Citation/Reference	Defraene B., van Waterschoot T., Diehl M., Moonen M., ``Subjective audio quality evaluation of embedded-optimization-based distortion precompensation algorithms'', *JASA Express Letters*, vol. 140, no. 1, Jul. 2016, pp. 101-106
Archived version	Final publisher's version / pdf
Published version	http://dx.doi.org/10.1121/1.4955025
Journal homepage	http://scitation.aip.org/content/asa/journal/jasael.
Author contact	Toon.vanwaterschoot@kuleuven.be Klik hier als u tekst wilt invoeren.
IR	https://lirias.kuleuven.be/handle/123456789/533050

Subjective audio quality evaluation of embedded-optimization-based distortion precompensation algorithms

Bruno Defraene,[a] Toon van Waterschoot,[b] Moritz Diehl,[c] and Marc Moonen

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

bruno.defraene@gmail.com, toon.vanwaterschoot@esat.kuleuven.be, moritz.diehl@imtek.uni-freiburg.de, marc.moonen@esat.kuleuven.be

Abstract: Subjective audio quality evaluation experiments have been conducted to assess the performance of embedded-optimization-based precompensation algorithms for mitigating perceptible linear and nonlinear distortion in audio signals. It is concluded with statistical significance that the perceived audio quality is improved by applying an embedded-optimization-based precompensation algorithm, both in case (i) nonlinear distortion and (ii) a combination of linear and nonlinear distortion is present. Moreover, a significant positive correlation is reported between the collected subjective and objective PEAQ audio quality scores, supporting the validity of using PEAQ to predict the impact of linear and nonlinear distortion on the perceived audio quality.

[CFG]

Date Received: September 27, 2015 **Date Accepted:** February 15, 2016

1. Introduction

Audio signal distortion introduced by, e.g., non-ideal recording, transmission, or reproduction devices, has been reported in numerous studies to negatively affect the perceived audio quality. *Linear* distortion involves changes in the relative amplitudes and phases of the frequency components constituting the original audio signal, and is perceived as changing the timbre or coloration of the audio signal (Moore and Tan, 2003). *Nonlinear* distortion involves the introduction of frequency components that are not present in the original audio signal, and is perceived as harshness or noisiness, or as crackles and clicks (Tan *et al.*, 2003).

In order to mitigate the perceptible effects of audio signal distortion, a *precompensation* algorithm can be applied to the audio signal *before* the distortion is introduced, e.g., prior to reproduction through a distorting loudspeaker. This approach typically requires *a priori* knowledge of a model for the distortion process. Amongst popular audio signal distortion models are linear finite impulse response (FIR) filters for modeling linear distortion processes, memoryless nonlinearities for modeling nonlinear distortion processes, as well as cascades of these two into Hammerstein or Wiener models for modeling a combination of linear and nonlinear distortion (Lashkari, 2005).

This paper focuses on a recently proposed class of audio signal distortion precompensation algorithms, which are based on so-called *embedded optimization*. Particular to this approach is that the signal precompensation problem is formulated and solved as a per-frame numerical optimization problem aimed at maximizing the resulting audio quality, which can be achieved by properly including a psychoacoustic model in the objective function. By applying embedded-optimization-based precompensation algorithms significant improvements have been reported in terms of *objective* measures of audio quality [including PEAQ (International Telecommunications Union, 1998)], both in case (i) nonlinear distortion (Defraene *et al.*, 2012) and (ii) a combination of linear and nonlinear distortion (Defraene *et al.*, 2014) is present. Because of the

[a]Author to whom correspondence should be addressed.

[b]Also at: ESAT-ETC, Advanced Integrated Sensing Lab (AdvISe), Department of Electrical Engineering, KU Leuven, Kleinhoefstraat 4, 2440 Geel, Belgium.

[c]Also at: Institute of Microsystems Engineering (IMTEK), University of Freiburg, Georges-Köhler-Allee 102, D-79110 Freiburg, Germany.

limited applicability and accuracy of objective audio quality measures, a formal subjective listening test has been performed to properly evaluate the performance of these algorithms. The goal of the subjective listening test is two-fold. The first goal is to assess the *subjective* audio quality improvement of applying embedded-optimization-based precompensation algorithms, both for audio signals subject to nonlinear distortion, and subject to a combination of linear and nonlinear distortion. The second goal is to assess the correlation between the objective and subjective audio quality scores, thus assessing the validity of using objective audio quality measures for predicting the impact of linear and nonlinear distortion on the perceived audio quality.

