BAYESIAN SUPPRESSION OF MEMORYLESS NONLINEAR AUDIO DISTORTION

H. T. Carvalho, F. R. Ávila†, L. W. P. Biscainho**

* Federal University of Rio de Janeiro
SMT/PEE/COPPE
21941-972 Rio de Janeiro, Brazil

† Rio de Janeiro State University
PROSAICO/PEL
20550-900 Rio de Janeiro, Brazil

ABSTRACT

Even if nonlinear distortion may be deliberately applied to audio signals for esthetic or technical reasons, it is common to hear annoying defects in accidentally saturated or amateurishly processed audio—which calls for some means to automatically undo the impairment. This paper proposes an algorithm to blindly identify the nonlinear distortion suffered by an audio signal and reconstruct its original form. Designed to deal with memoryless impairments, the model adopted for the nonlinear distortion is a curve composed of an invertible sequence of linear segments, capable of following the typical shape of compressed audio, and whose parameters are easily interpretable and thus constrainable. The solution builds on the posterior statistical distribution of the curve parameters given the degraded signal, and yields perceptually impressive results for real signals distorted by arbitrary curves.

***Index Terms*—** Nonlinear distortion, Bayesian signal processing, blind system identification, audio processing

1. INTRODUCTION

A modified signal is said to be distorted whenever its shape changes (magnitude scaling and time shifting are not distortions). Any distortion which creates new frequency content is nonlinear, since linear distortion can always be modeled by magnitude and phase changes imposed on the original frequency content of the signal.

Guitar pedal effects are an artistically motivated example of purposeful introduction of nonlinear distortions in audio. Scaling down to less radical procedures, compression is an essential resource in recording studios for either esthetic (e.g. enhancement of drum sound) or technical (e.g. signal-to-noise ratio improvement) reasons. On the other hand, indiscriminate over-compression and accidental saturation due to exceeding recording level can produce signals impaired by audibly annoying defects. Automatic tools to undo the underlying distortions would be useful in such circumstances.

A well-explored topic in control engineering [1], nonlinear system identification is still relatively under-researched in audio processing. Earlier attempts to tackle nonlinear audio distortion focused on particular applications where the distortion curve has a well-defined algebraic form, such as in movie soundtracks [2] or reproduction by horn loudspeakers [3]. Alternative solutions to deal with more general curve shapes appeared in [4] and [5].

The goal of the algorithm proposed herein is to estimate a generic memoryless distortion in the form of a piecewise linear model with unknown slopes. The original audio signal is described by an autoregressive (AR) model [6] of predefined order with unknown coefficients, excited by Gaussian white noise of unknown variance. By specifying an adequate prior for both signal and distortion parameters, one finds a formula for their posterior distribution, given the distorted signal. Then, a strategy based on Markov-chain Monte Carlo (MCMC) which combines the Metropolis-Hastings algorithm with Gibbs sampling [7] is employed to draw samples from the parameters' posterior, from which the piecewise linear curve can be inferred. Tests of the proposed solution performed over artificially distorted real audio signals show that high-quality estimates for smooth distorting functions are reached with high probability. Even if not specialized, the method is applicable to speech signals.

This paper continues in Section 2 with the description of both signal and distortion models. In Section 3, the parameters' posterior distribution is deduced. Section 4 explains the design of the MCMC algorithm that leads to the parameters estimation. Section 5 demonstrates by simulations the effectiveness of the proposed algorithm in undoing the targeted nonlinear distortions. Conclusions are drawn in Section 6.

2. SIGNAL AND DISTORTION MODELS

Figure 1 shows the overall model adopted for the memoryless nonlinear distortion of an audio signal. The leftmost block represents the generation of the original audio signal in the form of an autoregressive model in which the all-pole filter $A(z)$ is excited by the Gaussian white noise sequence e_n; the rightmost block represents the distortion curve. The two most important assumptions of the model are that the AR model is constant only over short sections of signal, and that the nonlinear curve is the same throughout the signal. Accord-

The authors thank CAPES, CNPq and FAPERJ Brazilian agencies for funding this work.

