EMBEDDED-OPTIMIZATION-BASED LOUDSPEAKER COMPENSATION USING A GENERIC HAMMERSTEIN LOUDSPEAKER MODEL

Bruno Defraene, Toon van Waterschoot, Moritz Diehl, and Marc Moonen

Dept. E.E./ESAT, SCD-SISTA, KU Leuven / iMinds-KU Leuven Future Health Department
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

ABSTRACT

This paper presents an embedded-optimization-based algorithm for loudspeaker compensation using a generic Hammerstein loudspeaker model, i.e. a cascade of a memoryless nonlinearity and a linear finite impulse response filter. An optimization procedure is embedded into the algorithm to carry out the loudspeaker compensation on a frame-by-frame basis. In order to minimize the perceptible distortion incurred in the loudspeaker, a perceptually meaningful optimization criterion is constructed by using a psychoacoustic model. The resulting per-frame optimization problems are solved efficiently using a gradient optimization method. Objective evaluation experiments show that the proposed loudspeaker compensation algorithm provides a significant audio quality improvement, and this for all considered amplitude levels.

Index Terms— Loudspeaker compensation, audio quality, Hammerstein model, embedded optimization.

1. INTRODUCTION

Loudspeakers have a non-ideal response and consequently introduce linear as well as nonlinear distortions in the reproduced audio signal. These distortions in most cases result in a degradation of the perceived audio quality. Nonlinear distortion is a notably prominent problem in small and low-cost loudspeakers, which are ubiquitous in mobile devices, especially so at high playback levels [1].

Loudspeaker compensation techniques aim at reducing the effects caused by the non-ideal loudspeaker response. The idea is to apply a digital compensation operation in cascade with the audio reproduction channel to counteract the response errors and nonlinearities introduced by the loudspeaker. Traditionally, loudspeakers have been modeled by *linear* systems such as FIR filters, IIR filters, warped filters or Kautz filters. The aim of linear loudspeaker compensation techniques is then to identify/approximate and apply the inverse digital filter to the audio signal prior to playback [2]. Nonlinear behaviour can be taken into account by using nonlinear loudspeaker models such as Hammerstein models, Wiener-Hammerstein models and Volterra models. The aim of nonlinear loudspeaker compensation techniques is then to invert the nonlinear system under consideration [3].

This paper presents an embedded-optimization-based algorithm for performing loudspeaker compensation using a generic Hammerstein loudspeaker model. An optimization procedure is embedded into the algorithm to carry out the loudspeaker compensation on a frame-by-frame basis. In order to minimize the perceptible distortion incurred in the loudspeaker, a psychoacoustic model is incorporated which captures knowledge about the human perception of sound. The proposed algorithm extends the compensation method in [4], which focuses solely on compensating a Hammerstein model containing a hard clipping nonlinearity. The proposed algorithm enables to compensate a *generic* Hammerstein model, i.e. a cascade of *any* memoryless nonlinearity and a linear finite impulse response filter.

The paper is organized as follows. In Section 2, the embedded-optimization-based loudspeaker compensation method is presented and the incorporation of a psychoacoustic model is discussed. In Section 3, a gradient optimization method for solving the per-frame optimization problems is described. In Section 4, the results from objective evaluation experiments are discussed. In Section 5, some concluding remarks are presented.

2. HAMMERSTEIN MODEL COMPENSATION

2.1. Hammerstein model description

The loudspeaker is modeled by a Hammerstein model, i.e. a cascade of a memoryless nonlinearity and a linear finite im-

This research work was carried out at the ESAT Laboratory of KU Leuven, and was supported by Research Council KUL: PFV/10/002 Optimization in Engineering Center OPTEC, GOA/10/09 MaNet, GOA/10/11 Global real- time optimal control of autonomous robots and mechatronic systems; Flemish Government: IOF / KP / SCORES4CHEM, iMinds 2013; FWO: PhD/postdoc grants, projects: G.0320.08 (convex MPC), G.0377.09 (Mechatronics MPC); IWT: PhD Grants, projects: SBO LeCoPro; Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017), IUAP P7/23 (BESTCOM, Belgian network on stochastic modeling analysis design and optimization of communication systems, 2012-2017); EU: FP7-DREAMS (MC ITN-316969), FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC ST HIGHWIND (259 166), Eurostars SMART, ACCM. The scientific responsibility is assumed by its authors.