This paper is organized as follows. In Sec. 2, the research hypotheses are formulated. In Sec. 3, the experimental design and setup of the subjective listening test are discussed. In Sec. 4, the test results are reported and the formulated hypotheses are statistically tested. In Sec. 5, some concluding remarks are presented.

2. Research hypotheses

The research hypotheses, that may or may not be rejected based on the outcome of the subjective listening test, are formulated as follows:

- Hypothesis 1: The perceived audio quality of audio signals with and without embedded-optimization-based precompensation prior to a nonlinear distortion process, is identical.
- Hypothesis 2: The perceived audio quality of audio signals with and without embedded-optimization-based precompensation prior to a combined linear and nonlinear distortion process, is identical.
- Hypothesis 3: There is no correlation between subjective perceived audio quality scores and objective perceived audio quality scores for audio signals subject to linear and nonlinear distortion.

3. Methods

3.1 Participants

A representative group of 19 test subjects having considerable musical listening and performance experience was selected to perform the listening test. All subjects were remunerated for their participation.

3.2 Stimuli

The stimuli presented to the test subjects consisted of 4 audio excerpts (detailed in Table 1), each of which were presented in 12 different processing scenarios:

- Processing scenarios S_1–S_3: Uncompensated symmetrical hard clipping nonlinearity, where the clipping level is selected such that the processed audio signal has an Objective Difference Grade (ODG) using the PEAQ Basic Model (International Telecommunications Union, 1998) of -1, -2, and -3, for the respective processing scenarios S_1, S_2, and S_3.
- Processing scenarios S_4–S_6: Precompensated symmetrical hard clipping nonlinearity using the embedded-optimization-based precompensation algorithm proposed in Defraene *et al*. (2012), with parameter values $N = 512$, $P = 128$, $\alpha = 0.04$, and the same clipping level U as used in the respective processing scenarios S_1, S_2, and S_3.
- Processing scenarios S_7–S_9: Uncompensated Hammerstein model consisting of
  - (i) Symmetrical hard clipping nonlinearity with the same clipping level U as used in the respective processing scenarios S_1, S_2, and S_3.
  - (ii) Linear FIR filter ($L = 128$) with impulse response $h[n]$ designed using the frequency sampling method `fir2` in MATLAB, having a required magnitude response $[1, 0.95, 0.75, 0.50, 0.20, 0]^T$ at the frequencies $[0, 0.2, 0.4, 0.6, 0.8, 1]^T \times f_{\text{Nyquist}}$.
- Processing scenarios S_{10}–S_{12}: Precompensated Hammerstein model, with the same Hammerstein model settings as in the respective processing scenarios S_7, S_8, and S_9,

Table 1. Audio excerpts used for subjective audio quality evaluation. Musical texture is either monophonic (single melodic line) or polyphonic (multiple melodic lines).

Nr.	Name	Texture	Style	Duration [s]
1	rhcp.wav	polyphonic	rock	9.8
2	chopin.wav	monophonic	classical	17.8
3	poulenc.wav	polyphonic	classical	17.8
4	crefsax.wav	monophonic	classical	10.9

and using the embedded-optimization-based precompensation algorithm proposed in Defraene *et al.* (2014), with parameter values $N = 512$, $\alpha = 0.01$, $\gamma_m^0 = \sqrt{\mu_m / C_m}$, $K = 500$.

It should be noted that the use of idealistic distortions rather than real-life loudspeaker distortions in this experiment allows to accurately control the objective distortion level, achieving a given objective PEAQ ODG score by setting the clipping level. Likewise, the FIR filter was also designed to have a significant impact on the objective distortion level, rather than to optimally reflect an actual loudspeaker response. This objective distortion level would be much more difficult to control for real-life loudspeaker distortions.

3.3 Procedure

The resulting $N_{\mathrm{ps}} = 4 \times 12 = 48$ pairs of stimuli (each consisting of the original unprocessed audio signal and the corresponding processed audio signal) were presented to the test subjects. For each pair of stimuli, the test subjects were asked to rate the perceived audio quality degradation of the presented processed signal with the original audio signal as a reference, using the ITU-T Degradation Category Rating (DCR) (International Telecommunications Union, 1996) scale depicted in Fig. 1. The listening tests were performed in a soundproof and well-illuminated test room. Stimuli were presented to the test subjects through high-quality circumaural headphones [Sennheiser HD 439 (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany): dynamic, closed transducer, frequency response 17–22 500 Hz, Sound Pressure Level 112 dB, Total Harmonic Distortion <0.1%] connected to a soundcard-equipped laptop [Sony Vaio VGN-CR41 (Sony Corp., Minato, Tokyo, Japan) Intel (Intel Corp., Santa Clara, CA) Core 2 duo T5550 processor @1.83 Ghz, 3 GB RAM, Realtek (Realtek Semiconductor Corp., Hsinchu, Taiwan) sound card]. Self-developed software was used to automate stimulus presentation and response collection. The playback level was fixed at a comfortable level.

