sigma}_{\theta\varphi}(t|t) & \tilde{\Sigma}_\theta(t|t) \end{bmatrix}$$

Note that since

$$\begin{aligned} \eta(t|t-1) &= c_0^T \tilde{\Sigma}(t|t-1) c_0 + \kappa_0 \\ &= \hat{\varphi}^T(t-1|t-1) \tilde{\Sigma}_\theta(t-1|t-1) \hat{\varphi}(t-1|t-1) \\ &\quad + \hat{\theta}^T(t-1|t-1) \tilde{\Sigma}_\varphi(t-1|t-1) \hat{\theta}(t-1|t-1) \\ &\quad + 2\hat{\theta}^T(t-1|t-1) \tilde{\Sigma}_{\varphi\theta}(t-1|t-1) \hat{\varphi}(t-1|t-1) + \kappa_0 + 1 \end{aligned}$$

both parameter and data uncertainties are taken into account when determining the decision threshold in (12).

Only minor changes are needed to adjust the EKF filter (8) so that it could cope with outliers. Actually, consider replacing the constant variance ratio κ_0 in (8) with the following time-varying one

$$\kappa(t) = \begin{cases} \kappa_0 & \text{if } d(t) = 0 \\ \infty & \text{if } d(t) = 1 \end{cases} \quad (13)$$

Quite obviously, by putting $\kappa(t) = \infty$ in (8) one indicates that the corresponding measurement is corrupted with noise of "infinite variance" ($\sigma_v^2(t) = \infty$) and hence that it bears no information about the recovered signal. Note that $\kappa(t) = \infty$ entails

$$\begin{aligned} \tilde{L}(t) &= 0 \\ \hat{x}(t|t) &= \hat{x}(t|t-1) \\ \tilde{\Sigma}(t|t) &= \tilde{\Sigma}(t|t-1) \end{aligned} \quad (14)$$

with the effect that the data point $y(t)$ - even though available - is ignored by the estimation routine.

Remark

Knowledge of the variance of the driving noise $\{e(t)\}$ may be essential in order to tune the EKF algorithm properly. First, the value of σ_e^2 is needed to establish the noise-to-signal ratio κ_0 in (8). Second, and more importantly, it acts as a scaling factor in the expression for the variance of the prediction error (c.f. (11)) and may hence strongly influence efficiency of the outlier detector (12).

Since the variance of the driving noise is related to the power of the signal $s(t)$ it would be pretty naive to assume that a single "global" estimate of σ_e^2 can be found suitable for all time instants t . In the absence of outliers the following "local" exponentially weighted maximum likelihood estimator of σ_e^2 can be used (cf. Bohlin [2])

$$\sigma_e^2(t) = \frac{r(t)}{k(t)} \quad (15)$$

where

$$r(t) = \sum_{i=1}^t \lambda^{t-i} \frac{\epsilon^2(i)}{\eta(i|i-1)} = \lambda r(t-1) + \frac{\epsilon^2(t)}{\eta(t|t-1)}$$

$$k(t) = \sum_{i=1}^t \lambda^{t-i} = \lambda k(t-1) + 1 \quad (16)$$

and λ , $0 < \lambda < 1$ is the so-called forgetting constant determining the "effective memory" of the estimator.

Replacing (16) with $r(t) = r(t-1)$ and $k(t) = k(t-1)$ whenever $d(t) = 1$ one obtains a simple estimation algorithm which can be used to track slow variations in σ_e^2 in the presence of outliers.

2.3 Smoothing/reconstruction

The optimal, in the mean square sense, estimate of $s(\tau)$ given the data set $Y(t)$ takes the well-known form [1]

$$\hat{s}(\tau|t) = E[s(\tau)|Y(t)]$$

If $t > \tau$ the quantity $\hat{s}(\tau|t)$ is called the smoothed estimate of $s(\tau)$ as it combines at instant τ not only all "past" measurements but also a certain number of "future" ones (of course this is possible if a decision delay of $t - \tau$ sampling periods is introduced).

