KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Bruno Defraene, Toon van Waterschoot, Moritz Diehl, and Marc Moonen **Embedded-optimization-based loudspeaker precompensation using a Hammerstein loudspeaker model** *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 22, no. 11, Nov. 2014, pp. 1648-1659.
Archived version	Author manuscript: the content is identical to the content of the submitted paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/TASLP.2014.2344862
Journal homepage	http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6870515
Author contact	toon.vanwaterschoot@esat.kuleuven.be *+ 32 (0)16 321927*
IR	ftp://ftp.esat.kuleuven.be/pub/SISTA/vanwaterschoot/abstracts/13-239.html

(article begins on next page)

Embedded-Optimization-Based Loudspeaker Precompensation Using a Hammerstein Loudspeaker Model

Bruno Defraene, *Member, IEEE*, Toon van Waterschoot, *Member, IEEE*, Moritz Diehl, *Member, IEEE*, and Marc Moonen, *Fellow, IEEE*

***Abstract*—This paper presents an embedded-optimization-based loudspeaker precompensation algorithm using a Hammerstein loudspeaker model, i.e. a cascade of a memoryless nonlinearity and a linear finite impulse response filter. The loudspeaker precompensation consists in a per-frame signal optimization. In order to minimize the perceptible distortion incurred in the loudspeaker, a psychoacoustically motivated optimization criterion is proposed. The resulting per-frame signal optimization problems are solved efficiently using first-order optimization methods. Depending on the invertibility and the smoothness of the memoryless nonlinearity, different first-order optimization methods are proposed and their convergence properties are analyzed. Objective evaluation experiments using synthetic loudspeaker models and real loudspeakers show that the proposed loudspeaker precompensation algorithm provides a significant audio quality improvement, especially so at high playback levels.**

***Index Terms*—Embedded optimization, gradient optimization, hammerstein model, loudspeaker precompensation, sound perception.**

I. Introduction

ACHIEVING a high perceived audio quality is a main concern in the development of any audio reproduction system. In general, loudspeakers have a non-ideal response introducing both linear and nonlinear distortion in the reproduced audio signal. Linear distortion involves changes in the relative amplitudes and phases of the frequency components constituting the original audio signal. Nonlinear distortion involves the introduction of frequency components that are not present in the original audio signal, and is a notably prominent problem in small and low-cost loudspeakers, which are ubiquitous in mobile devices, especially so at high playback levels [1].

The presence of linear and nonlinear distortion has been found to result in a significant degradation of the perceived audio quality, both when present separately [2] and simultaneously [3]. Linear distortion is typically perceived as affecting timbre or tone quality, whereas nonlinear distortion is typically perceived as adding harshness or noisiness, or as the perception of sounds that are not present in the original signal, such as crackles or clicks.

Loudspeaker precompensation techniques aim at reducing the effects caused by the non-ideal loudspeaker characteristics. The idea is to apply a digital precompensation operation in cascade with the audio reproduction channel to counteract the loudspeaker response errors and nonlinearities. Traditionally, loudspeakers have been modeled by *linear* systems such as FIR filters, IIR filters, warped filters or Kautz filters. The aim of linear loudspeaker precompensation (also known as equalization) techniques is then to identify/approximate and apply an inverse digital filtering to the audio signal prior to playback [4]. Nonlinear behavior 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 precompensation techniques is then to invert the nonlinear system under consideration [5].

This paper presents an embedded-optimization-based precompensation algorithm for linear as well as nonlinear distortions incurred in loudspeakers, using a Hammerstein loudspeaker model. The loudspeaker precompensation consists in a per-frame signal optimization. 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.

This paper builds on the embedded-optimization-based loudspeaker precompensation algorithms recently proposed in [6][7], of which the main ideas will be reviewed in this paper, thereby introducing three major novel contributions:

1) The psychoacoustic principles and the resulting psychoacoustic model applied for perceptible distortion minimization are elaborated in detail.
2) Different gradient optimization methods are proposed for efficiently solving the per-frame signal optimization problems and their convergence properties are analyzed. Each

Manuscript received November 29, 2013; revised June 11, 2014; accepted June 12, 2014. Date of publication July 31, 2014; date of current version August 07, 2014. This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 Optimization in Engineering Center (OPTEC), Concerted Research Action GOA-MaNet, GOA/10/11, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P7/19 "Dynamical systems control and optimization" (DYSCO) 2012-2017, and Flemish Government IOF/KP/SCORES4CHEM, SBO LeCoPro, the EU FP7- EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC ST HIGHWIND (259 166), Eurostars SMART, and ACCM. The scientific responsibility is assumed by its authors. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Thushara D. Abhayapala.

The authors are with the KU Leuven Department of Electrical Engineering, ESAT-STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, B-3001 Leuven, Belgium (e-mail: bruno.defraene@esat.kuleuven.be; toon.vanwaterschoot@esat.kuleuven.be; moritz.diehl@esat.kuleuven.be; marc.moonen@esat.kuleuven.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2014.2344862

2329-9290 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

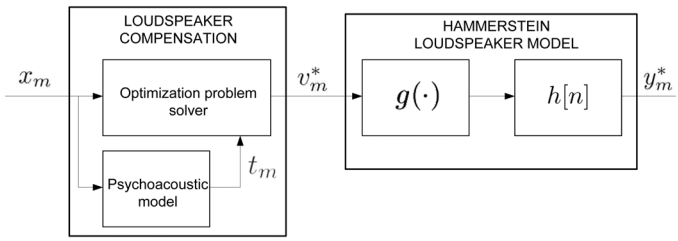

Fig. 1. Embedded-optimization-based precompensation: schematic overview.

optimization method is tailored to a given class of memoryless nonlinearities.

3) Extensive audio quality evaluation experiments are performed and reported, both using synthetic Hammerstein loudspeaker models (comprising different classes of memoryless nonlinearities), and using identified Hammerstein loudspeaker models.

This paper is organized as follows. In Section II, the Hammerstein loudspeaker model is introduced and an embedded-optimization-based precompensation algorithm is proposed. In Section III, three optimization methods for efficiently solving the per-frame signal optimization problems are proposed, where each optimization method is designed for a given class of memoryless nonlinearities. In Section IV, the proposed loudspeaker precompensation algorithm is evaluated using synthetic Hammerstein loudspeaker models and real loudspeakers. In Section V, some concluding remarks are presented.

II. Embedded-Optimization-Based Precompensation

A. Hammerstein Model Description

The loudspeaker is modeled by a Hammerstein model, i.e. a cascade of a memoryless nonlinearity and a linear finite impulse response (FIR) filter. The FIR filter has an impulse response $h[n], n = 0 \ldots L$. The memoryless nonlinearity $g(x)$ is represented as a linear combination of P basis functions,

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

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

A per-frame processing of the digital input audio signal $x[n]$ will be applied, employing input frames $\mathbf{x}_m = [x_{m,1}, \ldots, x_{m,N}]^T \in \mathbb{R}^N, m = 0, 1 \ldots M$, with $N \geq L$, and $x_{m,i} = x[mN + i]$. 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),

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

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

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

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

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,

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

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

B. Embedded-Optimization-Based Precompensation

Fig. 1 shows the operation of the proposed embedded-optimization-based loudspeaker precompensation technique. Before it is fed into the loudspeaker, the input frame $\mathbf{x}_m$ passes through the loudspeaker precompensation block. For a given input frame $\mathbf{x}_m$, the loudspeaker precompensation consists of the following steps:

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 Section II-C).
2) Calculate a precompensated 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 precompensated 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 $\mathbf{x}_m$,

$$\mathbf{v}_m^* = \underset{\mathbf{v}_m \in \mathbb{R}^N}{\arg\min} \frac{1}{2N} \sum_{i=0}^{N-1} w_{m,i} |Y_m(e^{j\omega_i}) - X_m(e^{j\omega_i})|^2 \tag{6}$$

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 strictly positive weights of a perceptual weighting function $\mathbf{w}_m = [w_{m,0}, \ldots, w_{m,N-1}]^T \in \mathbb{R}^N$. The perceptual weighting function depends on the masking threshold $\mathbf{t}_m$ and is obeying the symmetry property $w_{m,i} = w_{m,N-i}$ for $i = 1, 2, \ldots, \frac{N}{2} - 1$ (see Section II-C).