Prior to the listening test, the subjects were provided with written instructions, which were verbally reviewed by the experimenter. Before the first pair of stimuli was presented, the subjects were familiarized with the effects of linear and nonlinear distortion on audio signals, by successively listening to an original sample audio signal and its distorted version. The presentation order of the pairs of stimuli was randomized using an altered Latin square scheme (Bech and Zacharov, 2007), thus eliminating possible bias effects due to order effects and sequential dependencies.

4. Results

The listening test had an average duration of 35 min per subject. The raw data resulting from the listening test consists of a categorical DCR response by each of the 19 subjects, for each of the 48 presented pairs of stimuli. Figure 2 shows histograms of the obtained DCR responses for the audio signals having ODG $= -1$ after hard symmetrical clipping (histograms for other ODGs were omitted due to space restrictions). It is observed that the response histograms for the processing scenarios with precompensation [Figs. 2(b) and 2(d)] have a higher probability mass in the two leftmost bins compared to the corresponding response histograms for processing scenarios without precompensation [Figs. 2(a) and 2(c)]. These categorical DCR responses were first converted to integers according to the scale in Fig. 1. The following statistical analysis was performed on the obtained numerical set of DCR responses.

4.1 Testing hypothesis 1

Let us denote the population DCR responses corresponding to audio signals processed by the uncompensated and by the precompensated symmetrical hard clipping nonlinearity by random variables $R_{\mathrm{clip}}^{\mathrm{unc}}$ and $R_{\mathrm{clip}}^{\mathrm{pre}}$, respectively. Based on the sample DCR responses, we tested the following statistical hypothesis H_0^1 against its alternative H_a^1:

$$H_0^1 \,:\, \tilde{R}_{\mathrm{clip}}^{\mathrm{unc}} = \tilde{R}_{\mathrm{clip}}^{\mathrm{pre}}, \tag{1}$$

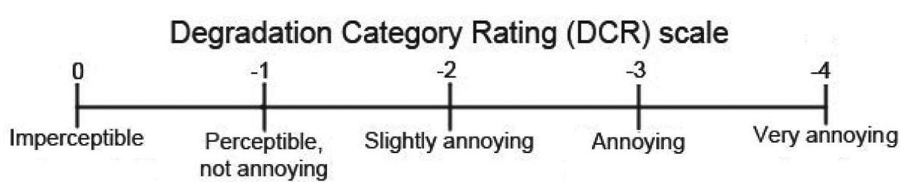

Fig. 1. ITU-T DCR scale (adapted from ITU, 1996).