pulse response (FIR) filter. The FIR filter has an impulse response $h[n], n = 0 \ldots L-1$. The memoryless nonlinearity $g(x)$ is represented as a linear combination of P basis functions,

$$g(x) = \sum_{p=1}^{P} c_p \psi_p(x) = \mathbf{c}^T \boldsymbol{\psi}(x) \quad (1)$$

where the basis functions are stacked in a vector $\boldsymbol{\psi}(x) = [\psi_1(x), ..., \psi_P(x)]^T$ and the corresponding coefficients are stacked in a vector $\mathbf{c} = [c_1, ..., c_P]^T$.

A frame-by-frame processing of the digital input audio signal $x[n]$ will be applied, employing input frames $\mathbf{x}_m = [x_{m,1}, ..., x_{m,N}]^T \in \mathbb{R}^N, m = 0, 1 \ldots M$, with $N \geq L-1$. The output $\mathbf{g}(\mathbf{x}_m)$ of the memoryless nonlinearity for a given input frame $\mathbf{x}_m$ is straightforwardly constructed using the relation (1),

$$\begin{aligned} \mathbf{g}(\mathbf{x}_m) &= [g(x_{m,1}), ..., g(x_{m,N})]^T \\ &= \boldsymbol{\Psi}(\mathbf{x}_m)\mathbf{c} \end{aligned} \quad (2)$$

where the basis function vectors for the different samples are assembled in a matrix $\boldsymbol{\Psi}(\mathbf{x}_m) = [\boldsymbol{\psi}(x_{m,1}), ..., \boldsymbol{\psi}(x_{m,N})]^T$.

The output frame $\mathbf{y}_m$ of the Hammerstein model can then be written as

$$\mathbf{y}_m = \mathbf{H}_m \mathbf{g}(\mathbf{x}_m) + \tilde{\mathbf{H}}_m \mathbf{g}(\mathbf{x}_{m-1}) \quad (3)$$

where the matrices $\mathbf{H}_m \in \mathbb{R}^{N \times N}$ and $\tilde{\mathbf{H}}_m \in \mathbb{R}^{N \times N}$ implement a convolution operation with the FIR filter $h[n]$ as follows

$$\mathbf{H}_m = \begin{bmatrix} h[0] & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ h[1] & h[0] & 0 & \ldots & \ldots & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ h[L-1] & & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & h[L-1] & \ldots & h[1] & h[0] \end{bmatrix} \quad (4)$$

$$\tilde{\mathbf{H}}_m = \begin{bmatrix} 0 & \ldots & 0 & h[L-1] & \ldots & h[2] & h[1] \\ 0 & \ldots & 0 & 0 & h[L-1] & \ldots & h[2] \\ \vdots & & & & \ddots & \ddots & \vdots \\ \vdots & & & & & \ddots & h[L-1] \\ \vdots & & & & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 \end{bmatrix}. \quad (5)$$

2.2. Embedded-optimization-based compensation

Figure 1 shows the operation of the proposed Hammerstein loudspeaker model compensation technique. Before it is fed into the loudspeaker, the input frame $\mathbf{x}_m$ passes through the loudspeaker compensation block. For a given frame $\mathbf{x}_m$, the loudspeaker compensation consists of the following steps:

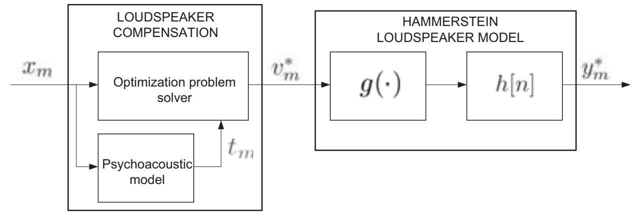

Fig. 1. Embedded-optimization-based Hammerstein loudspeaker model compensation: schematic overview

1. Calculate the global masking threshold $\mathbf{t}_m \in \mathbb{R}^{\frac{N}{2}+1}$ of the input frame $\mathbf{x}_m$ using a psychoacoustic model (see Subsection 2.3).