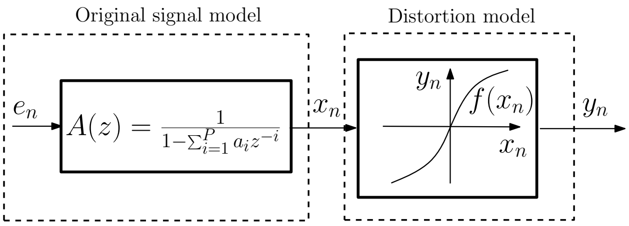

Fig. 1: Audio signal corrupted by memoryless nonlinearity.

ingly, in order to analyze the signal statistical behavior, one splits the full N-sample signal into B frames of L samples. Typically, audio signals are considered stationary over a period of around 20 ms; thus, the choice of $L = 1000$ was adequate for the sampling rate of 44.1 kHz employed in this work. The original signal sequence x_n and the distorted signal sequence y_n are stacked into vectors $\mathbf{x}$ and $\mathbf{y}$, with frames in sub-vectors $\mathbf{x}_j$ and $\mathbf{y}_j$, respectively, for $j \in \{1, \ldots, B\}$. Analogous definitions apply to $\mathbf{e}$ and $\mathbf{e}_j$.

2.1. Statistical Model for the Original Signal

In [6] one finds a usual statistical description for audio signals, briefly reviewed here for the sake of completeness.

For each block $j \in \{1, \ldots, B\}$, the excitation vector $\mathbf{e}_j$ is filled by samples of Gaussian white noise with variance $\sigma_{e_j}^2$. Then, $\mathbf{x}_j$ is calculated through the order-P AR-model recursion $x_n = e_n + a_1 x_{n-1} + a_2 x_{n-2} + \ldots + a_P x_{n-P}$.

Model parameters $\mathbf{a}_j$ and $\sigma_{e_j}^2$ can be independently sampled from their respective prior distributions, chosen like in [6]: for $\sigma_{e_j}^2$, an Inverted-Gamma distribution with very small parameters α and β, which approximates its Jeffrey's non-informative prior; and for $\mathbf{a}_j$, a $\mathbf{0}$-mean Gaussian distribution with covariance matrix $\sigma_a^2 \mathbf{I}_P$, where $\mathbf{I}_P$ denotes the $P \times P$ identity matrix and σ_a^2 is sufficiently high w.r.t the usual values of $\mathbf{a}$ to turn the prior non-informative.

In order to simplify the notation for the subsequent analysis, multivariate variables $\mathbf{a}$ and $\boldsymbol{\sigma}_e^2$ are formed by concatenating all $\mathbf{a}_j$ and $\sigma_{e_j}^2$, respectively, in their temporal order.

2.2. Model for the Distortion

In Figure 1, the degraded signal $\mathbf{y}_j$ is produced by inputting the original signal $\mathbf{x}_j$ to the nonlinear curve $f(x_n)$. The option for parameterizing the inverse curve $g(y_n) = f^{-1}(y_n)$ makes the calculation of the likelihood function simpler. This function is assumed to be: anti-symmetric, which is reasonable in audio distortions; monotonically increasing to assure invertibility; and identity in the vicinity of the origin to prevent re-scaling the magnitude of the original signal.

Figure 2 illustrates the model for the inverse curve $g(y_n)$. Without loss of generality, the degraded signal is assumed to lie within the interval $[-1, 1]$. The range from 0 to 1 (and likewise its negative counterpart) is split into M contiguous intervals of length $\Delta y = 1/(M+1)$. Each interval I_i, $i \in \{-M, -1, 1, \ldots, M\}$, is described by an affine function with slope $m_i = \tan(\theta_i)$, with $m_i > 0$ (monotonic increase) and $m_i = m_{-i}$ (anti-symmetry). Starting from $m_1 = 1$ (identity around the origin) at the positive side, the first point of each subsequent segment coincides with the last point of the former (curve continuity). Referring to Figure 2, any point y^* belonging to the i-th interval is mapped to a point x^* by

