BLIND IDENTIFICATION OF HAMMERSTEIN NONLINEAR DISTORTION MODELS

Guillaume Picard

Institut National de l'Audiovisuel
4, avenue de l'Europe
94366 Bry-sur-Marne, France
`gpicard` at `ina.fr`

Olivier Cappé

Centre National de la Recherche Scientifique
LTCI / GET-ENST, 46 rue Barrault
75634 Paris cedex 13, France
`cappe` at `tsi.enst.fr`

ABSTRACT

Compensation of nonlinear distortions is an issue of importance for the restoration of degraded audio material. It however remains a very challenging task, especially in cases when only a single instance of the degraded audio signal is available. Compared to other sources of distortion such as additive noise (hiss) or signal gaps, even the fundamental limits achievable by the restoration are yet unknown.

In this contribution, we consider a particular distortion model of the Hammerstein type (instantaneous nonlinear distortion followed by an all pole filter) for which only the output signal is observed. We argue that the tasks of identifying the distortion model and restoring the signal should be handled separately and focus on the former one. The proposed method estimates the distortion model from a large number of signal frames using a sub-optimal iterative framework.

1. INTRODUCTION

Identification and compensation of nonlinear distortions is a recurrent topic in signal processing, particularly for audio applications. The most challenging situation is the one faced when trying to restore degraded archived audio material where the characteristics of both the original signal and those of the source of distortion (amplifier, recording media, etc.) are usually unknown.

The idealized model in which the unknown degraded signal is a (strict sense) white noise has been extensively studied in mainstream signal processing and automatic control and some results and methods are available (often with further simplifying assumptions concerning the nature of the unknown distortion) [1, 2]. In audio however, it is not clear how such methods could be of any use except in situations where the source of distortion is available and may be excited with arbitrary signals. At present, the most impressive results demonstrated in settings closer to actual audio scenarios have been obtained using computer intensive numerical Bayesian methods based Markov Chain Monte Carlo (MCMC) simulations [3].

In this contribution, we follow a middle path between these two options by restricting to signal and distortion models which are sufficiently simple to guarantee that they are indeed identifiable by robust methods from the sole observation of the distorted output. On the other hand, we do not want to assume that the input signal is white. To cope with the fact that the spectrum of the input signal is mostly unknown we will take profit of the observation that, when considered on a large time scale, audio signals are non stationary and thus displays a variety of spectral contents, and most importantly, of signal levels.

2. MODEL AND HYPOTHESES

A first observation is that instantaneous distortion models are obviously not appropriate for audio except in very specific situations (such as digital clipping). On the other hand, models based on general nonlinear expansions such as Volterra series [1] are clearly too general (to allow identifiability) and cannot cope with the type of time dependence observed in audio unless one uses a large number of lagged terms in the expansion (although they may be usefull for high quality devices where the distortion effect is quasi instantaneous [4]).

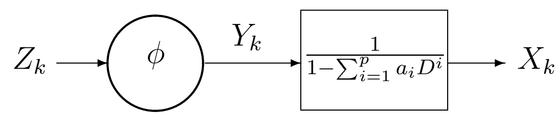

Figure 1: *Distortion model.*

A model that is reasonable in many situation given our knowledge of the physical behavior of some typical distorting devices is the Hammerstein model shown in figure 1. In this model, the signal first undergoes a nonlinear instantaneous transformation by a function ϕ followed by an all pole filtering (here we used the notation D to denote the delay operator). Our observation is the signal X_k and both the original signal Z_k and the intermediate one Y_k are unavailable. Although we do not describe these experiments here (for reason of space), we first fitted the Hammerstein model to some input-output (ie. using test signals) measurements of actual audio devices (a tube amplifier and magnetic recorder). The results showed that the distortion effect observed for large amplitude signals was satisfyly modeled using a Hammerstein model with p (AR order) about 10-15 and $d = 7$ or 9 (see Section 3 for the meaning of d).