2. Calculate a compensated input frame $\mathbf{v}_m^* \in \mathbb{R}^N$ as the solution of an optimization problem, such that the corresponding output frame $\mathbf{y}_m^*$ is perceptually as close as possible to $\mathbf{x}_m$.

The compensated input frame $\mathbf{v}_m^*$ is calculated from the knowledge of the input frame $\mathbf{x}_m$ and its masking threshold $\mathbf{t}_m$. The objective function reflects the amount of perceptible distortion added between $\mathbf{y}_m$ and $\mathbf{x}_m$,

$$\mathbf{v}_m^* = \arg\min_{\mathbf{v}_m \in \mathbb{R}^N} \frac{1}{2N} \sum_{i=0}^{N-1} w_m(i) |Y_m(e^{j\omega_i}) - X_m(e^{j\omega_i})|^2 \quad (6)$$

where $\omega_i = (2\pi i)/N$ represents the discrete frequency variable, $X_m(e^{j\omega_i})$ and $Y_m(e^{j\omega_i})$ are the discrete frequency components of $\mathbf{x}_m$ and $\mathbf{y}_m$ respectively, and $w_m(i)$ are the weights of a perceptual weighting function to be defined in subsection 2.3.

Using the Hammerstein model input-output relation (3), optimization problem (6) can conveniently be rewritten as

$$\begin{aligned} \mathbf{v}_m^* &= \arg\min_{\mathbf{v}_m \in \mathbb{R}^N} \frac{1}{2} (\mathbf{y}_m - \mathbf{x}_m)^T \underbrace{\mathbf{D}^T \mathbf{W}_m \mathbf{D}}_{\triangleq \mathbf{Q}_m} (\mathbf{y}_m - \mathbf{x}_m) \\ &= \arg\min_{\mathbf{v}_m \in \mathbb{R}^N} \frac{1}{2} \left(\mathbf{H}_m \mathbf{g}(\mathbf{v}_m) + \tilde{\mathbf{H}}_m \mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m\right)^T \mathbf{Q}_m \\ &\qquad \left(\mathbf{H}_m \mathbf{g}(\mathbf{v}_m) + \tilde{\mathbf{H}}_m \mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m\right) \\ &= \arg\min_{\mathbf{v}_m \in \mathbb{R}^N} \frac{1}{2} \mathbf{g}(\mathbf{v}_m)^T \mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m \mathbf{g}(\mathbf{v}_m) \\ &\qquad + \left(\mathbf{H}_m^T \mathbf{Q}_m^T (\tilde{\mathbf{H}}_m \mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m)\right)^T \mathbf{g}(\mathbf{v}_m) \end{aligned} \quad (7)$$

where $\mathbf{D} \in \mathbb{C}^{N \times N}$ is the unitary Discrete Fourier Transform (DFT) matrix, $\mathbf{W}_m \in \mathbb{R}^{N \times N}$ is a diagonal weighting matrix with positive diagonal elements $w_m(i)$, obeing the symmetry property $w_m(i) = w_m(N-i)$ for $i = 1, 2, ..., \frac{N}{2} - 1$[1].

[1]Note that the optimization problem is quadratic in the nonlinearly transformed optimization variable $\mathbf{g}(\mathbf{v}_m) \in \mathbb{R}^N$. Also, the optimization problem is not convex, so the global optimality of the solution can not be guaranteed.

2.3. Perceptual weighting

The rationale behind applying perceptual weights in the summation (6) is the fact that distortion at certain frequencies is more perceptible than distortion at other frequencies. Two phenomena of human auditory perception are responsible for this. A first phenomenon is the *absolute threshold of hearing*, which is defined as the required intensity (dB) of a pure tone such that an average listener will just hear the tone in a noiseless environment. The absolute threshold of hearing is a function of the tone frequency. A second phenomenon is *simultaneous masking*, where the presence of certain spectral energy (the masker) masks the simultaneous presence of weaker spectral energy (the maskee). Combining both phenomena, the global masking threshold gives the amount of distortion energy (dB) at each frequency that can be masked by the signal.