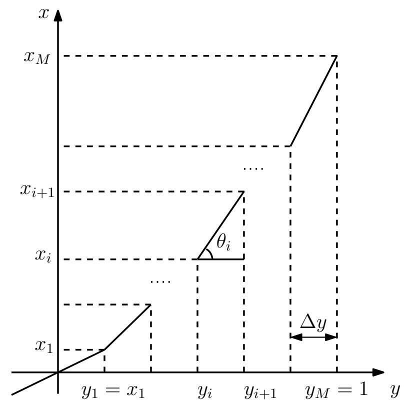

Fig. 2: Piecewise linear model.

$$x^* = g_i(y^*) = x_i + \operatorname{sign}(y^*)(y^* - y_i)m_i, \qquad (1)$$

where $x_0 = 0$, and $\operatorname{sign}(y^*)$ denotes the sign of y^*. Noting by induction that $x_i = \operatorname{sign}(y_i)\Delta y \sum_{i=0}^{i-1} m_i$, one can represent the undistorted signal vector as

$$\mathbf{x} = \mathbf{u} + \mathbf{R}\mathbf{m}, \qquad (2)$$

where $\mathbf{m} = [m_2\, m_3\, \cdots\, m_M]^T$, and both vector $\mathbf{u}$ and matrix $\mathbf{R}$ can be computed from elements of $\mathbf{y}$ and known constants.

Parameter $\mathbf{m}$ can be sampled from a prior distribution. Recalling restrictions $m_1 = 1$ and $m_2, m_3, \ldots, m_M > 0$, one can denote their admissible set by Ω. Since there is no information about the angular-coefficient values, one can assume a Gaussian prior restricted to this region:

$$p(\mathbf{m}) \propto \mathrm{e}^{-\frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m}} \mathbf{1}_\Omega(\mathbf{m}), \qquad (3)$$

where $\mathbf{1}_\Omega(\mathbf{m})$ is the indicator function of Ω and variance σ_m^2 is made sufficiently high to turn the priori non-informative.

3. PARAMETERS' POSTERIOR DISTRIBUTION

Since the goal of the identification procedure is to find the most likely model parameters based on the degraded signal $\mathbf{y}$, the first step in designing the algorithm is to build the posterior distribution $p(\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2|\mathbf{y})$, which according to Bayes' Theorem is the product of the likelihood function by the parameters' prior distribution, up to a multiplicative constant:

$$p(\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2) p(\mathbf{m}) p(\mathbf{a}, \boldsymbol{\sigma}_e^2). \qquad (4)$$

The likelihood $p(\mathbf{y}|\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2)$ can be calculated by exploring the Gaussianity of e_n and invoking the formula for random variable transformation.

First, a single block $\mathbf{x}_j$ of the original signal is considered. To keep notation light, subscript j and conditioning on $\mathbf{a}_j$ and $\sigma_{e_j}^2$ will be omitted for a moment. Since $x_n = e_n + \sum_{i=1}^{P} a_i x_{n-i}$,

$$p(x_n|x_{n-1}, \ldots, x_1) = p_{e_n}\left(x_n - \sum_{i=1}^{P} a_i x_{n-i}\right), \quad (5)$$

for $n \geq P+1$. Using induction and (5), one can show that

$$p(x_L, \ldots, x_{P+1}|x_P, \ldots, x_1) = \prod_{n=P+1}^{L} p(x_n|x_{n-1}, \ldots, x_1) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{\frac{L-P}{2}} \mathrm{e}^{-\frac{1}{2\sigma_e^2}\mathbf{e}^T\mathbf{e}}; \quad (6)$$

the excitation vector for the current frame $\mathbf{e} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is constructed using elements of $\mathbf{a}$. This distribution is dependent on the first P samples of the current frame, but since in general $L \gg P$, their influence on the whole block can be disregarded [6], so that

$$p(x_L, \ldots, x_1) \approx p(x_L, \ldots, x_{P+1}|x_P, \ldots, x_1). \quad (7)$$

Combining together the information of the B blocks and bringing back the omitted variables, one can write

$$p(\mathbf{x}|\mathbf{a}, \boldsymbol{\sigma}_e^2) = \prod_{j=1}^{B} p(\mathbf{x}_j|\mathbf{a}_j, \sigma_{e_j}^2) \approx \prod_{j=1}^{B}\left(\frac{1}{2\pi\sigma_{e_j}^2}\right)^{\frac{L-P}{2}} \mathrm{e}^{-\frac{1}{2\sigma_{e_j}^2}\mathbf{x}_j^T\mathbf{A}_j^T\mathbf{A}_j\mathbf{x}_j}. \quad (8)$$

The "approximation" will be written as an equality hereafter.