It is straightforward to note that the Kalman filter (5) can serve as a fixed lag smoother. Actually, we have

$$\hat{\varphi}(t|t) \cong \begin{bmatrix} E[s(t)|Y(t)] \\ \vdots \\ E[s(t-p+1)|Y(t)] \end{bmatrix} = \begin{bmatrix} \hat{s}(t|t) \\ \vdots \\ \hat{s}(t-p+1|t) \end{bmatrix}$$

and hence the last variable comprising the estimated regression vector at instant t can be used as a smoothed estimate of the signal at instant $t - p + 1$.

Since the mean square error $E[(s(\tau) - \hat{s}(\tau|t))^2]$ is a nonincreasing function of the smoothing lag $t - \tau$ (it is strictly decreasing if $\sigma_v^2 > 0$) one may be interested in obtaining a smoother with a lag q greater than $p - 1$. We note that this task can be easily accomplished by taking $\varphi(t) = [s(t), s(t-1), \dots, s(t-p), s(t-p-1), \dots, s(t-q)]^T$, $q \geq p$ and redefining the variables $A(t)$ and c in (4) appropriately.

Finally, we note that in the case where $d(t) = 1$, i.e., where the sample is called in question by the outlier detection routine, the quantity $\hat{s}(t|t+q) = E[s(t)|Y(t+q)] \cong \hat{\varphi}_{q+1}(t+q|t+q)$ can be regarded as the best (in the mean square sense) *reconstruction* of $s(t)$ based on the data available at instant $t+q$.

3 Experimental results

Results of two numerical tests carried out for real audio signals show how the proposed algorithm works in practice. In the first experiment a segment of a noiseless J.S. Bach recording (baroque orchestra with harpsichord - c.f. Fig.1a) was corrupted with three kinds of artificially generated disturbances: white Gaussian noise (Fig.1b), positive noise pulses of a constant magnitude (Fig.1d) and a mixture of impulsive and background noise (Fig.1f). The results yielded by the EKF algorithm, shown in figures 1c, 1e and 1g, respectively, are quite impressive. In particular, all noise pulses were correctly identified as outliers and replaced with reconstructed samples.