Consider the input frame $\mathbf{x}_m$ to act as the masker, and $\mathbf{y}_m - \mathbf{x}_m$ as the maskee. By selecting the weights $w_m(i)$ to be exponentially decreasing with the value of the global masking threshold of $\mathbf{x}_m$ at frequency i, the objective function reflects the amount of perceptible distortion introduced,

$$w_m(i) = \begin{cases} 10^{-\alpha t_m(i)} & \text{if } 0 \leq i \leq \frac{N}{2} \\ 10^{-\alpha t_m(N-i)} & \text{if } \frac{N}{2} < i \leq N-1 \end{cases} \tag{8}$$

where $\mathbf{t}_m$ is the global masking threshold of $\mathbf{x}_m$ (in dB), calculated using the MPEG-1 Layer 1 psychoacoustic model [5].

3. OPTIMIZATION ASPECTS

3.1. Regularized optimization problem

To preserve inter-frame continuity, overlapping frames will be used, where frames overlap with K samples, with $K \leq N - L + 1$. The input frame vector $\mathbf{v}_m$ can then be partitioned as follows,

$$\mathbf{v}_m = \begin{bmatrix} \mathbf{v}_m^{\mathrm{F}} \\ \mathbf{v}_m^{\mathrm{M}} \\ \mathbf{v}_m^{\mathrm{L}} \end{bmatrix} \tag{9}$$

where $\mathbf{v}_m^{\mathrm{F}} \in \mathbb{R}^K$ contains the first (overlapping with the previous frame) samples, $\mathbf{v}_m^{\mathrm{M}} \in \mathbb{R}^{N-2K}$ contains the middle (non-overlapping) samples, and $\mathbf{v}_m^{\mathrm{L}} \in \mathbb{R}^K$ contains the last (overlapping with the next frame) samples. We can write down a slightly modified version of optimization problem (7) accounting for the frame overlap structure,

$$\mathbf{v}_m^* = \underset{\mathbf{v}_m \in \mathbb{R}^N}{\arg\min} \; \frac{1}{2}\, \mathbf{g}(\mathbf{v}_m)^T \underbrace{\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m}_{\triangleq \mathbf{A}_m} \mathbf{g}(\mathbf{v}_m) + \underbrace{(\mathbf{H}_m^T \mathbf{Q}_m^T (\tilde{\mathbf{H}}_m \mathbf{g}\left(\begin{bmatrix} \mathbf{v}_{m-1}^{\mathrm{F},*} \\ \mathbf{v}_{m-1}^{\mathrm{M},*} \end{bmatrix}\right) - \mathbf{x}_m))^T}_{\triangleq \mathbf{b}_m} \mathbf{g}(\mathbf{v}_m) \tag{10}$$

where $\tilde{\mathbf{H}}_m$ is of similar structure as in (5), but now $\tilde{\mathbf{H}}_m \in \mathbb{R}^{N \times (N-K)}$.

Algorithm 1 Hammerstein loudspeaker model compensation using gradient optimization method

---

Input $\mathbf{x}_m \in \mathbb{R}^N$, $\mathbf{v}_{m-1}^* \in \mathbb{R}^N$, $h[n] \in \mathbb{R}^L$, $\mathbf{c} \in \mathbb{R}^P$, $\boldsymbol{\psi}(x) \in C(\mathbb{R} \to \mathbb{R}^P)$, $\mathbf{D} \in \mathbb{C}^{N \times N}$, $\alpha, \beta, \gamma, \lambda, k_{\max}$

Output $\mathbf{v}_m^* \in \mathbb{R}^N$

```
Compute masking threshold t_m for x_m using [5]
Compute weights w_m using (8)
Q_m = D^T diag(w_m) D
Construct H_m and H~_m using (4)-(5)
Initialize v_m^0 = x_m
k = 0
while k < k_max do
    Compute ∇f(v_m^k) using (13)-(15)
    while f(v_m^k − s_m^k ∇f(v_m^k)) > f(v_m^k) − β s_m^k ||∇f(v_m^k)||_2^2 do
        s_m^k = γ s_m^k
    end while
    v_m^{k+1} = v_m^k − s_m^k ∇f(v_m^k)
    k = k + 1
end while
v_m^* = v_m^k
```