Now, using (2), the desired likelihood can be computed:

$$p(\mathbf{y}|\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2) = \prod_{i=1}^{M} m_i^{N_i} \times \prod_{j=1}^{B}\left(\frac{1}{2\pi\sigma_{e_j}^2}\right)^{\frac{L-P}{2}} \mathrm{e}^{-\frac{1}{2\sigma_{e_j}^2}(\mathbf{u}_j+\mathbf{R}_j\mathbf{m})^T\mathbf{A}_j^T\mathbf{A}_j(\mathbf{u}_j+\mathbf{R}_j\mathbf{m})}, \quad (9)$$

where N_i is the number of signal samples affected by m_i, and $\mathbf{u}_j$ and $\mathbf{R}_j$ are the blockwise counterparts of $\mathbf{u}$ and $\mathbf{R}$.

At last, by replacing (9) and the prior distributions of $\mathbf{m}$, $\mathbf{a}$ and $\boldsymbol{\sigma}_e^2$ in (4), one can obtain an explicit expression for the posterior distribution:

$$p(\mathbf{m}, \mathbf{a}, \boldsymbol{\sigma}_e^2|\mathbf{y}) \propto \left[\prod_{i=1}^{M} m_i^{N_i} \prod_{j=1}^{B} \frac{1}{(\sigma_{e_j}^2)^{\frac{L-P}{2}}} \mathrm{e}^{-\frac{1}{2\sigma_{e_j}^2}(\mathbf{u}_j+\mathbf{R}_j\mathbf{m})^T\mathbf{A}_j^T\mathbf{A}_j(\mathbf{u}_j+\mathbf{R}_j\mathbf{m})}\right] \times \left[\mathrm{e}^{-\frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m}}\mathbf{1}_\Omega(\mathbf{m})\right]\left[\mathrm{e}^{-\frac{1}{2\sigma_a^2}\mathbf{a}^T\mathbf{a}}\right]\left[\prod_{j=1}^{B}\frac{1}{(\sigma_{e_j}^2)^{\alpha+1}}\mathrm{e}^{-\frac{\beta}{\sigma_{e_j}^2}}\right]. \quad (10)$$

4. POSTERIOR SIMULATION VIA MCMC

Even though $\mathbf{x}$ can be written as a linear function of parameters $\mathbf{m}$, the expression for the posterior is very complicated, and not guaranteed to be uni-modal—which means that no simple deterministic maximization algorithm can avoid being trapped on a local maximum. The solution adopted in this work is to explore the distribution by drawing samples of it using an MCMC algorithm, from which the unknown parameters are later inferred.

Starting from guesses $\mathbf{m}^{(0)}$ and $\mathbf{a}^{(0)}$, the Gibbs sampling algorithm summarized below sequentially draw samples from the total conditional distributions of the unknown variables.

Algorithm 1 Nonlinear model identification algorithm

```
1: (m^(0), a^(0)) ← Initialize from y
2: for i = 1 to # iterations do
3:     σ_e^2(i) ~ p(σ_e^2 | y, a^(i-1), m^(i-1))
4:     a^(i) ~ p(a | y, σ_e^2(i), m^(i-1))
5:     m^(i) ~ p(m | y, a^(i), σ_e^2(i))
6: end for
```

Sampling of AR-model parameters: The conditional distributions for $\boldsymbol{\sigma}_e^2$ and $\mathbf{a}$ are obtained by ignoring the portions of (10) unrelated to them, which results in Inverted-Gamma for each $\sigma_{e_j}^2$ and Gaussian for each $\mathbf{a}_j$—both easily samplable distributions.