Optimization problem (6) can then be rewritten as

$$\mathbf{v}_m^* = \underset{\mathbf{v}_m \in \mathbb{R}^N}{\arg\min} \frac{1}{2} (\mathbf{y}_m - \mathbf{x}_m)^T \underbrace{\mathbf{D}^H \mathbf{W}_m \mathbf{D}}_{\triangleq \mathbf{Q}_m} (\mathbf{y}_m - \mathbf{x}_m) \tag{7}$$

where $\mathbf{D} \in \mathbb{C}^{N\times N}$ is the unitary Discrete Fourier Transform (DFT) matrix

$$\mathbf{D} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-j\omega_1} & e^{-j\omega_2} & \dots & e^{-j\omega_{N-1}} \\ 1 & e^{-j\omega_2} & e^{-j\omega_4} & \dots & e^{-j\omega_{2(N-1)}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-j\omega_{N-1}} & e^{-j\omega_{2(N-1)}} & \dots & e^{-j\omega_{(N-1)(N-1)}} \end{bmatrix} \tag{8}$$

and $\mathbf{W}_m \in \mathbb{R}^{N\times N}$ is a diagonal weighting matrix with diagonal elements $w_{m,i}$. By inserting the Hammerstein model input-output relation (3), the optimization problem (7) can be rewritten as follows,

$$\begin{aligned} \mathbf{v}_m^* &= \arg\min_{\mathbf{v}_m\in\mathbb{R}^N} f_0(\mathbf{v}_m) \\ &= \arg\min_{\mathbf{v}_m\in\mathbb{R}^N} \frac{1}{2}(\mathbf{H}_m\mathbf{g}(\mathbf{v}_m) + \tilde{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m)^T\mathbf{Q}_m \\ &\quad (\mathbf{H}_m\mathbf{g}(\mathbf{v}_m) + \tilde{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m) \\ &= \arg\min_{\mathbf{v}_m\in\mathbb{R}^N} \frac{1}{2}\mathbf{g}(\mathbf{v}_m)^T \underbrace{\mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m}_{\triangleq\mathbf{A}_m}\mathbf{g}(\mathbf{v}_m) \\ &\quad + \underbrace{(\mathbf{H}_m^T\mathbf{Q}_m(\tilde{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m))}_{\triangleq\mathbf{b}_m}{}^T\mathbf{g}(\mathbf{v}_m). \end{aligned} \tag{9}$$

In general, the objective function $f_0(\mathbf{v}_m)$ of this optimization problem is not convex in the optimization variable $\mathbf{v}_m$. In Section III, three distinct optimization methods for efficiently solving optimization problem (9) are proposed, each of these being specifically designed for a given class of memoryless nonlinearities $\mathbf{g}$ in the Hammerstein loudspeaker model.

C. Perceptual Weighting Function

The rationale behind applying perceptual weights in the summation (6) is the fact that distortion at different frequencies is differently perceived, and that the relative perceptibility is partly signal-dependent. 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 and has been measured experimentally [9].
- A second phenomenon is *simultaneous masking*, where the presence of strong spectral energy (the masker) masks the simultaneous presence of weaker spectral energy (the maskee), or in other words, renders it imperceptible.

Combining both phenomena, the *instantaneous global masking threshold* $\mathbf{t}_m$ of the input audio signal $\mathbf{x}_m$ gives the amount of signal energy (dB) at each frequency that can be masked by the signal. As such, the masking threshold $\mathbf{t}_m$ gives an indication of the relative perceptibility of signal components at different frequencies.

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 $\mathbf{t}_m$ of the signal $\mathbf{x}_m$ at frequency i, i.e.

$$w_{m,i} = \begin{cases} 10^{-\alpha t_{m,i}} & \text{if } 0 \le i \le \frac{N}{2} \\ 10^{-\alpha t_{m,N-i}} & \text{if } \frac{N}{2} < i \le N-1 \end{cases} \tag{10}$$

the objective function reflects the amount of perceptible distortion introduced. Appropriate values of the compression parameter α have been determined to lie in the range [0.01, 0.04] through informal listening tests performed by the authors. This range of parameters was observed to result in the highest subjective audio quality improvement for the evaluated set of audio excerpts, such that it can be concluded that (6) indeed accurately reflects the amount of perceptible distortion. Note that the beneficial effect on the resulting audio quality of minimizing the psychoacoustically motivated criterion (6) has been previously demonstrated in [8] in the context of audio signal clipping.

The instantaneous global masking threshold $\mathbf{t}_m$ of a given audio signal is calculated using part of the ISO/IEC 11172-3 MPEG-1 Layer 1 psychoacoustic model 1. A complete description of the operation of this psychoacoustic model is beyond the scope of this paper (we refer the reader to [10] and [11]). We outline the relevant steps [8] in the computation of $\mathbf{t}_m$ and illustrate the result of each step on an example audio signal (see Fig. 2):

1) *Spectral analysis and SPL normalization*: In this step a high-resolution spectral estimate of the audio signal is calculated, with spectral components expressed in terms of sound pressure level (SPL). After a normalization operation and a Hann windowing operation on the input signal frame, the PSD estimate is obtained through a 512-point DFT. Fig. 2(a) shows the time-domain input signal, Fig. 2(b) shows the resulting spectral estimate.
2) *Identification of tonal and non-tonal maskers*: It is known from psychoacoustic research that the tonality of a masking component has an influence on its masking properties [12]. For this reason it is important to discriminate between tonal maskers (defined as local maxima of the signal spectrum) and non-tonal maskers. The output of the DFT is used to determine the relevant tonal and non-tonal maskers in the spectrum of the audio signal. In a first phase, tonal maskers are identified at local maxima of the PSD: energy from three adjacent spectral components centered at the local maximum is combined to form a single tonal masker. In a second phase, a single non-tonal masker per critical band is formed by addition of all the energy from the spectral components within the critical band that have not contributed to a tonal masker.
3) *Decimation of maskers*: In this step, the number of maskers is reduced using two criteria. First, any tonal or non-tonal masker below the absolute threshold of hearing is discarded. Next, any pair of maskers occurring within a distance of 0.5 Bark is replaced by the stronger of the two. Figs. 2(c) and 2(d) depict the identified tonal and non-tonal maskers respectively, after decimation.
4) *Calculation of individual masking thresholds*: an individual masking threshold is calculated for each masker in the decimated set of tonal and non-tonal maskers, using fixed psychoacoustic rules. Essentially, the individual masking threshold depends on the frequency, loudness

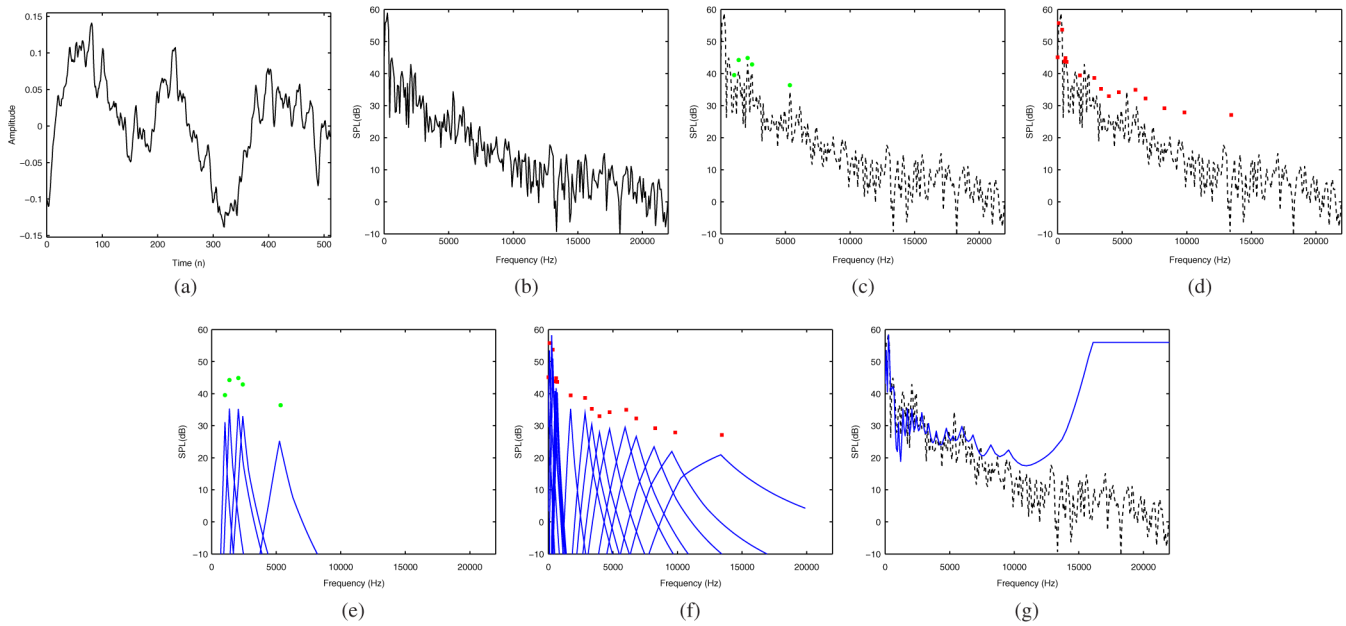

Fig. 2. Different steps in the computation of the global masking threshold using the ISO/IEC 11172-3 MPEG-1 Layer 1 psychoacoustic model 1: (a)-(b) Time domain and normalized frequency domain representations of the input audio signal (c)-(d) Tonal maskers (circles), non-tonal maskers (squares) and input frequency spectrum (dotted line) (e)-(f) Individual masking thresholds related to tonal and non-tonal maskers respectively (g) Global masking threshold (solid line) and input frequency spectrum (dotted line). Figure adapted from [8].

level and tonality of the masker. Fig. 2(e) and 2(f) show the individual masking thresholds associated with tonal and non-tonal maskers, respectively.