To enforce continuity between the first samples $\mathbf{v}_m^{\mathrm{F}}$ of the current frame and the last samples $\mathbf{v}_{m-1}^{\mathrm{L},*}$ of the previously optimized frame, the optimization problem (10) is regularized, i.e. a regularization term is added in the objective function,

$$\begin{aligned} \mathbf{v}_m^* &= \underset{\mathbf{v}_m \in \mathbb{R}^N}{\arg\min}\, f(\mathbf{v}_m) \\ &= \underset{\mathbf{v}_m \in \mathbb{R}^N}{\arg\min}\; \frac{1}{2}\, \mathbf{g}(\mathbf{v}_m)^T\, \mathbf{A}_m\, \mathbf{g}(\mathbf{v}_m) + \mathbf{b}_m^T\, \mathbf{g}(\mathbf{v}_m) \\ &\qquad + \frac{\lambda}{2} \|\mathbf{v}_m^F - \mathbf{v}_{m-1}^{L,*}\|_2^2 \end{aligned} \tag{11}$$

where λ is a regularization parameter.

3.2. Gradient optimization method

The proposed optimization method for solving optimization problem (11) is an iterative gradient method. Introducing the notation $\mathbf{v}_m^k$ for the kth iterate of the mth frame, the $(k+1)$th iteration of the gradient method consists in taking a step of stepsize s_m^k along the negative gradient direction,

$$\mathbf{v}_m^{k+1} = \mathbf{v}_m^k - s_m^k \nabla f(\mathbf{v}_m^k) \tag{12}$$

where the gradient $\nabla f(\mathbf{v}_m^k)$ is computed as

$$\nabla f(\mathbf{v}_m^k) = \mathrm{diag}(\nabla \mathbf{g}(\mathbf{v}_m^k)) \left[\mathbf{H}_m^T \mathbf{Q}_m^T \left(\mathbf{H}_m \mathbf{g}(\mathbf{v}_m^k) + \tilde{\mathbf{H}}_m \mathbf{g}\left(\begin{bmatrix} \mathbf{v}_{m-1}^{\mathrm{F},*} \\ \mathbf{v}_{m-1}^{\mathrm{M},*} \end{bmatrix} \right) - \mathbf{x}_m \right) \right] + \lambda \begin{bmatrix} \mathbf{v}_m^{\mathrm{F},k} - \mathbf{v}_{m-1}^{\mathrm{L},*} \\ \mathbf{0} \end{bmatrix} \tag{13}$$

where the gradient of the memoryless nonlinearity $\nabla \mathbf{g}(\mathbf{x}_m) \in \mathbb{R}^N$ is computed as

$$\begin{aligned} \nabla \mathbf{g}(\mathbf{x}_m) &= [\nabla g(x_{m,1}), ..., \nabla g(x_{m,N})]^T \\ &= \nabla \boldsymbol{\Psi}(\mathbf{x}_m)\mathbf{c} \end{aligned} \tag{14}$$

and where the matrix $\nabla\mathbf{\Psi}(\mathbf{x}_m) \in \mathbb{R}^{N\times P}$ is defined as

$$\nabla\mathbf{\Psi}(\mathbf{x}_m) = \begin{bmatrix} \nabla\psi_1(x_{m,1}) & \dots & \nabla\psi_P(x_{m,1}) \\ \vdots & \ddots & \vdots \\ \nabla\psi_1(x_{m,N}) & \dots & \nabla\psi_P(x_{m,N}) \end{bmatrix}. \quad (15)$$

The stepsize s_m^k is determined using a backtracking line search for satisfying the Armijo condition [6]. The resulting Hammerstein loudspeaker model compensation algorithm, including the detailed description of the backtracking line search, is given in Algorithm 1.

4. SIMULATION RESULTS

A Hammerstein loudspeaker model was used with the following specifications:

- A memoryless nonlinearity with $P = 3$ basis functions $\boldsymbol{\psi}(x) = [x\ x^3\ x^5]^T$ and a corresponding coefficient vector $\mathbf{c} = [0.6\ 0.3\ 0.4]^T$.
- An FIR filter ($L = 128$) with impulse response $h[n]$, designed using the frequency sampling method `fir2` in Matlab, having a required magnitude response $[1\ 0.95\ 0.75\ 0.50\ 0.20\ 0]^T$ for the frequencies $[0\ 0.2\ 0.4\ 0.6\ 0.8\ 1]^T \times f_{\text{Nyquist}}$.