Sampling of nonlinear distortion parameters: By ignoring the portions of (10) unrelated to $\mathbf{m}$, its conditional distribution is found to be

$$p(\mathbf{m}|\mathbf{y}, \mathbf{a}, \boldsymbol{\sigma}_e^2) \propto \prod_{i=1}^{M} m_i^{N_i} \times \prod_{j=1}^{B} \mathrm{e}^{-\frac{1}{2\sigma_{e_j}^2}(\mathbf{u}_j+\mathbf{R}_j\mathbf{m})^T\mathbf{A}_j^T\mathbf{A}_j(\mathbf{u}_j+\mathbf{R}_j\mathbf{m}) - \frac{1}{2\sigma_m^2}\mathbf{m}^T\mathbf{m}} \mathbf{1}_\Omega(\mathbf{m}). \quad (11)$$

Due to the first product and geometric restrictions both involving $\mathbf{m}$, this is not a well-known type of distribution. In order to draw samples from it, a strategy known as Metropolis within Gibbs [7] was adopted: instead of directly sampling from the posterior distribution of $\mathbf{m}$ during the Gibbs sampling, samples $\mathbf{m}^*$ are drawn from some proposal distribution $q(\mathbf{m}|\mathbf{m}^{(i-1)})$ and accepted as the next sample of $\mathbf{m}$, denoted as $\mathbf{m}^{(i)}$, with probability

$$\alpha = \min\left(1, \frac{p(\mathbf{m}^*|\mathbf{y}, \mathbf{a}^{(i)}, \boldsymbol{\sigma}_e^{2(i)})q(\mathbf{m}^{(i-1)}|\mathbf{m}^*)}{p(\mathbf{m}^{(i-1)}|\mathbf{y}, \mathbf{a}^{(i)}, \boldsymbol{\sigma}_e^{2(i)})q(\mathbf{m}^*|\mathbf{m}^{(i-1)})}\right). \quad (12)$$

The chosen proposal distribution was a Gaussian centered on $\mathbf{m}_0$ that maximizes $\xi(\mathbf{m}) = \ln p(\mathbf{m}|\mathbf{y}, \mathbf{a}, \boldsymbol{\sigma}_e^2)$, and with covariance matrix given by minus the second derivative of $\xi(\mathbf{m})$ computed at $\mathbf{m}_0$, denoted by $H(\mathbf{m}_0)$. In order to impose the geometric restrictions over $\mathbf{m}$, the support of the proposal distribution itself was restricted to the set Ω. This procedure is an improvement of the so-called Laplace Approximation [8], since it does not simply substitutes the target distribution by a Gaussian one, but there is some probability that such samples are rejected.

While matrix $H(\mathbf{m}_0)$ can be analytically computed, finding $\mathbf{m}_0$ is quite difficult and requires the use of some numerical approximation method. Since $\xi(\mathbf{m})$ is approximately quadratic, a few iterations of Newton's method [9] may suffice to that end.

5. RESULTS

The performance of the proposed algorithm was evaluated through tests where a real audio signal was artificially degraded by: (A) a piecewise linear distortion curve as in the model, in order to assess the method's accuracy and convergence; and (B) a more general smooth distortion curve, in order to assess the method's generality.

The signal chosen for the tests is a 3-s monophonic recording in PCM format, sampled at 44.1 kHz with 16-bit precision, where a flute plays a single note with vibrato. The simplicity of the signal allows one to perceive very clearly the distortion, even when it is not too severe.

The tests were run in a personal computer with a quad-core processor operating at 2.3 GHz clock, and 8 GB of RAM.