The next step consists in dissociating the model identification and signal restoration steps. The reason for this has to do with the type of functions ϕ that are typically encountered which exhibit a strong saturation effect for high amplitudes. If we consider the extreme saturation curve which corresponds to hard clipping, it is clear that the recovery of the input signal is a very difficult and ambiguous task (often referred to as an ill-posed inverse problem) which can only be tackled using strong prior assumptions on the input signal. A very good solution in this setting would be to use MCMC methods on small frames of signal where the signal may be assumed to be stationary with an AR prior [5]. On the other hand, estimation of the clipping curve is a much simpler problem. Yet, it is a problem that can only be solved by looking at large portions of the signal so as to actually observe clipping. In this contribution we focus on the identification side of the problem and consider an approach which effectively profits from the observation of a large section of signal (typically several seconds or more) on which the stationary assumption cannot hold.

The idealized model that we will be used here is that when splitted into n signal frames of length t, which we denote by X_1^i, ..., X_t^i for $i = 1, \ldots, n$, the output signals may be viewed as the response to n independent input Gaussian signals $Z_1^i, \ldots, Z_t^i$ for $i = 1, \ldots, n$ with identical power spectral density (psd) $f(\omega)$ known up to a scale σ_i (which is different for each signal index $i = 1, \ldots, n$ and unknown). This is a very crude model which pretends that the only source of non-stationarity in audio signals has to do with level changes and not with spectral changes (this issue will be discussed further in section 4).

It is clear that at this point of simplification, we could estimate the unknown parameters $a_1, \ldots, a_p$ (the AR coefficients), ϕ (or at least a parameterized version of it) and $\sigma_1, \ldots, \sigma_n$ (input levels) using a maximum likelihood approach. To do so however, we would have to parameterize ϕ^{-1} rather than ϕ (because of the change of variable formula) which would be a very bad move granted the type of saturation curve that we are expecting for ϕ (that is ϕ^{-1} is a function whose derivative is arbitrary large on the borders of the range of ϕ). We will thus use a sub-optimal approach based only on the empirical second order properties and marginal distribution of $(Z_1^i, \ldots, Z_t^i)$ for $i = 1, \ldots, n$.

For the estimation to be feasible we will need the additional assumptions that ϕ may be parametrized as finite order (d) polynomial and that it is monotonic. The reason for this second assumption is that

1. $(a_i)_{1 \le i \le p}$ is identifiable from second order properties only when the corresponding causal filter is stable and $f(\omega)$ is known. If ϕ is a polynomial of order d, the $(a_i)_{1 \le i \le p}$ are identifiable for all ϕ in this class when at least $d + 1$ input signals with different scales σ_i are available.
2. ϕ is then identifiable from the marginal distributions up to an unknown scale if ϕ is monotonic (ϕ is never identifiable unless assumed monotonic). The scales σ_i ($i = 1, \ldots, n$) are then identifiable up to an unknown global scale.

3. PARAMETER IDENTIFICATION

In the following we assume that $\phi(x) = \sum_{j=1}^{d} \frac{\alpha_j}{j!} x^j$. There is unfortunately no way of constraining simply ϕ to be monotonic as required. We will however use only odd order monomials in order to force ϕ to be antisymmetric which is highly sensible for audio applications.

The proposed parameter estimation procedure corresponds to a suboptimal approach which iterates between two fundamental steps : one in which the all-pole parameters are estimated from the second order properties of the observed signals $(X_k^i)_{1 \le k \le t}$ and second step in which the scales and nonlinear transformation parameters are estimated from the marginal distribution of the whitened output signals.