A test database consisting of 8 audio excerpts was compiled (see Table 1 for details). Each audio signal in the test database was fed into the Hammerstein loudspeaker model, once with and once without performing a compensation step. The following parameters were used in the compensation step: $N = 512$, $O = 128$, $\alpha = 0.04$, $\beta = 0.1$, $\gamma = 0.6$, $\lambda = 10^{-4}$ and $k_{\max} = 250$. These simulations were performed at four different relative average amplitude levels $a = \{0.25, 0.50, 0.75, 1.00\}$.

The objective audio quality improvement for each audio excerpt was assessed by computing the ΔODG measure,

$$\Delta\text{ODG} = \text{ODG}(\mathbf{x}, \mathbf{y}^*) - \text{ODG}(\mathbf{x}, \mathbf{y}) \quad (16)$$

where $\mathbf{x}$ is the input signal, $\mathbf{y}$ is the uncompensated output signal, $\mathbf{y}^*$ is the compensated output signal, and $\text{ODG}(\mathbf{r}, \mathbf{d})$ is an objective measure [7] which predicts the audio quality of a signal $\mathbf{d}$ with respect to a signal $\mathbf{r}$ on a scale of $[0, -4]$, where 0 corresponds to an imperceptible degradation, and -4 corresponds to a very annoying degradation.

In Figure 2, the ΔODG scores are shown. We observe a positive audio quality improvement for all audio excerpts, and this for all considered relative average amplitude levels. For most audio excerpts, increasing audio quality improvement scores are observed for increasing amplitude levels. This is to be expected, as it is exactly at higher amplitude levels that the Hammerstein nonlinearity is severely affecting the audio signal, and that compensating for it is considerably improving the resulting audio quality.

In Figure 3, the ΔODG scores for all audio excerpts at relative average amplitude level $a = 0.50$ are shown, for different maximum iteration numbers $k_{\max} = \{50, 100, 250\}$. We observe that increasing the maximum iteration number results in an increased audio quality improvement. On average, 86 percent of the total audio quality improvement (after 250 iterations) is achieved after only 100 iterations. Likewise, an average of 72 percent of the total audio quality improvement is achieved after only 50 iterations.

5. CONCLUSION

In this paper, an embedded-optimization-based algorithm for loudspeaker compensation using a generic Hammerstein loudspeaker model was presented. By incorporating a psychoacoustic model, a percepually meaningful optimization criterion was constructed. In order to efficiently solve the per-frame optimization problems, a gradient optimization method was proposed. From objective evaluation experiments, it was shown that the proposed loudspeaker compensation algorithm results in a significant audio quality improvement, and this for all considered amplitude levels.

6. REFERENCES

[1] K. Lashkari, "A modified Volterra-Wiener-Hammerstein model for loudspeaker precompensation," in *Proc. 39th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 2005, pp. 344–348.

[2] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filters: Direct least squares design," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–13, 2007.

[3] K. Shi, G. T. Zhou, and M. Viberg, "Compensation for nonlinearity in a Hammerstein system using the coherence function with application to nonlinear acoustic echo cancellation," *IEEE Transactions on Signal processing*, vol. 55, no. 12, pp. 5853–5858, 2007.

[4] B. Defraene, T. van Waterschoot, M. Diehl, and M. Moonen, "Perception-based nonlinear loudspeaker compensation through embedded convex optimization," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug. 2012, pp. 2472 –2476.

[5] ISO/IEC, "11172-3 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio," 1993.

[6] L. Armijo, "Minimization of functions having lipschitz continuous first partial derivatives.," *Pacific Journal of mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[7] International Telecommunications Union Recommendation BS.1387, "Method for objective measurements of perceived audio quality," 1998.

[8] Francis Poulenc, "Sonata for flute and piano (Cantilena)," French Flute Music, Naxos 8.557328, 2005.