5) *Calculation of global masking threshold*: Finally, the global masking threshold is calculated by a power-additive combination of the tonal and non-tonal individual masking thresholds, and the absolute threshold of hearing. This is illustrated in Fig. 2(g).

III. Optimization Methods

The efficient solution of the per-frame optimization problem (9) is crucial for the proposed loudspeaker precompensation algorithm to be applicable in real-time environments. In this section, three optimization methods for efficiently solving the considered optimization problem are proposed, each optimization method being specifically designed for a given class of memoryless nonlinearities $\mathbf{g}$ in the Hammerstein loudspeaker model.

A. Classes of Memoryless Nonlinearities

In order to define the considered classes of memoryless nonlinearities, the notions of *invertibility* and *smoothness* for memoryless nonlinearities are first introduced.

Definition 1 (Invertible memoryless nonlinearity): A memoryless nonlinearity $y = g(x)$ is invertible on the domain $D \subseteq \mathbb{R}$ if there exists a unique function $x = g^{-1}(y)$ for which

$$g^{-1}(g(x)) = x, \quad \forall x \in D. \tag{11}$$

The function $g^{-1}(y)$ is called the inverse memoryless nonlinearity of $g(x)$ on the domain D. The output $\mathbf{g^{-1}}(\mathbf{y}_m)$ of the inverse memoryless nonlinearity for a given frame $\mathbf{y}_m$ is constructed as

$$\mathbf{g^{-1}}(\mathbf{y}_m) = [g^{-1}(y_{m,1}), \ldots, g^{-1}(y_{m,N})]^T. \tag{12}$$

Definition 2 (Smooth memoryless nonlinearity): A memoryless nonlinearity $y = g(x)$ is smooth on the domain $D \subseteq \mathbb{R}$ if its derivatives of all orders exist and are continuous on D.

Based on the above definitions, the following three classes of memoryless nonlinearities[1] can be determined:

- Class I is formed by the *invertible memoryless nonlinearities*, such as the one depicted in Fig. 3(a).
- Class II is formed by the *non-invertible smooth memoryless nonlinearities*, such as the one depicted in Fig. 3(b).
- Class III is formed by the *non-invertible non-smooth hard clipping memoryless nonlinearities*, such as the one depicted in Fig. 3(c). This class of memoryless nonlinearities has a characteristic of the form

$$g(x) = \begin{cases} x, & |x| \leq U \\ \mathrm{sgn}(x)U, & |x| > U \end{cases} \tag{13}$$

with U defined as the *clipping level*.

For each of these classes of memoryless nonlinearities, a specific optimization method is proposed for solving the optimization problem (9) in a fast and reliable way.

B. Class I: Invertible Memoryless Nonlinearities

In case the memoryless nonlinearity $\mathbf{g}$ in the Hammerstein model is invertible, the precompensation of the linear FIR filter $\mathbf{h}$ and the precompensation of the memoryless nonlinearity $\mathbf{g}$ can be decoupled. In a first step, the linear FIR filter can be precompensated for, resulting in the nonlinearly transformed frame $\mathbf{g}(\mathbf{v}_m^*)$. In a second step, the memoryless nonlinearity $\mathbf{g}$ can be precompensated for by applying the inverse memoryless nonlinearity $\mathbf{g^{-1}}$ to $\mathbf{g}(\mathbf{v}_m^*)$ in order to compute the precompensated input frame $\mathbf{v}_m^*$, i.e.

$$\mathbf{v}_m^* = \mathbf{g^{-1}}(\mathbf{g}(\mathbf{v}_m^*)). \tag{14}$$

[1]In the remainder of this paper, all considered memoryless nonlinearities are assumed to operate on the domain $D = [-1, 1]$.

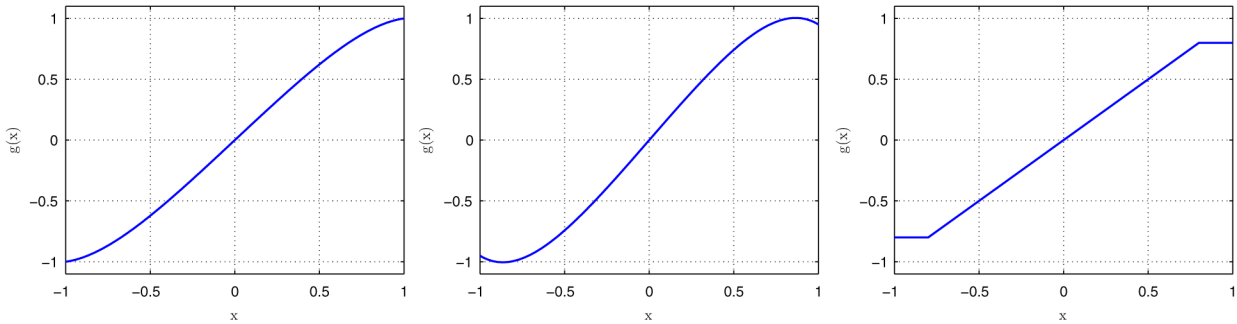

Fig. 3. Examples for different classes of memoryless nonlinearities on the domain $D = [-1, 1]$: (a) Invertible memoryless nonlinearity $g(x) = 1.3x - 0.2x^3 - 0.1x^5$ (b) Non-invertible smooth memoryless nonlinearity $g(x) = 1.6x - 0.4x^3 - 0.25x^5$ (c) Non-invertible non-smooth hard clipping memoryless nonlinearity $g(x) = \max(\min(x, 0.8), -0.8)$.

In order to carry out the first precompensation step, the objective function in (9) can be seen to be a strictly convex quadratic function of the nonlinearly transformed variable $\mathbf{g}(\mathbf{v}_m)$, i.e.

$$\begin{aligned}\mathbf{g}(\mathbf{v}_m^*) &= \underset{\mathbf{g}(\mathbf{v}_m)\in\mathbb{R}^N}{\arg\min}\ f_1(\mathbf{g}(\mathbf{v}_m)) \\ &= \underset{\mathbf{g}(\mathbf{v}_m)\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\mathbf{g}(\mathbf{v}_m)^T\mathbf{A}_m\mathbf{g}(\mathbf{v}_m) + \mathbf{b}_m^T\mathbf{g}(\mathbf{v}_m)\end{aligned} \tag{15}$$

with $\mathbf{A}_m \succ 0$. Indeed, as it was shown that $\mathbf{Q}_m \succ 0$ in [8], it follows that $\mathbf{A}_m = \mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m \succ 0$, since congruence transformations have the property of preserving positive definiteness.

The necessary and sufficient condition for $\mathbf{g}(\mathbf{v}_m^*)$ to be the unique global minimizer of this strictly convex optimization problem, is

$$\nabla f_1(\mathbf{g}(\mathbf{v}_m^*)) = \mathbf{0} \tag{16}$$

from which the following closed form solution $\mathbf{g}(\mathbf{v}_m^*)$ can be derived,

$$\begin{aligned}\mathbf{g}(\mathbf{v}_m^*) &= -\mathbf{A}_m^{-1}\mathbf{b}_m \\ &= -(\mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m)^{-1}(\mathbf{H}_m^T\mathbf{Q}_m(\check{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m)) \\ &= -\mathbf{H}_m^{-1}(\check{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m)\end{aligned} \tag{17}$$

where the symmetry of $\mathbf{Q}_m$ is used and the invertibility of $\mathbf{H}_m$ is assumed to establish the last equality. It can be easily shown that $f_1(\mathbf{g}(\mathbf{v}_m^*)) = 0$, implying that a perfect precompensation $\mathbf{y}_m^* = \mathbf{x}_m$ is achieved. Note that the perceptual weighting $\mathbf{W}_m$ of the frequency domain error terms (which are all zero in the solution) is unable to affect this solution, explaining the absence of $\mathbf{Q}_m$ in (17).