3.1. AR parameters

A fundamental Hilbert space result known as Mehler's formula states that if Z_1 and Z_2 are jointly Gaussian variables with common unit variance and correlation coefficient ρ, $\mathrm{E}[H_k(Z_1)H_l(Z_2)] = k!\rho^k$ if $k = l$ (and equals 0 otherwise), where H_k denote the family of Hermite polynomials which are orthogonal with respect to the Gaussian measure $\mathcal{N}(0, 1; z) = 1/\sqrt{2\pi} \exp(-z^2/2)$ on $\mathbb{R}$ [6]. We may thus compute the autocovariance function of a signal $(Z_k)_{1 \le k \le t}$ distorted by ϕ as

$$\mathrm{E}(Y_l Y_{l+k}) = \sum_{j=1}^{d} \frac{\left(c_j^\sigma\right)^2}{j!} \rho(k)^j \tag{1}$$

where $\rho(k)$ is the autocorrelation function of the input and σ^2 its variance and $c_1^\sigma, \ldots c_d^\sigma$ are the coefficients of $\phi(\sigma z)$ in the basis of the Hermite polynomials $H_1(z), \ldots H_d(z)$. These can be obtained by back-substitution from the monomial coefficients α_1, ..., α_d when the order d is fixed. For instance, for $d = 5$ these are given by

$$\begin{pmatrix} c_1^\sigma \\ c_3^\sigma \\ c_5^\sigma \end{pmatrix} = \begin{pmatrix} \sigma & 1/2\sigma^3 & 1/8\sigma^5 \\ 0 & \sigma^3 & 1/2\sigma^5 \\ 0 & 0 & \sigma^5 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_3 \\ \alpha_5 \end{pmatrix}$$

The above method is used to compute the autocovariance functions, or equivalently, the power spectral density f_Y^i of the distorted signals $(Y_k^i)_{1 \le k \le t}$ ($i = 1, \ldots, n$) from the common average autocorrelation function ρ and the input scales $\sigma_1, \ldots, \sigma_n$. It can be shown that the joint Whittle criterion for the observed signals is optimized for the parameter vector $a_1, \ldots, a_p$ which solves the Yule-Walker like system

$$\sum_{i=1}^{n} \begin{pmatrix} r^i(0) & r^i(1) & \cdots & r^i(p-1) \\ r^i(1) & r^i(0) & \cdots & r^i(p-2) \\ \vdots & & & \\ r^i(p-1) & r^i(p-2) & \cdots & r^i(0) \end{pmatrix} \times \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} r^i(1) \\ r^i(2) \\ \vdots \\ r^i(p) \end{pmatrix}$$

where r^i are the Fourier coefficients of I_X^i / f_Y^i, with $I_X^i(\omega) = |\sum_{k=1}^{t} X_k^i \, \mathrm{e}^{-j\omega k}|^2/t$ denoting the periodogram of the ith signal.

In practice, the r^i are approximated by computing the periodogram I_X^i at Fourier frequencies $\omega_k = 2\pi k/t$, dividing by $f_Y^i(\omega_k)$ and performing inverse FFT.

3.2. Distortion and scales

For estimating the parameters of ϕ ($\alpha_1, \ldots, \alpha_d$) and the signal scales $\sigma_1, \ldots, \sigma_n$, we rely solely on the marginal distribution of the observable signal and use a least-square criterion. The ideas used here are inspired by [2] and [7].

Let $\hat{Y}_k^i = X_k^i - \sum_{j=1}^{p} a_j X_{k-j}^i$ denote the ith whitened signal given the current estimate of the all-pole parameters. From this we compute the order statistics

$$\hat{Y}_{(1)}^i < \hat{Y}_{(2)}^i < \cdots < \hat{Y}_{(t)}^i$$

for each signal $i = 1, \ldots, n$ (note that in practice, we only have $t - p$ samples of the whitened signal but we will ignore this fact to keep the notations simple). It turns out that we have a rather good idea of the statistical behavior of these order statistics thanks to the following Lemma.