[9] Red Hot Chili Peppers, "Californication," Californication, Warner Bros. Records 9362473862, 1999.

[10] An Pierlé & White Velvet, "Mexico," An Pierlé & White Velvet, PIAS Recordings 941.0170.020, 2006.

[11] Frederic Chopin, "Waltz op. 69 no. 2," Favourite Piano Works (Vladimir Ashkenazy), Decca 02894448302, 1995.

[12] Kraftwerk, "Tour de France Etape 1," Minimum-Maximum, EMI Music 724356061620, 2005.

Table 1. Audio excerpts database used for comparative audio quality evaluation (16 bit mono at 44.1 kHz)

Nr.	Name	Texture	Composition	Style	Duration [s]	$\text{Sample}_{\text{start}}$	$\text{Sample}_{\text{end}}$	Origin
1	poulenc.wav	polyphonic	instrumental	classical	17.8	400000	1183000	[8]
2	rhcp.wav	polyphonic	instrumental	rock	9.8	468996	900000	[9]
3	pierle.wav	polyphonic	instrumental+vocal	pop	11.7	2234000	2750000	[10]
4	chopin.wav	monophonic	instrumental	classical	17.8	50000	836200	[11]
5	kraftw.wav	polyphonic	instrumental	electronic	17.2	7480000	8240000	[12]
6	breftri.wav	monophonic	instrumental	classical	19.7	1	869675	[7]
7	crefsax.wav	monophonic	instrumental	classical	10.9	1	479026	[7]
8	grefcla.wav	monophonic	instrumental	classical	6.9	1	302534	[7]

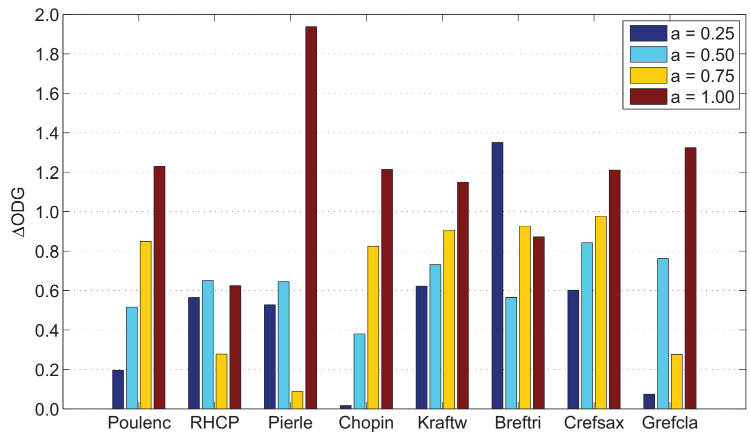

Fig. 2. Audio quality improvement scores for different audio excerpts, at relative average amplitude levels $a = \{0.25, 0.50, 0.75, 1.00\}$.

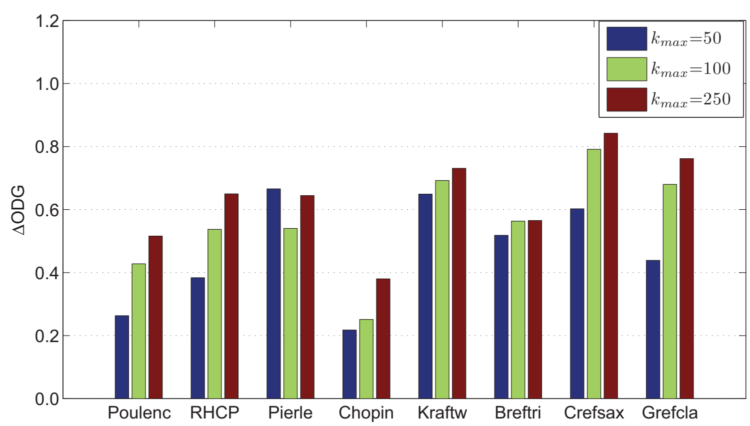

Fig. 3. Audio quality improvement scores for different audio excerpts at relative average amplitude level $a = 0.50$, for different maximum iteration numbers $k_{\max} = \{50, 100, 250\}$.