C. Class II: Non-Invertible Smooth Memoryless Nonlinearities

In case the memoryless nonlinearity $\mathbf{g}$ in the Hammerstein model is non-invertible, the precompensation of the linear FIR filter $\mathbf{h}$ and the precompensation of the memoryless nonlinearity $\mathbf{g}$ should be performed simultaneously by solving optimization problem (9). From the smoothness property of $\mathbf{g}$ and its constituting basis functions ψ, it follows that for any $\mathbf{x}_m \in D^N$ the gradient $\nabla\mathbf{g}(\mathbf{x}_m) \in \mathbb{R}^N$ can be computed as

$$\begin{aligned}\nabla\mathbf{g}(\mathbf{x}_m) &= [\nabla g(x_{m,1}), \ldots, \nabla g(x_{m,N})]^T \\ &= \nabla\boldsymbol{\Psi}(\mathbf{x}_m)\mathbf{c}\end{aligned} \tag{18}$$

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

$$\nabla\boldsymbol{\Psi}(\mathbf{x}_m) = \begin{bmatrix} \nabla\psi_1(x_{m,1}) & \ldots & \nabla\psi_P(x_{m,1}) \\ \vdots & \ddots & \vdots \\ \nabla\psi_1(x_{m,N}) & \ldots & \nabla\psi_P(x_{m,N}) \end{bmatrix}. \tag{19}$$

The proposed optimization method for solving the non-convex optimization problem (9) 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 iterative gradient method consists in taking a step along the negative gradient direction,

$$\mathbf{v}_m^{k+1} = \mathbf{v}_m^k - s_m^k\nabla f_0(\mathbf{v}_m^k) \tag{20}$$

where s_m^k is the stepsize and the gradient $\nabla f_0(\mathbf{v}_m^k)$ is computed as

$$\begin{aligned}\nabla f_0(\mathbf{v}_m^k) = \operatorname{diag}(\nabla\mathbf{g}(\mathbf{v}_m^k))\big[\mathbf{H}_m^T\mathbf{Q}_m\big(\mathbf{H}_m\mathbf{g}(\mathbf{v}_m^k) \\ +\check{\mathbf{H}}_m\mathbf{g}(\mathbf{v}_{m-1}^*) - \mathbf{x}_m\big)\big]\end{aligned} \tag{21}$$

and where $\nabla\mathbf{g}(\mathbf{v}_m^k)$ is established using (18)-(19).

The stepsize s_m^k is determined using a backtracking line search for satisfying the Armijo sufficient decrease condition [13],

$$f_0(\mathbf{v}_m^k - s_m^k\nabla f_0(\mathbf{v}_m^k)) \le f_0(\mathbf{v}_m^k) - \beta s_m^k\|\nabla f_0(\mathbf{v}_m^k)\|_2^2 \tag{22}$$

with $\beta \in (0, \frac{1}{2})$ the relaxation of the gradient. Starting from $s_m^k = 1$, successive reductions of s_m^k by a factor $\eta \in (0, 1)$ are performed until the condition (22) holds. The resulting algorithm is given in Algorithm 1.

D. Class III: Non-Invertible Hard Clipping Memoryless Nonlinearities

In case the memoryless nonlinearity $\mathbf{g}$ in the Hammerstein model is a non-invertible hard clipping memoryless nonlinearity with clipping level U as defined in (13), optimization problem (9) can be rewritten as follows,

$$\begin{aligned}\mathbf{v}_m^* &= \underset{\mathbf{v}_m\in\mathbb{R}^N}{\arg\min}\ f_1(\mathbf{v}_m) \quad \text{s.t.} \quad -\mathbf{u} \le \mathbf{v}_m \le \mathbf{u} \\ &= \underset{\mathbf{v}_m\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\mathbf{v}_m^T\mathbf{A}_m\mathbf{v}_m + \mathbf{b}_m^T\mathbf{v}_m \\ &\quad \text{s.t.} \quad -\mathbf{u} \le \mathbf{v}_m \le \mathbf{u}\end{aligned} \tag{23}$$

Algorithm 1 Precompensation of a Hammerstein loudspeaker model with class II memoryless nonlinearity using the gradient optimization method

Input $\mathbf{x}_m \in \mathbb{R}^N$, $\mathbf{v}^*_{m-1} \in \mathbb{R}^N$, $\mathbf{h} \in \mathbb{R}^{L+1}$, $\mathbf{c} \in \mathbb{R}^P$, $\boldsymbol{\psi}(x) \in C(\mathbb{R} \to \mathbb{R}^P)$, $\mathbf{D} \in \mathbb{C}^{N\times N}$, $\alpha \in \mathbb{R}^+$, $\beta \in (0, \frac{1}{2})$, $\eta \in (0,1)$, $K \in \mathbb{N}^+$

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

Compute masking threshold $\mathbf{t}_m$ for $\mathbf{x}_m$ [using [11]]
Compute weights $\mathbf{w}_m$ [using (10)]
Construct $\mathbf{Q}_m = \mathbf{D}^H \mathrm{diag}(\mathbf{w}_m)\mathbf{D}$
Construct $\mathbf{H}_m$ and $\tilde{\mathbf{H}}_m$ from $\mathbf{h}$ [using (4)–(5)]
Initialize $\mathbf{v}^0_m = \mathbf{x}_m$
Initialize $k = 0$
while $k < K$ **do**
  Compute $\nabla f_0(\mathbf{v}^k_m)$ [using (18)-(19) and (21)]
  Initialize $s^k_m = 1$
  while $f_0(\mathbf{v}^k_m - s^k_m \nabla f_0(\mathbf{v}^k_m)) > f_0(\mathbf{v}^k_m) - \beta s^k_m \|\nabla f_0(\mathbf{v}^k_m)\|_2^2$ **do**
    $s^k_m = \eta s^k_m$
  end while
  $\mathbf{v}^{k+1}_m = \mathbf{v}^k_m - s^k_m \nabla f_0(\mathbf{v}^k_m)$
  $k = k + 1$
end while
$\mathbf{v}^*_m = \mathbf{v}^k_m$

where $\mathbf{u} = U\mathbf{1}_N$ contains the per-sample clipping levels and $\mathbf{1}_N \in \mathbb{R}^N$ is an all ones vector. Note that by constraining $\mathbf{v}_m$ to be within the linear range $[-U, U]$ of the memoryless nonlinearity $\mathbf{g}$, it is possible to introduce the equivalence $\mathbf{g}(\mathbf{v}_m) = \mathbf{v}_m$ in the objective function. The strictly convex quadratic objective function ($\mathbf{A}_m \succ 0$) and the affine constraints make optimization problem (23) a strictly convex *quadratic program*.

The proposed optimization method for solving the strictly convex optimization problem (23) is an iterative *projected gradient method*. In each iteration of a projected gradient method, first a step is taken along the negative gradient direction of the objective function, after which the result is orthogonally projected onto the convex feasible set, thereby maintaining feasibility of the iterates. A low computational complexity per iteration is the main asset of projected gradient methods, provided that the orthogonal projection onto the convex feasible set and the gradient of the objective function can easily be computed. The main steps in the $(k+1)$th iteration of the standard projected gradient method can be written as follows:

- Take a step of stepsize s^k_m along the negative gradient direction:

$$\tilde{\mathbf{v}}^{k+1}_m = \mathbf{v}^k_m - s^k_m \nabla f_1(\mathbf{v}^k_m) \tag{24}$$

where the gradient $\nabla f_1(\mathbf{v}^k_m)$ is computed as

$$\nabla f_1(\mathbf{v}^k_m) = \mathbf{H}^T_m \mathbf{Q}_m(\mathbf{H}_m \mathbf{v}^k_m + \tilde{\mathbf{H}}_m \mathbf{v}^*_{m-1} - \mathbf{x}_m). \tag{25}$$

Algorithm 2 Precompensation of a Hammerstein loudspeaker model with class III memoryless nonlinearity using the optimal projected gradient optimization method

Input $\mathbf{x}_m \in \mathbb{R}^N$, $\mathbf{v}^*_{m-1} \in \mathbb{R}^N$, $\mathbf{h} \in \mathbb{R}^{L+1}$, $\mathbf{D} \in \mathbb{C}^{N\times N}$, $U \in (0,1]$, $\alpha \in \mathbb{R}^+$, $\gamma^0_m \in (0,1)$, $K \in \mathbb{N}^+$

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

Compute masking threshold $\mathbf{t}_m$ for $\mathbf{x}_m$ [using [11]]
Compute weights $\mathbf{w}_m$ [using (10)]
Construct convex feasible set $\Omega = \{\mathbf{v}_m \in \mathbb{R}^N | \ -\mathbf{u} \leq \mathbf{v}_m \leq \mathbf{u}\}$
Construct $\mathbf{Q}_m = \mathbf{D}^H \mathrm{diag}(\mathbf{w}_m)\mathbf{D}$
Construct $\mathbf{H}_m$ and $\tilde{\mathbf{H}}_m$ from $\mathbf{h}$ [using (4)–(5)]
Establish Lipschitz constant C_m [using (35)]
Establish convexity parameter μ_m [using (36)]
Compute $\kappa_m = \frac{C_m}{\mu_m}$
Initialize $\mathbf{v}^0_m = \mathbf{c}^0_m = \mathbf{x}_m$
Initialize $k = 0$
while $k < K$ **do**
  $\tilde{\mathbf{v}}^{k+1}_m = \mathbf{c}^k_m - \frac{1}{C_m}\nabla f_1(\mathbf{c}^k_m)$ [using (25)]
  $\mathbf{v}^{k+1}_m = \Pi_\Omega(\tilde{\mathbf{v}}^{k+1}_m)$ [using (27)]
  Calculate γ^{k+1}_m from $(\gamma^{k+1}_m)^2 = (1-\gamma^{k+1}_m)(\gamma^k_m)^2 + \kappa_m \gamma^{k+1}_m$
  $\delta^k_m = \frac{\gamma^k_m(1-\gamma^k_m)}{(\gamma^k_m)^2 + \gamma^{k+1}_m}$
  $\mathbf{c}^{k+1}_m = \mathbf{v}^{k+1}_m + \delta^k_m(\mathbf{v}^{k+1}_m - \mathbf{v}^k_m)$
  $k = k + 1$
end while
$\mathbf{v}^*_m = \mathbf{v}^k_m$