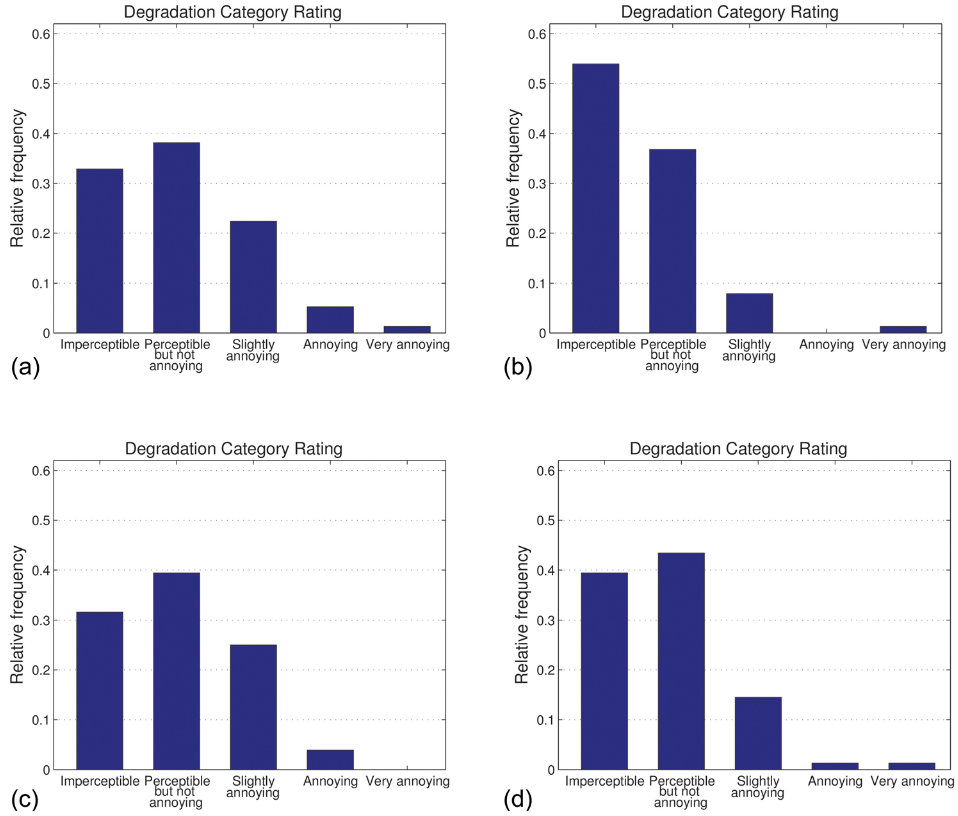

Fig. 2. (Color online) Histograms of DCR responses for audio signals without (left) and with (right) embedded-optimization-based precompensation, for input ODG = −1. (a) Uncompensated symmetrical hard clipping nonlinearity (scenario S_1). (b) Precompensated symmetrical hard clipping nonlinearity (scenario S_4). (c) Uncompensated Hammerstein model (scenario S_7). (d) Precompensated Hammerstein model (scenario S_{10}).

$$H_a^1 \; : \; \tilde{R}_{\text{clip}}^{\text{unc}} < \tilde{R}_{\text{clip}}^{\text{pre}}, \tag{2}$$

where $\tilde{R}$ is the population median of the random variable R. This statistical hypothesis was tested for all three considered ODGs using one-tailed Wilcoxon-Mann-Whitney tests (Wilcoxon, 1945) with significance level $\alpha = 0.05$. The resulting one-sided P-values are synthesized in the first column of Table 2. From the obtained P-values, we conclude that the null hypothesis [Eq. (1)] can be rejected in favor of the alternative [Eq. (2)] for all considered ODGs.

4.2 Testing hypothesis 2

Let us denote the population DCR responses corresponding to audio signals processed by the uncompensated and the precompensated Hammerstein model by random variables $R_{\text{hamm}}^{\text{unc}}$ and $R_{\text{hamm}}^{\text{pre}}$, respectively. Based on the sample DCR responses, we tested the following statistical hypothesis H_0^2 against its alternative H_a^2:

$$H_0^2 \; : \; \tilde{R}_{\text{hamm}}^{\text{unc}} = \tilde{R}_{\text{hamm}}^{\text{pre}}, \tag{3}$$

$$H_a^2 \; : \; \tilde{R}_{\text{hamm}}^{\text{unc}} < \tilde{R}_{\text{hamm}}^{\text{pre}}. \tag{4}$$

This statistical hypothesis was tested for all three considered ODGs using one-tailed Wilcoxon-Mann-Whitney tests with significance level $\alpha = 0.05$. The resulting one-sided P-values are synthesized in the second column of Table 2. From the obtained P-values,

Table 2. P-values from one-tailed Wilcoxon-Mann-Whitney tests on sample DCR responses. Significant P-values with respect to $\alpha = 0.05$ in bold.

Null hypothesis →	H_0^1	H_0^2
ODG = −1	**0.0006**	0.0616
ODG = −2	**<0.0001**	**<0.0001**
ODG = −3	**<0.0001**	**<0.0001**

we conclude that the null hypothesis [Eq. (3)] can be rejected in favor of the alternative [Eq. (4)] for ODGs of −2 and −3.

4.3 Testing hypothesis 3

The PEAQ ODG measure has been designed to objectively assess the perceptibility of degradations commonly encountered in audio codecs. However, the nature of signal distortions introduced by the type of linear and nonlinear distortions under study can be rather different compared to signal distortions introduced by audio codecs. Therefore, we investigate the validity of using PEAQ ODG as an objective audio quality measure in these alternative scenarios. The correlation between subjective and objective scores is the most obvious criterion to validate an objective method. Let us denominate the mean DCR responses over all 19 test subjects as MDCR responses. Then we can calculate the sample Pearson correlation coefficient $\hat{\rho}$ between the subjective MDCR responses and the objective ODG scores as follows:

$$\hat{\rho} = \frac{\sum_{i=1}^{N_{\mathrm{ps}}} \left(\mathrm{MDCR}_i - \overline{\mathrm{MDCR}}\right)\left(\mathrm{ODG}_i - \overline{\mathrm{ODG}}\right)}{\sqrt{\sum_{i=1}^{N_{\mathrm{ps}}} \left(\mathrm{MDCR}_i - \overline{\mathrm{MDCR}}\right)^2}\sqrt{\sum_{i=1}^{N_{\mathrm{ps}}} \left(\mathrm{ODG}_i - \overline{\mathrm{ODG}}\right)^2}}, \tag{5}$$

where

$$\overline{\mathrm{MDCR}} = \sum_{i=1}^{N_{\mathrm{ps}}} \mathrm{MDCR}_i, \tag{6}$$

$$\overline{\mathrm{ODG}} = \sum_{i=1}^{N_{\mathrm{ps}}} \mathrm{ODG}_i. \tag{7}$$

Based on the resulting sample Pearson correlation coefficient value $\hat{\rho} = 0.67$, we tested the following statistical hypothesis H_0^3 against its alternative H_a^3:

$$H_0^3 \,:\, \rho = 0, \tag{8}$$

$$H_a^3 \,:\, \rho > 0, \tag{9}$$

where ρ is the population Pearson correlation coefficient. This statistical hypothesis was tested with significance level $\alpha = 0.05$ by using a one-tailed t-test having $N_{\mathrm{ps}} - 2$ degrees of freedom for the test statistic value $t = |\hat{\rho}|(\sqrt{N_{\mathrm{ps}} - 2}/\sqrt{1 - \hat{\rho}^2})$. The resulting one-sided P-value is $1.206 \times 10^{-7} < \alpha$, which means that the null hypothesis [Eq. (8)] can be confidently rejected in favor of the alternative [Eq. (9)].

5. Conclusions

A subjective evaluation has been conducted to assess the performance of embedded-optimization-based precompensation algorithms for mitigating perceptible linear and nonlinear distortion in audio signals. For audio signals subject to nonlinear distortion, it is concluded that the resulting audio quality is significantly improved by applying an embedded-optimization-based precompensation algorithm, and this for all considered levels of nonlinear distortion. For audio signals subject to a combination of linear and nonlinear distortion, it is concluded that the resulting audio quality is significantly improved by applying an embedded-optimization-based precompensation algorithm, and this for moderate (ODG = −2) to high (ODG = −3) levels of distortion. Moreover, a significant positive correlation has been reported between the subjective and objective PEAQ audio quality scores, supporting the validity of using PEAQ to objectively predict the impact of linear and nonlinear distortion on the perceived audio quality.

Acknowledgments

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 (OPTEC), Concerted Research Action GOA/10/09 MaNet, and the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office: IUAP P7/19 "Dynamical systems control and optimization" (DYSCO) 2012–2017. The scientific responsibility is assumed by its authors.

References and links

Bech, S., and Zacharov, N. (**2007**). *Perceptual Audio Evaluation—Theory, Method and Application* (Wiley, New York).

Defraene, B., van Waterschoot, T., Diehl, M., and Moonen, M. (**2014**). "Embedded-optimization-based loudspeaker precompensation using a Hammerstein loudspeaker model," IEEE Trans. Audio Speech Lang. Process. **22**(11), 1648–1659.

Defraene, B., van Waterschoot, T., Ferreau, H. J., Diehl, M., and Moonen, M. (**2012**). "Real-time perception-based clipping of audio signals using convex optimization," IEEE Trans. Audio Speech Lang. Process. **20**(10), 2657–2671.

International Telecommunications Union (**1996**). "Methods for subjective determination of transmission quality," ITU Recommendation, P. 800.

International Telecommunications Union (**1998**). "Method for objective measurements of perceived audio quality," ITU Recommendation, B. S. 1387.

Lashkari, K. (**2005**). "A modified Volterra-Wiener-Hammerstein model for loudspeaker precompensation," in *Proceedings of the 39th Asilomar Conference on Signals, Systems, and Computers*, pp. 344–348.

Moore, B. C., and Tan, C.-T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am. **114**(1), 408–421.

Tan, C.-T., Moore, B. C., and Zacharov, N. (**2003**). "The effect of nonlinear distortion on the perceived quality of music and speech signals," J. Audio Eng. Soc. **51**(11), 1012–1031.

Wilcoxon, F. (**1945**). "Individual comparisons by ranking methods," Biometrics Bull. **1**(6), 80–83.