Lemma [6] For a Gaussian $\mathcal{N}(0,1)$ iid signal frame $Z_1,\ldots,Z_t$, the order statistics $Y_{(1)},\ldots,Y_{(t)}$ for the distorted signal $Y_k = \phi(Z_k)$ satisfies when $t, j \nearrow \infty$ such that $j/t \to \theta \in [0,1]$,

$$\sqrt{t}(Y_{(j)} - \phi(Q(\theta))) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \dot{\phi}(Q(\theta))^2 \dot{Q}(\theta)^2 \theta(1-\theta)\right)$$

where the symbol above refers to convergence in distribution, the dot correspond to the first derivative and Q denotes the inverse of the Gaussian distribution function, that is the function $(0,1) \to (-\infty,\infty)$ such that

$$\int_{-\infty}^{Q(\theta)} 1/\sqrt{2\pi}\exp(-z^2/2)dz = \theta.$$

In other words, $Y_{(j)}$ is approximately Gaussian with mean $\phi(Q(j/t))$ and known variance (except for the very first or last order statistics). The use of the variance information given by the above lemma is somewhat problematic since it depends on the derivative $\dot{\phi}$ of ϕ. Although the behavior of this variance would deserve more comments, we will here consider that the variations of this variance with θ are sufficiently small to be ignored. This suggest the use of a nonlinear least squares procedure with objective function

$$\sum_{i=1}^{n} \left\| \underbrace{\begin{pmatrix} \hat{Y}^i_{(1)} \\ \hat{Y}^i_{(2)} \\ \vdots \\ \hat{Y}^i_{(t)} \end{pmatrix}}_{\hat{\mathbf{Y}}^i} - \underbrace{\begin{pmatrix} q_1 & \ldots & q_1^d/d! \\ q_2 & \ldots & q_2^d/d! \\ \vdots & & \vdots \\ q_t & \ldots & q_t^d/d! \end{pmatrix}}_{\mathbf{Q}} \underbrace{\mathrm{diag}\begin{pmatrix} \sigma_i \\ \sigma_i^2 \\ \vdots \\ \sigma_i^d \end{pmatrix}}_{\mathbf{D}(\sigma_i)} \underbrace{\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix}}_{\boldsymbol{\alpha}} \right\|^2$$

where $q_i = Q(i/(t+1))$ (the division by $t+1$ rather than t avoids problems for the last statistic) and diag denotes the diagonal matrix with given diagonal elements. With the notations introduced above, the optimization problem to be solved consists in minimizing

$$\sum_{i=1}^{n} \left\| \hat{\mathbf{Y}}^i - \mathbf{Q}\mathbf{D}(\sigma_i)\boldsymbol{\alpha} \right\|^2 \qquad (2)$$

with respect to the unknown vector $\boldsymbol{\alpha}$ and the scales $\sigma_1,\ldots,\sigma_n$. The criterion is quadratic with respect to $\boldsymbol{\alpha}$ (the parameters of the transformation ϕ) but highly nonlinear with respect to the scale parameters.

Optimization of this criterion has to be done carefully. In particular, the complete relaxation procedure which optimize with respect to $\boldsymbol{\alpha}$ and then all the σ_i separately was found to be overly sensitive to initialization. Robust identification results have been obtained by applying a Quasi-Newton optimization procedure to the reduced criterion obtained by replacing $\boldsymbol{\alpha}$ in (2) by its least-squares estimate. The resulting criterion only depends on the signal scales $\sigma_1,\ldots,\sigma_n$ which have to be somehow initialized. In the following, we use the estimated standard deviation of the ith output signal as an initial guess of σ_i (which would be correct if there was no distortion). Since, the scales $\sigma_1,\ldots,\sigma_n$ and the distortion ϕ are only identifiable up to a shared unknown constant, σ_1 is kept fixed during the optimization. In the simulations of Section 4 below, we used the knowledge of the standard deviation of the actual first input signal to recover the correct scaling for ϕ (in order to allow comparisons) but, in general, ϕ will only be determined up to a constant factor.

4. EXPERIMENTS

4.1. Simulated Signals

We first consider as input signals $n = 10$ AR(1) sequences (with AR parameter 0.9) of length $t = 1024$ with linearly increasing standard deviations such that $\sigma_{10}/\sigma_1 = 3.5$. The distortion ϕ is a fifth order odd polynomial ($\alpha_1 = 1, \alpha_3 = -2.1, \alpha_5 = 3.6$) displayed in the left plot of figure 2 (bold curve). The all-pole filter is of order $p = 2$ and its frequency response is shown in the right plot of figure 2 (bold curve).