- Project $\tilde{\mathbf{v}}^{k+1}_m$ orthogonally onto the convex feasible set $\Omega = \{\mathbf{v}_m \in \mathbb{R}^N | \ -\mathbf{u} \leq \mathbf{v}_m \leq \mathbf{u}\}$ of optimization problem (23),

$$\mathbf{v}^{k+1}_m = \Pi_\Omega(\tilde{\mathbf{v}}^{k+1}_m) = \arg\min_{\mathbf{v}_p \in \Omega} \frac{1}{2}\|\mathbf{v}_p - \tilde{\mathbf{v}}^{k+1}_m\|_2^2 \tag{26}$$

An orthogonal projection $\Pi_\Omega(\tilde{\mathbf{v}}^{k+1}_m)$ onto Ω can be shown to correspond to performing a componentwise hard clipping operation with clipping level U, i.e. $\Pi_\Omega(\tilde{\mathbf{v}}^{k+1}_m) = \mathbf{g}(\tilde{\mathbf{v}}^{k+1}_m) = [g(\tilde{v}^{k+1}_{m,1}), \ldots, g(\tilde{v}^{k+1}_{m,N})]^T$, with

$$g(\tilde{v}^{k+1}_{m,i}) = \begin{cases} \tilde{v}^{k+1}_{m,i}, & |\tilde{v}^{k+1}_{m,i}| \leq U \\ \mathrm{sgn}(\tilde{v}^{k+1}_{m,i})U, & |\tilde{v}^{k+1}_{m,i}| > U. \end{cases} \tag{27}$$

Variants of the standard projected gradient method have been proposed which have better convergence properties. Here, a projected gradient optimization method is adopted that reaches an optimal convergence for the class of convex optimization problems with *strongly convex* objective functions. This method was first proposed in [14] and has been applied in diverse applications, e.g. real-time clipping of audio signals [15]. Let us assume knowledge of the Lipschitz constant C_m of the gradient ∇f_1 and the convexity parameter μ_m of f_1 on the set Ω. Then, in each iteration k of the optimal projected gradient method, a

standard projected gradient step is performed on a weighted sum $\mathbf{c}_m^k$ of two previous feasible iterates $\mathbf{v}_m^k$ and $\mathbf{v}_m^{k-1}$,

$$\mathbf{v}_m^{k+1} = \Pi_\Omega \left(\mathbf{c}_m^k - \frac{1}{C_m} \nabla f_1(\mathbf{c}_m^k) \right) \tag{28}$$

where

$$\mathbf{c}_m^k = \mathbf{v}_m^k + \delta_m^k (\mathbf{v}_m^k - \mathbf{v}_m^{k-1}) \tag{29}$$

where δ_m^k is computed as

$$\delta_m^k = \frac{\gamma_m^{k-1}(1-\gamma_m^{k-1})}{(\gamma_m^{k-1})^2 + \gamma_m^k} \tag{30}$$

and where γ_m^k is computed as the solution of

$$(\gamma_m^k)^2 = (1-\gamma_m^k)(\gamma_m^{k-1})^2 + \kappa_m \gamma_m^k \tag{31}$$

and $\kappa_m = \frac{C_m}{\mu_m}$ is the condition number.

In order to establish C_m and μ_m for optimization problem (23), the next two lemmas are proposed.

Lemma 1: ([14]) Let function f_1 be twice continuously differentiable on set Ω. The gradient ∇f_1 is Lipschitz continuous on set Ω with Lipschitz constant C if and only if

$$\nabla^2 f_1(\mathbf{z}) \preceq C\mathbf{I}, \forall \mathbf{z} \in \Omega \tag{32}$$

Lemma 2: ([14]) Let function f_1 be twice continuously differentiable on set Ω. The function f_1 is strongly convex on set Ω with convexity parameter μ if and only if there exists $\mu > 0$ such that

$$\nabla^2 f_1(\mathbf{z}) \succeq \mu\mathbf{I}, \forall \mathbf{z} \in \Omega \tag{33}$$

Let us denote and order the eigenvalues of $\mathbf{A}_m$ as follows,

$$\lambda_1(\mathbf{A}_m) \leq \lambda_2(\mathbf{A}_m) \leq \cdots \leq \lambda_{N-1}(\mathbf{A}_m) \leq \lambda_N(\mathbf{A}_m). \tag{34}$$

Using Lemma 3.1 and Lemma 3.2, it is then shown that the Lipschitz constant C_m can be computed as the largest eigenvalue of $\mathbf{A}_m$,

$$C_m = \lambda_N(\mathbf{A}_m) = \lambda_N(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m) \tag{35}$$

and that the convexity parameter μ_m can be computed as the smallest eigenvalue of $\mathbf{A}_m$,

$$\mu_m = \lambda_1(\mathbf{A}_m) = \lambda_1(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m). \tag{36}$$

The resulting algorithm is given as Algorithm 2.

E. Algorithmic Complexity Bounds

It is possible to derive polynomial upper and lower bounds on the *algorithmic complexity*, i.e. the number of necessary iterations of the optimization method proposed in Section III-D as a function of the solution accuracy. For the class of convex optimization problems with strongly convex objective functions, bounds on the algorithmic complexity of the projected gradient method used in Algorithm 2 are derived as follows:

Theorem 1: ([14]) *Let f_1 be a strongly convex, continuously differentiable function with Lipschitz constant C_m and convexity parameter $\mu_m = \frac{C_m}{\kappa_m}$ (where the condition number $\kappa_m > 1$), and let Ω be a convex feasible set. Then the minimum and maximum number of necessary iterations of the optimal projected gradient method described in Algorithm 2 to find a solution $\mathbf{v}_m^k$ satisfying $f_1(\mathbf{v}_m^k) - f_1(\mathbf{v}_m^*) \leq \epsilon$ are given by*

$$k_m^{\min} = \frac{\sqrt{\kappa_m} - 1}{4} \log \left(\frac{\mu_m \|\mathbf{v}_m^0 - \mathbf{v}_m^*\|^2}{2\epsilon} \right) \tag{37}$$

$$k_m^{\max} = \sqrt{\kappa_m} \log \left(\frac{C_m \|\mathbf{v}_m^0 - \mathbf{v}_m^*\|^2}{\epsilon} \right). \tag{38}$$

It has been shown that congruence transformations like $\mathbf{Q}_m \rightarrow \mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m$ do preserve the signs but not the magnitudes of the matrix eigenvalues. Interestingly, bounds on the change in magnitude of the eigenvalues under a congruence transformation have been established,

Theorem 2: ([16]) Let $\mathbf{Q}_m \in \mathbb{R}^{N\times N}$ be symmetric and $\mathbf{H}_m \in \mathbb{R}^{N\times N}$ nonsingular. Let the eigenvalues of $\mathbf{Q}_m$ and $\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m$ be arranged in nondecreasing order (34). *Let $\sigma_1(\mathbf{H}_m) \geq \cdots \geq \sigma_N(\mathbf{H}_m) > 0$ be the singular values of $\mathbf{H}_m$. Then each eigenvalue $\lambda_i(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m)$ $i = 1, \ldots, N$, is bounded as follows,*

$$\sigma_N^2(\mathbf{H}_m)\lambda_i(\mathbf{Q}_m) \leq \lambda_i(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m) \leq \sigma_1^2(\mathbf{H}_m)\lambda_i(\mathbf{Q}_m). \tag{39}$$

Corollary 1: The condition number $\kappa(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m) = \frac{\lambda_N(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m)}{\lambda_1(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m)}$ is bounded as follows,

$$\kappa(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m) \geq \max\left(1, \kappa^{-1}(\mathbf{H}_m^T \mathbf{H}_m)\kappa(\mathbf{Q}_m)\right) \tag{40}$$

$$\kappa(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m) \leq \kappa(\mathbf{H}_m^T \mathbf{H}_m)\kappa(\mathbf{Q}_m). \tag{41}$$

The previous theorem and its corollary now allow to rewrite the lower and upper iteration bounds (37)–(38) as a function of the extreme eigenvalues and singular values of $\mathbf{H}_m$ and $\mathbf{Q}_m$.