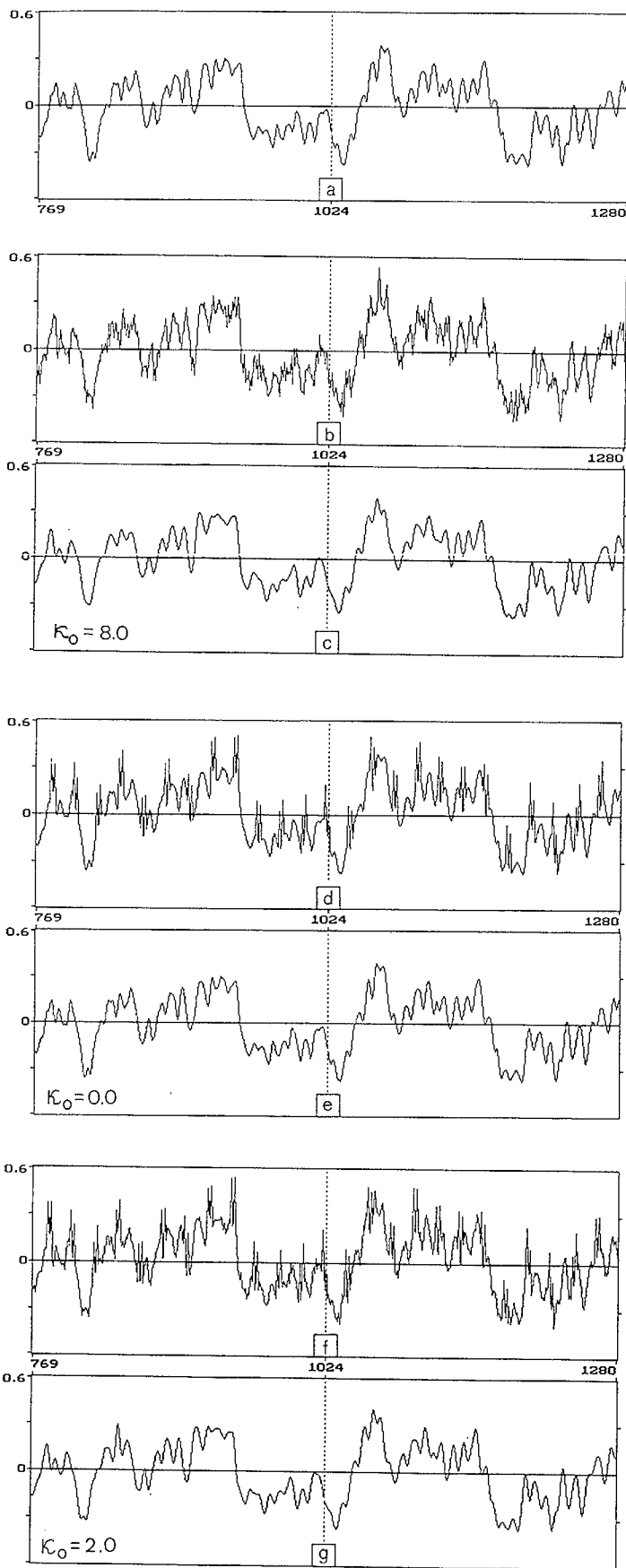


Fig.1 A segment of a noiseless audio signal (a) and its corrupted versions obtained by adding white background noise (b), noise pulses (d) and a mixture of impulsive and background noise (f). The results yielded by the EKF algorithm are shown in figures (c), (e) and (g), respectively.

In the second experiment a noisy segment of Ch. Gounod's song "Ave Maria" (symphonic orchestra - c.f. Fig.2a) played back from an old 78-r/min recording was analysed. The proposed method seems to have dealt favourably with all kinds of disturbances present in the recording such as background noise, "clicks" (noise pulses), "flats" (distortions introduced by the digital amplifier) and "scratches" (large artifacts caused by the local groove damage) - see Fig.2b.

In both experiments signals were sampled at the rate of 24kHz using the 16 bit A/D converter. The adopted order of autoregression was $p = 8$ and the memory-controlling parameter ξ was set to 0.00001.

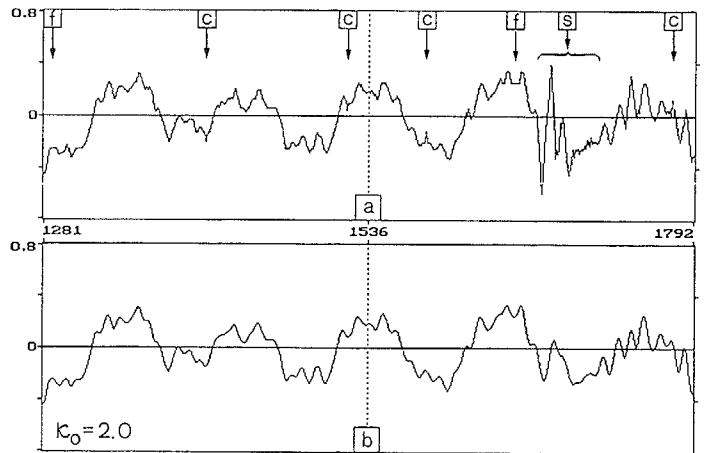


Fig.2 A segment of a noisy audio signal from an old 78-r/min phonograph record (a) and its restored version (b). Arrows mark the most significant disturbances such as "clicks" \square , "flats" \square and "scratches" \square .

References

- [1] B.D.O. Anderson and J.B. Moore, "Optimal filtering", Prentice-Hall, 1979.
- [2] T. Bohlin, "Four cases of identification of changing systems", System Identification: Advances and Case Studies, P.K. Mahra, D.G. Lainiotis, eds., pp. 441-518, 1976.
- [3] A.J.E.M. Janssen et al., "Adaptive interpolation of discrete-time signals that can be modelled as autoregressive processes", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 317-330, 1986.
- [4] R.H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations", Technometrics, vol. 22, pp. 389-395, 1980.
- [5] M. Niedźwiecki, "Statistical reconstruction of multivariate time series", IEEE Trans. Signal Processing, vol. 41, pp. 451-457, 1993.
- [6] M. Niedźwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes", IEEE Trans. Auto. Contr., vol. 35, pp. 329-334, 1990.
- [7] S.V. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings", J. Audio Eng. Soc., vol. 40, pp. 791-801, 1992.