5.1. PIECEWISE LINEAR DISTORTION

The signal was artificially degraded by a piecewise linear distortion described by angular coefficients

$$[1/2 \quad 1/4 \quad 1/6 \quad 1/8]^T. \tag{13}$$

Since the proposed algorithm recovers the inverse function, i.e. the function which restore the original signal, the algorithm is expected to output the vector

$$\mathbf{m} = [2 \quad 4 \quad 6 \quad 8]^T. \tag{14}$$

In order to save processing time, considering that the nonlinear distortion is the same throughout the entire signal, a single block with $L = 1000$ samples in the region of maximum magnitude of the degraded signal, expected to provide information about all $\mathbf{m}$ entries, was processed. The estimation algorithm ran for 100 iterations, outputting the average of the last 50 samples for each parameter, after a burn-in period of 50 iterations. The algorithm initialization was performed by: estimating the order-P vector $\mathbf{a}^{(0)}$ from the corrupted signal $\mathbf{y}$ using a well known method [10]; and choosing $\mathbf{m}^{(0)}$ with ones in all its entries, meaning no distortion.

In Figure 3, where red squares mark their real values, one can verify the successful convergence of angular coefficients.

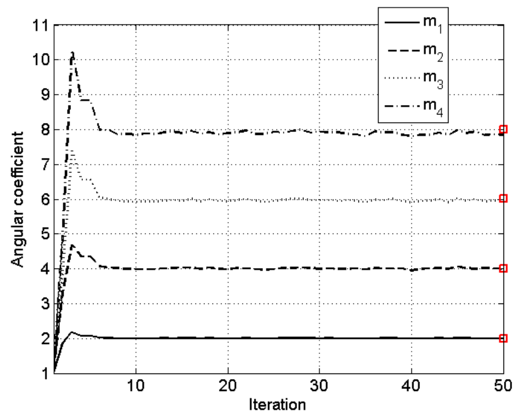

Fig. 3: Convergence of $\mathbf{m}$.

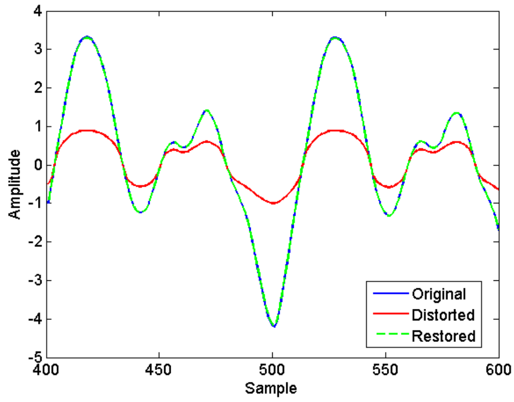

Fig. 4: Comparison of original, distorted and restored signals.

After application of the estimated inverse distortion to the corrupted signal, the restored signal was audibly undistinguishable from their original version, as Figure 4 visually confirms. Measured mean iteration duration was around 0.12 s.

5.2. SMOOTH NONLINEAR DISTORTION

The signal was artificially degraded by a distorting curve described by $\arctan(\lambda x)/\lambda$, with $\lambda \in \{1, \ldots, 5\}$ (the higher the value of λ, the stronger the degradation).

In order to obtain a good estimate of the smooth curve, a larger number of segments was required: $M = 90$ were sufficient even in the most degraded examples, and was adopted in all cases to provide fair comparisons. Consequently, more information was necessary to feed the restoration procedure: 10 blocks of $L = 1000$ samples were processed, which enlarged the overall mean iteration duration to 1.56 s. As in Section 5.1: the estimate was averaged from the last 50 of 100 algorithm iterations; and initialization was performed as in the previous test, but now with an AR-model order $P = 40$, which is sufficient to accurately describe any test signal block.

Once more, the signals restored according to the estimated distortion curves sounded undistinguishable from the original version. The objective evaluation tool Rnonlin [11], tailored to grade from 0 to 1 the similarity between a nonlin-

	Rnonlin note	
Value of λ	**Distorted signal**	**Restored signal**
1	0.9730	0.9996
2	0.9289	0.9994
3	0.8949	0.9991
4	0.8695	0.9976
5	0.8487	0.9943

Table 1: (Rnonlin grade) $\times\lambda$ for distorted and restored signals.