Theorem 3: The minimum and maximum number of necessary iterations of the optimal projected gradient method described in Algorithm 2 to find $\mathbf{v}_m^k$ satisfying $f_1(\mathbf{v}_m^k) - f_1(\mathbf{v}_m^*) \leq \epsilon$ are given by

$$k_m^{\min} = \frac{\sqrt{\max\left(1, \kappa^{-1}(\mathbf{H}_m^T \mathbf{H}_m)\kappa(\mathbf{Q}_m)\right)} - 1}{4} \log \left(\frac{\lambda_1(\mathbf{H}_m^T \mathbf{H}_m)\lambda_1(\mathbf{Q}_m)\|\mathbf{v}_m^0 - \mathbf{v}_m^*\|^2}{2\epsilon} \right) \tag{42}$$

$$k_m^{\max} = \sqrt{\kappa(\mathbf{H}_m^T \mathbf{H}_m)\kappa(\mathbf{Q}_m)} \log \left(\frac{\lambda_N(\mathbf{H}_m^T \mathbf{H}_m)\lambda_N(\mathbf{Q}_m)\|\mathbf{v}_m^0 - \mathbf{v}_m^*\|^2}{\epsilon} \right). \tag{43}$$

Proof: In the lower iteration bound (37) established in Theorem 1, incorporate the lower bound (39) on $\mu_m = \lambda_1(\mathbf{H}_m^T \mathbf{Q}_m \mathbf{H}_m)$ established in Theorem 2, and

TABLE I
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	[17]
2	rhcp.wav	polyphonic	instrumental	rock	9.8	468996	900000	[18]
3	pierle.wav	polyphonic	instrumental+vocal	pop	11.7	2234000	2750000	[19]
4	chopin.wav	monophonic	instrumental	classical	17.8	50000	836200	[20]
5	kraftw.wav	polyphonic	instrumental	electronic	17.2	7480000	8240000	[21]
6	crefsax.wav	monophonic	instrumental	classical	10.9	1	479026	[22]
7	grefcla.wav	monophonic	instrumental	classical	6.9	1	302534	[22]
8	arefcla.wav	monophonic	instrumental	classical	7.4	1	302534	[22]

the lower bound (40) on $\kappa_m = \kappa(\mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m)$ established in Corollary 1. This yields the lower iteration bound (42). Analogously, in the upper iteration bound (38) established in Theorem 1, incorporate the upper bound (39) on $C_m = \lambda_N(\mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m)$ established in Theorem 2, and the upper bound (41) on $\kappa_m = \kappa(\mathbf{H}_m^T\mathbf{Q}_m\mathbf{H}_m)$ established in Corollary 1. This yields the upper iteration bound (43). □

IV. AUDIO QUALITY EVALUATION

The proposed loudspeaker precompensation algorithm is evaluated in terms of the resulting audio quality. Section IV-A reports audio quality evaluation experiments using synthetic Hammerstein loudspeaker models, comprising different classes of memoryless nonlinearities. Section IV-B reports audio quality evaluation experiments using audio signals played back through real loudspeakers. This requires an off-line identification of the Hammerstein loudspeaker model parameters, which is discussed in Section IV-C.

A. Audio Quality Evaluation Using Synthetic Hammerstein Loudspeaker Models

Evaluation experiments were performed using two synthetic Hammerstein loudspeaker models. The first Hammerstein model consists of a non-invertible smooth memoryless nonlinearity (class II), followed by a linear FIR filter. The second Hammerstein model consists of a non-invertible hard clipping memoryless nonlinearity (class III), followed by the same linear FIR filter. The following specifications apply:

- Class II memoryless nonlinearity: $P = 3$ basis functions $\boldsymbol{\psi}(x) = [x\ x^3\ x^5]^T$, a corresponding coefficient vector $\mathbf{c} = [1.6\ -0.4\ -0.25]^T$, depicted in Fig. 3(b).
- Class III memoryless nonlinearity: of the form (13), with clipping level $U = 0.8$, depicted in Fig. 3(c).
- Linear FIR filter: $L = 128$, 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 I for details). Each audio signal in the test database was fed into each of the considered Hammerstein loudspeaker models, once with and once without performing precompensation. The following settings were used in the precompensation:

- First Hammerstein model: precompensation using Algorithm 1, with $N = 512$, $\alpha = 0.01$, $\beta = 0.1$, $\eta = 0.6$, and $K = 500$.
- Second Hammerstein model: precompensation using Algorithm 2, with $N = 512$, $\alpha = 0.01$, $\gamma_m^0 = \sqrt{\frac{\mu_m}{C_m}}$, and $K = 500$.

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

$$\Delta\text{ODG} = \text{ODG}(\mathbf{x}, \mathbf{y}^*) - \text{ODG}(\mathbf{x}, \mathbf{y}) \tag{44}$$

where $\mathbf{x}$ is the input signal, $\mathbf{y}$ is the output signal when no precompensation is applied, $\mathbf{y}^*$ is the output signal when precompensation is applied, and $\text{ODG}(\mathbf{r}, \mathbf{d})$ is an objective measure [22] that 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.

The simulations were performed for four distinct average amplitude levels of the input signals, selected such that the corresponding $\text{ODG}(\mathbf{x}, \mathbf{y}) = \{-0.5, -1.0, -1.5, -2.0\}$. This signal-specific selection of the average amplitude levels ensures a uniform audio quality degradation for different output signals when no precompensation is applied.

The resulting ΔODG scores for the first and second Hammerstein models are shown in Fig. 4 and Fig. 5, respectively. For the first Hammerstein model in Fig. 4, a positive audio quality improvement is observed for all audio excerpts, and this for all but one considered average amplitude level. For most audio excerpts, increasing audio quality improvement scores are observed for increasing average amplitude levels. Indeed, it is exactly at higher amplitude levels that the Hammerstein memoryless nonlinearity is severely affecting the audio signal, and that precompensation is capable of improving the resulting audio quality considerably.

For the second Hammerstein model in Fig. 5, a similar pattern is observed with predominantly positive audio quality improvement scores that are increasing with the average amplitude levels. On average, the resulting audio quality improvement is seen to be slightly lower than for the first Hammerstein model. This difference can be attributed to the non-smoothness of the memoryless nonlinearity, which makes the second Hammerstein model more difficult to precompensate for, compared to the first Hammerstein model which comprises a smooth memoryless nonlinearity.

B. Audio Quality Evaluation Using Real Loudspeakers

Evaluation experiments were performed using two different loudspeakers having the following specifications:

- Loudspeaker 1: Boss MA-12 Micro Monitor, full-range, impedance 6Ω.

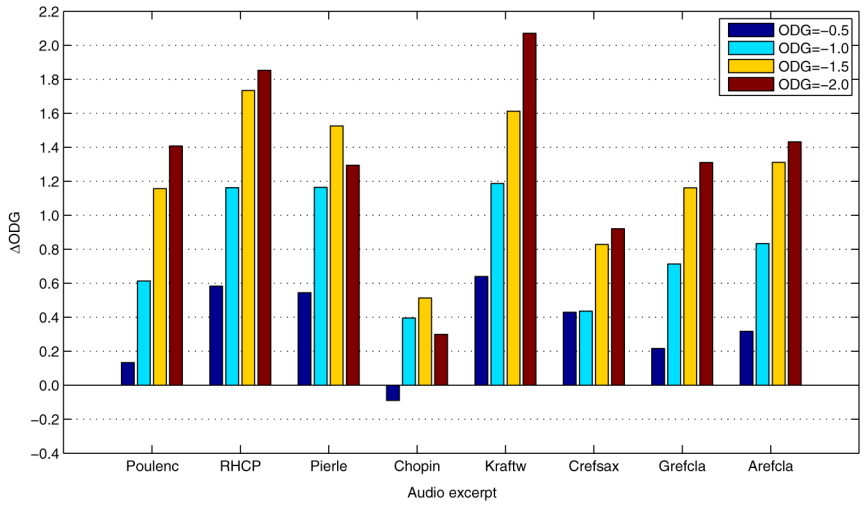

Fig. 4. Precompensation of a Hammerstein model with non-invertible smooth memoryless nonlinearity: objective audio quality improvement scores for different audio excerpts, at four distinct average amplitude levels corresponding to uncompensated input $\mathrm{ODGs} = \{-0.5, -1.0, -1.5, -2.0\}$.

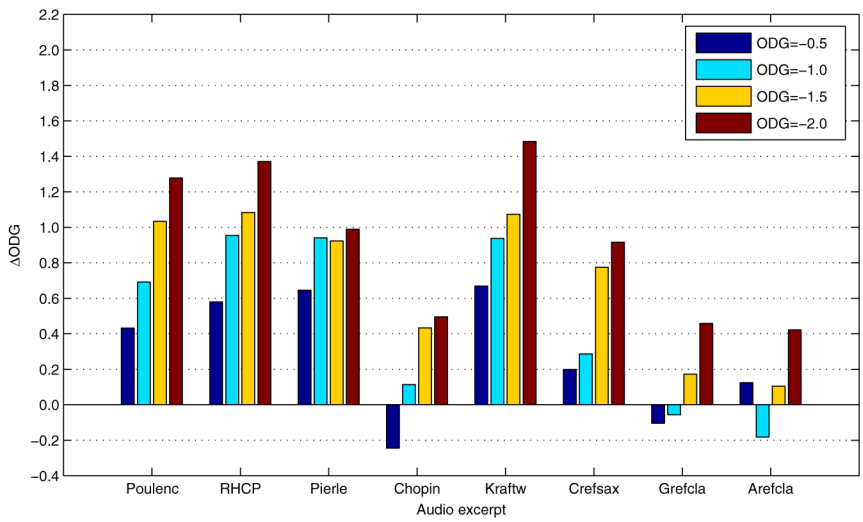

Fig. 5. Precompensation of a Hammerstein model with non-invertible hard clipping memoryless nonlinearity: objective audio quality improvement scores for different audio excerpts, at four distinct average amplitude levels corresponding to uncompensated input $\mathrm{ODGs} = \{-0.5, -1.0, -1.5, -2.0\}$.