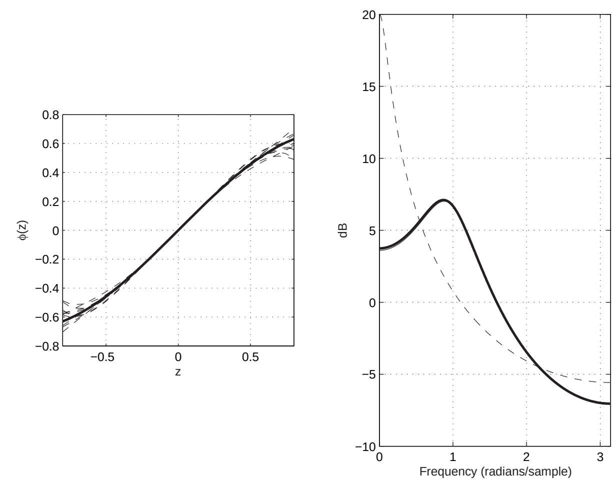

Figure 2: **Left:** Actual distortion curve (bold curve) and ten estimated distortion curves; **Right:** Common psd of the input sequences (light dashed curve), actual (bold curve) and estimated frequency response of the all-pole filter.

We used two iterations of the our two steps procedure starting with the estimation of the distortion parameters and input scales, the former being the only parameters which have to be initialized (see Section 3.2). In this simulated example, the procedure was found to be very robust with respect to its initialization and converges quickly to a solution which is indiscernible from the actual distortion source for the all-pole part (right plot in figure 2). The distortion curve is also fairly well estimated although the estimation errors tends to be larger at the borders of the interval (there are 10 superimposed estimated curves on the left plot of figure 2). Note that the average number of input samples with magnitude larger than 0.8 for all 10 independent draws of the input signal is 7 (out of 10×1024). Generally these samples are found only in the last signal, which has the larger standard deviation. This clearly means that the behavior of the estimated curve above 0.8 and below -0.8 has more to do with polynomial extrapolation than with actual estimation. In the range of the input signal however, the estimation of the distortion curve is quite good.

4.2. Audio signal

We now consider a musical record of vocal track and piano with ambient noise, synthetically distorted with the same nonlinear model as in the previous experiment. Now the maximal magnitude of the signal is 1.5 and there are about 0.9% of the samples with magnitude larger than 0.8. We consider $n = 340$ signal frames of length

$t = 1024$ corresponding to 7.9 seconds of signal. The evolution of the RMS magnitude of the signal for 70 consecutive frames taken in the middle of the musical extract is shown on figure 3 (with the circles). The "common" psd $f(\omega)$ is computed by averaging the frequency content of all input frames in the cepstral representation (using 32 cepstral coefficients). Cepstral averaging is recommended since it is both insensitive to the scale (if the first cepstral coefficients is discarded) and provides a smoothed version of the spectrums. The obtained common psd is displayed on figure 4 (right plot, dashed curve).

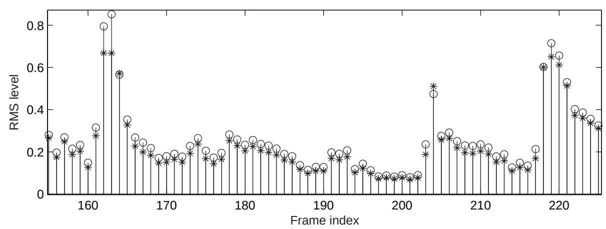

Figure 3: **Circles:** RMS levels of input signal frames; **Stars:** estimated levels.