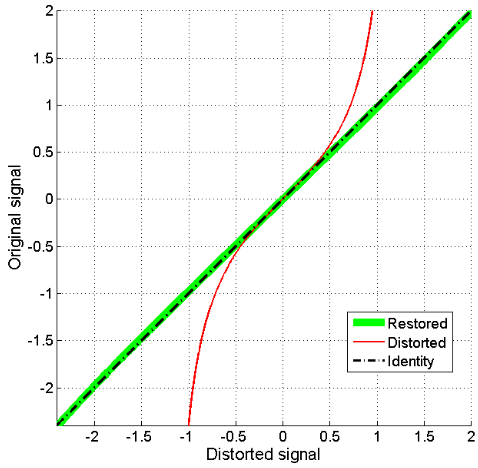

Fig. 5: Distortion curves for distorted and restored signals.

early distorted signal and its reference version, was used to check the results. Table 1 attests that the algorithm performed very well in all cases. As a visual confirmation of this good performance, Figure 5 compares the distorting and estimated curves for $\lambda = 3$, which seem nearly identical. All signals are available from `www.smt.ufrj.br/~hugo.carvalho`, along with additional examples.

6. CONCLUSION AND FURTHER WORKS

In this paper, an algorithm aiming to identify and remove memoryless nonlinear distortions typically found in audio signals was proposed. The method is based on Bayesian statistics; does not assume any knowledge on the original signal—which is modeled by a fixed-order AR model with unknown coefficients; and models the distortion curve as piecewise linear.

The main advantages of the proposed method are that the coefficients describing the nonlinear distortion are easily interpretable and constrainable, and that the model is general enough to deal quite well with smooth memoryless distorting curves even if they are quite severe. In all performed tests, the restored signal sounded indistinguishable from the original, which was confirmed by Rnonlin grades.

The need for a large number of segments to approximate a general nonlinear distortion should be further investigated. It was noted that the piecewise linear approximation leaves an almost white noise in the restored signal, to which the AR-parameter estimation is known to be very sensitive, and thus capable of impairing the estimate of $\mathbf{m}$. By increasing M, this residual can be made so low as to not affect any parameter estimate. On the other hand, more blocks must be processed to provide the necessary information, since a smaller number of signal samples are modified by each segment of the distortion curve; too few samples around $m = 1$ may even erroneously re-scale the signal magnitude due to a poor estimation of $\boldsymbol{\sigma}_e^2$. Another future target in this work is the inclusion of memory in the model.

REFERENCES

[1] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, Springer, New York, USA, 2001.

[2] T. B. Bakó and T. Dabóczi, "Restoration of nonlinearly distorted audio with the application to old motion pictures," in *Proc. AES 20th Int. Conf.*, Budapest, Hungary, Oct. 2001, AES, pp. 191–198.

[3] W. Klippel, "Compensation for nonlinear distortion of horn loudspeakers by digital signal processing," *J. Audio Engin. Soc.*, vol. 44, no. 11, pp. 964–972, Nov. 1996.

[4] G. Picard and O. Cappe, "Blind identification of hammerstein nonlinear distortion models," in *Proc. Workshop App. Sig. Proc. Audio Acoust. (WASPAA)*, New Paltz, USA, Oct. 2003, IEEE, pp. 17–20.

[5] F. R. Ávila and L. W. P. Biscainho, "ML estimation of memoryless nonlinear distortions in audio signals," in *Proc. Int. Conf. Acoust., Speech, Sig. Proc. (ICASSP)*, Florence, Italy, May 2014, IEEE, pp. 4493–4497.

[6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration - A Statistical Model Based Approach*, Cambridge University Press, Cambridge, UK, 1998.

[7] G. H. Givens and J. A. Hoeting, *Computational Statistics*, Wiley, Hoboken, USA, 2012.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, USA, 2007.

[9] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, USA, 2006.

[10] J. D. Cryer and K.-S. Chan, *Time Series Analysis: With Applications in R*, Springer, New York, USA, 2010.

[11] B. C. J. Moore, C. T. Tan, N. Zacharov, and V. V. Mattila, "Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion," *J. Audio Engin. Soc.*, vol. 52, no. 12, pp. 1228–1244, Dec. 2004.