- Loudspeaker 2: Dell Latitude E6400 laptop built-in speaker.

The loudspeakers were modeled using Hammerstein models consisting of a memoryless nonlinearity with $P = 3$ basis functions $\boldsymbol{\psi}(x) = [x\ x^3\ x^5]^T$ and a linear FIR filter of order $L = 128$. The identification of the loudspeaker Hammerstein model parameters $\mathbf{h} \in \mathbb{R}^{L+1}$ and $\mathbf{c} \in \mathbb{R}^P$ was performed using the normalized iterative algorithm proposed in [23], and detailed in Section IV-C. The identified Hammerstein models with parameter values $\{\hat{\mathbf{h}}, \hat{\mathbf{c}}\}$ for both loudspeakers are shown in Fig. 6, with Fig. 6(a) depicting the identified memoryless nonlinearity and Fig. 6(b) depicting the identified FIR filter frequency response. From Fig. 6(a), we observe that both loudspeakers possess a non-invertible smooth memoryless nonlinearity (class II). The memoryless nonlinearity of Loudspeaker 2 is seen to deviate from $g(x) = x$ over a wider amplitude input range than the memoryless nonlinearity of Loudspeaker 1. From Fig. 6(b), we observe that for both loudspeakers, the identified FIR filter attenuates low frequencies, this effect being more pronounced for Loudspeaker 2.

Each audio signal in the test database (see Table I) was played back by each of the two loudspeakers, once with and once without performing precompensation. The following settings were used in the precompensation using Algorithm 1:

- Identified Hammerstein model parameters $\mathbf{h}_m = \hat{\mathbf{h}}, \mathbf{c} = \hat{\mathbf{c}}$ are used.
- $N = 512$, $\alpha = 0.01$, $\beta = 0.1$, $\eta = 0.6$, and $K = 200$.

The simulations were performed for three distinct playback levels of the input signals, with relative levels $a = \{0.50, 0.75, 1.00\}$. The objective audio quality improvement for each audio signal was assessed by computing the $\Delta\mathrm{ODG}$ measure,

$$\Delta\mathrm{ODG} = \mathrm{ODG}(\mathbf{x}, \mathbf{y}^*) - \mathrm{ODG}(\mathbf{x}, \mathbf{y}) \tag{45}$$

where $\mathbf{x}$ is the input signal, $\mathbf{y}$ is the measured loudspeaker output signal when no precompensation is applied, and $\mathbf{y}^*$ is the measured loudspeaker output signal when precompensation is applied. The signals were played back in a soundproof room and the loudspeaker output signals were measured using a

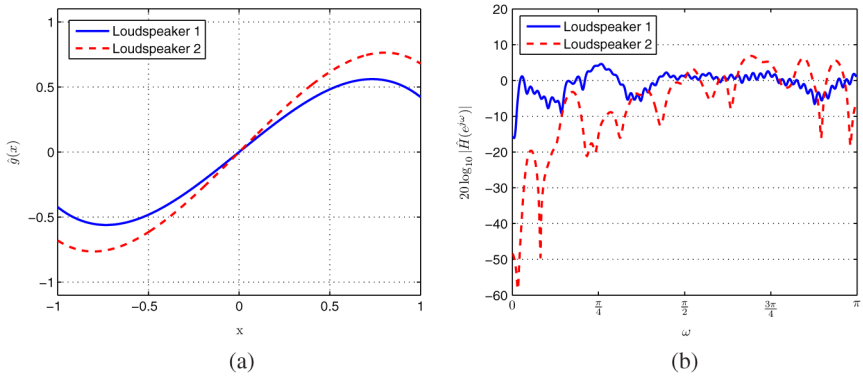

Fig. 6. Identified Hammerstein models for Loudspeaker 1 (solid lines) and Loudspeaker 2 (dashed lines): (a) memoryless nonlinearity (b) FIR filter frequency response.

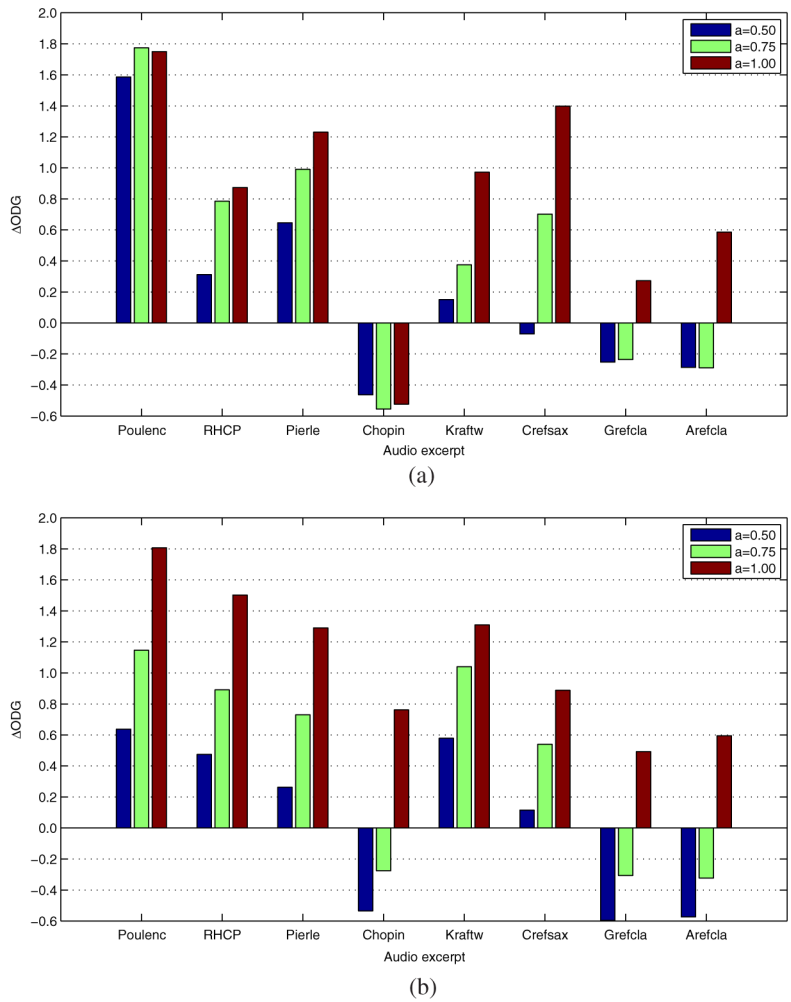

Fig. 7. Precompensation of two loudspeakers using Hammerstein loudspeaker models - objective audio quality improvement scores for different audio excerpts, at average amplitude levels $= \{0.50, 0.75, 1.00\}$: (a) Loudspeaker 1 (b) Loudspeaker 2.

microphone placed on the axis of the loudspeaker at a distance of 10 centimeter.

The resulting ΔODG scores for Loudspeaker 1 and Loudspeaker 2 are shown in Fig. 7(a) and Fig. 7(b), respectively. For Loudspeaker 1 in Fig. 7(a), positive audio quality improvement scores are observed for most audio excerpts. Increasing audio quality improvement scores are observed for increasing playback levels. The negative improvement scores for audio excerpts Chopin (all playback levels), Grefcla and Arefcla (low playback levels) seem to indicate that there is a signal-specific threshold on the loudspeaker playback level below which performing precompensation is unnecessary and can even harm the resulting audio quality.

For Loudspeaker 2 in Fig. 7(b), a similar pattern is observed with audio quality improvement scores that are increasing with the playback level. For the highest playback level $a = 1.00$, all precompensated audio excerpts result in an improved audio quality score. For the lower playback level, positive audio

quality improvement scores are observed, except for the audio excerpts Chopin, Grefcla and Arefcla mentioned previously.