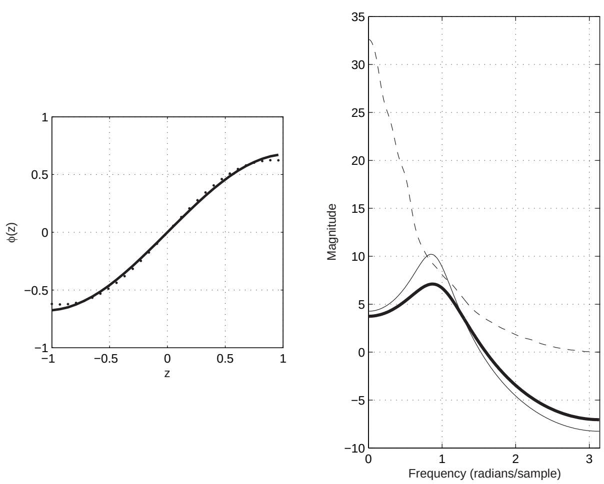

Figure 4: **Left:** Actual distortion curve (bold curve) and estimated distortion curve (dot curve); **Right:** Average psd of the input sequences (light dashed curve), actual (bold curve) and estimated frequency response of the all-pole filter.

The distortion model (both the static transfer function ϕ and the all-pole filter) are the same as in the previous example. We proceed as previously for the estimation, initializing the variances $\sigma_1, \ldots, \sigma_n$ to those of the output (distorted) signals.

From the results displayed on figure 4, we can first conclude that the instantaneous distortion is very well estimated (left plot in figure 4), especially granted that only about 0.7% of the input samples have a magnitude larger than 0.8. The second conclusion is that the estimation of the all-pole model (light curve in the right plot of figure 4) is correct somewhat less accurate.

A closer investigation reveals that the estimated all-pole filter is more sensitive than the estimated nonlinear curve to the spectral contents of the signals. In particular, the obtained results depends, to some extent, on the selected signal frames. However using a large section of the input signal pays and the estimated all-pole filter is both much more accurate and stable when using all 340 frames then when using smaller subsets of them (we first tried using about 35 frames which was less robust). Using more signal frames makes it possible to average out the spectral variations around the "common" psd (even though, the assumption that all signals share an identical psd is grossly wrong). The second obvious factor is that the all-pole filter is sensitive to the form of the average psd which is used by the identification algorithm. This corresponds to a strong characteristic of the problem which is already encountered for the blind deconvolution of *linear* effects [8]. This means that in practical situations, recovering the input signal up to an unknown linear filtering effect is about the best that we can hope for.

5. CONCLUSION

The work described above is based on the idea that robust identification of nonlinear distortion models should take profit of the observation of large sections of signals, and in particular, of the scale variations typical of audio signals. The experiments reported in the last section shows that the proposed method fulfills, at least, a significant part of this program. Possible improvements include refining the assumptions that are used for the input signals. In particular, one could use mixture models, as proposed in [9] for single sensor source separation, instead of a single common spectrum.

6. REFERENCES

[1] G. B. Giannakis and E. Serpedin, "A bibliography on nonlinear system identification," *Signal Processing*, vol. 81, no. 3, pp. 533–580, 2001.

[2] W. Greblicki and M. Pawlak, "Nonparametric identification of Hammerstein systems," *IEEE Trans. Information Theory*, vol. 35, 1989.

[3] P. Troughton and S. Godsill, "MCMC methods for restoration of nonlinearly distorted autoregressive signals," *IEEE Trans. Speech Audio Proc.*, vol. 81, no. 1, 2001.

[4] H. Schurer, A. G. J. Nijmeijer, M. A. Boer, C. H. Slump, and O. Herrmann, "Identification and compensation of the electrodynamic transducer nonlinearities," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, 1997.

[5] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration - A Statistical Model-Based Approach*. Springer-Verlag, 1998.

[6] A. Stuart and J. K. Ord, Eds., *Kendall's advanced theory of statistics*, 5th ed. London: Charles Griffin & Co, 1987, vol. 1 (distribution theory).

[7] R. Balchandran and R. Mammone, "Non-parametric estimation and correction of non-linear distortion in speech systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, 1998.

[8] T. G. Stockham, T. M. Canon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, no. 4, pp. 678–692, 1975.

[9] L. Benaroya and F. Bimbot, "Wiener based source separation with HMM/GMM using a single sensor," in *Proc. Int. Conf. Ind. Comp. Anal. and Blind Signal Sep.*, 2003.