C. Identification of Hammerstein Model Parameters

The purpose of identification is to estimate the unknown Hammerstein model parameters $\mathbf{h}_L = [h_1, \ldots, h_L]^T$ (excluding h_0) and $\mathbf{c} = [c_1, \ldots, c_P]^T$ based on an observed output signal $\mathbf{y} = [y_1, \ldots, y_N]^T$ and an observed input signal $\mathbf{x} = [x_1, \ldots, x_N]^T$. In our experimental set-up, we have used a test audio signal having 30 seconds duration as the loudspeaker input signal, and measured the loudspeaker output signal in a soundproof room using a microphone placed on the axis of the loudspeaker at a distance of 10 centimeter. The estimation is done by solving the following optimization problem,

$$\begin{aligned}\{\hat{\mathbf{h}}_L, \hat{\mathbf{c}}\} &= \arg\min_{\mathbf{h}_L, \mathbf{c}} f_2(\mathbf{h}_L, \mathbf{c}) \\ &= \arg\min_{\mathbf{h}_L, \mathbf{c}} \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \mathbf{h}_L^T \mathbf{\Psi}(\mathbf{x}_n)\mathbf{c}\right)^2 \end{aligned} \quad (46)$$

where $\mathbf{x}_n = [x_{n-1}, \ldots, x_{n-L}]^T$ and consequently $\mathbf{\Psi}(\mathbf{x}_n) = [\boldsymbol{\psi}(x_{n-1}), \ldots, \boldsymbol{\psi}(x_{n-L})]^T$. Given an initial estimate $\hat{\mathbf{h}}_L^0 \neq \mathbf{0}$, the normalized iterative algorithm [23] solves optimization problem (46) by performing the following steps in each iteration k:

1) An estimate $\bar{\mathbf{c}}^k$ is computed as the solution of the least-squares problem (46) for fixed $\mathbf{h}_L = \hat{\mathbf{h}}_L^{k-1}$,

$$\bar{\mathbf{c}}^k = \arg\min_{\mathbf{c} \in \mathbb{R}^P} f_2(\hat{\mathbf{h}}_L^{k-1}, \mathbf{c}). \quad (47)$$

2) An estimate $\bar{\mathbf{h}}_L^k$ is computed as the solution of the least-squares problem (46) for fixed $\mathbf{c} = \bar{\mathbf{c}}^k$,

$$\bar{\mathbf{h}}_L^k = \arg\min_{\mathbf{h}_L \in \mathbb{R}^L} f_2(\mathbf{h}_L, \bar{\mathbf{c}}^k). \quad (48)$$

3) A normalization is performed so that $||\hat{\mathbf{h}}_L^k|| = 1$,

$$\hat{\mathbf{h}}_L^k = \frac{\xi^k \bar{\mathbf{h}}_L^k}{||\bar{\mathbf{h}}_L^k||} \quad (49)$$

$$\hat{\mathbf{c}}^k = \xi^k \bar{\mathbf{c}}^k ||\bar{\mathbf{h}}_L^k|| \quad (50)$$

where $\xi^k = \pm 1$ is the sign of the first non-zero entry of $\bar{\mathbf{h}}_L^k$.

V. Conclusions

In this paper, an embedded-optimization-based loudspeaker precompensation algorithm using a Hammerstein loudspeaker model has been presented. The loudspeaker precompensation consists in a per-frame signal optimization. In order to minimize the perceptible distortion incurred in the loudspeaker, a perceptually meaningful optimization criterion has been constructed. Depending on the invertibility and the smoothness of the memoryless nonlinearity, different first order optimization methods have been proposed for solving the per-frame optimization problems. Objective evaluation experiments using synthetic and identified loudspeaker models have shown that the proposed loudspeaker precompensation algorithm provides a significant audio quality improvement, especially so at high playback levels.

References

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

[2] C.-T. Tan, B. C. J. Moore, and N. Zacharov, "The effect of nonlinear distortion on the perceived quality of music and speech signals," *J. Audio Eng. Soc.*, vol. 51, no. 11, pp. 1012–1031, Nov. 2003.

[3] B. C. 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 Eng. Soc.*, vol. 52, no. 12, pp. 1228–1244, 2004.

[4] M. Karjalainen and T. Paatero, "Equalization of loudspeaker and room responses using Kautz filters: Direct least squares design," *EURASIP J. Adv. Signal Process.*, vol. 2007, pp. 1–13, 2007.

[5] M. V. K. Shi and G. T. Zhou, "Compensation for nonlinearity in a hammerstein system using the coherence function with application to nonlinear acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5853–5858, Dec. 2007.

[6] B. Defraene, T. van Waterschoot, M. Diehl, and M. Moonen, "Perception-based nonlinear loudspeaker compensation through embedded convex optimization," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 2472–2476.

[7] B. Defraene, T. van Waterschoot, M. Diehl, and M. Moonen, "Embedded-optimization-based loudspeaker compensation using a generic hammerstein loudspeaker model," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, Sep. 2013.

[8] B. Defraene, T. van Waterschoot, H. J. Ferreau, M. Diehl, and M. Moonen, "Real-time perception-based clipping of audio signals using convex optimization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2657–2671, Dec. 2012.

[9] E. Terhardt, "Calculating virtual pitch," *Hear. Res.*, vol. 1, no. 2, pp. 155–182, 1979.

[10] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.

[11] 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.

[12] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, 2nd ed. New York, NY, USA: Springer, 1999.

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

[14] Y. Nesterov, *Introductory lectures on convex optimization*. New York, NY, USA: Springer, 2004.

[15] B. Defraene, T. van Waterschoot, M. Diehl, and M. Moonen, "A fast projected gradient optimization method for real-time perception-based clipping of audio signals," in *Proc. ICASSP '11*, Prague, Czech Republic, May 2011, pp. 333–336.

[16] A. Ostrowski, "A quantitative formulation of Sylvester's law of inertia," in *Proc. Nat. Acad. Sci. U.S.A.*, 1959, vol. 45, no. 5, p. 740.

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

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

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

[20] F. Chopin, *Waltz op. 69 no. 2*, 1995, Favourite Piano Works (Vladimir Ashkenazy), Decca 0289 4448302.

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

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

[23] E.-W. Bai and D. Li, "Convergence of the iterative Hammerstein system identification algorithm," *IEEE Trans. Autom. Control*, vol. 49, no. 11, pp. 1929–1940, Nov. 2004.



Bruno Defraene (S'10–M'14) was born in Halle, Belgium, in 1986. He received the M.Sc. degree (2009) and Ph.D. degree (2013) in electrical engineering from KU Leuven, Belgium. He has been a researcher at the Electrical Engineering Department of KU Leuven from 2009-2013, and has been a member of KU Leuvens Optimization in Engineering Center (OPTEC) throughout the same period. His research interests include audio and speech signal processing, acoustical signal enhancement, audio quality assessment, and optimization for signal processing applications. Since 2014, he is working as a Voice Researcher with NXP Software in Leuven, Belgium.



Toon van Waterschoot (S'04–M'12) received the M.Sc. degree (2001) and the Ph.D. degree (2009) in electrical engineering, both from KU Leuven, Belgium. He is currently a part-time Assistant Professor at KU Leuven, Belgium, and a part-time Postdoctoral Research Fellow of the Research Foundation - Flanders (FWO), Belgium. He has previously held positions as a Teaching Assistant with the Antwerp Maritime Academy, Belgium (2002), as a Research Assistant with KU Leuven, Belgium (2002–2009) and with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT), Belgium (2003-2007), and as a Postdoctoral Research Fellow at KU Leuven, Belgium (2009–2010) and at Delft University of Technology, The Netherlands (2010–2011). Since 2005, he has been a Visiting Lecturer at the Advanced Learning and Research Institute of the University of Lugano (Universit della Svizzera Italiana), Switzerland. He has been teaching courses related to digital signal processing and control theory. His research interests are in signal processing, machine learning, and numerical optimization, with application to acoustic modeling, audio analysis, and acoustic signal enhancement. Dr. van Waterschoot is serving as an Associate Editor for the J. the Audio Engineering Society and for the EURASIP Journal on Audio, Music, and Speech Processing, and as a Guest Editor for Signal Processing. He has been a Nominated Officer for the European Association for Signal Processing (EURASIP), and a Scientific Coordinator of the FP7-PEOPLE Marie Curie Initial Training Network on Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS). He has been serving as an Area Chair for Speech Processing at the European Signal Processing Conference (EUSIPCO 2010, 2013, and 2014), and will be the General Chair of the 2016 AES Conference to be held in Leuven, Belgium. He is a member of the Audio Engineering Society, the Acoustical Society of America, EURASIP, and IEEE.



Moritz Diehl (M'11) studied physics and mathematics in Heidelberg and Cambridge, and did his Ph.D. in 2001 in Heidelberg on real-time optimization for large scale nonlinear processes. From 2006–2013 he was a professor at KU Leuven where he also served as the principal investigator of the optimization in engineering center OPTEC. Since 2013 he has held the chair of systems theory, control, and optimization at the University of Freiburg, but also keeps an affiliation with KU Leuven. His research interests are in optimization methods for estimation and control, in particular real-time optimization on embedded hardware, and cover theory, software development as well as applications, particularly in robotics, mechatronics, and aerospace, in automotive, chemical and microsystems engineering, and in renewable energy systems.



Marc Moonen (M'94–SM'06–F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, wireless communications, DSL and audio signal processing. He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received a journal best paper award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING (with Geert Leus) and from Elsevier Signal Processing (with Simon Doclo). He was chairman of the IEEE Benelux Signal Processing Chapter (1998-2002), a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing, 2007-2008 and 2011–2012). He has served as Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003-2005), and has been a member of the editorial board of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, IEEE SIGNAL PROCESSING MAGAZINE, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking, and Signal Processing. He is currently a member of the editorial board of EURASIP Journal on Applied Signal Processing and Area Editor for Feature Articles in IEEE SIGNAL PROCESSING MAGAZINE.