**KU Leuven**

Departement Elektrotechniek ESAT-STADIUS/TR 13-35

# Declipping of audio signals using perceptual compressed sensing[1]

Bruno Defraene[2] [3], Naim Mansour[2], Steven De Hertogh[2], Toon van Waterschoot[2], Moritz Diehl[2] and Marc Moonen[2]

December 2013
*IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 12, Dec. 2013, pp. 2627-2637

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/bdefraen/reports/13-35.pdf*

[2]KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 321788, Fax +32 16 321970, WWW: *http://homes.esat.kuleuven.be/∼bdefraen*. E-mail: *bruno.defraene@esat.kuleuven.be*.

[3]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, Flemish Government iMinds 2013, 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.

# Declipping of Audio Signals Using Perceptual Compressed Sensing

Bruno Defraene, *Student Member, IEEE*, Naim Mansour, Steven De Hertogh, Toon van Waterschoot, *Member, IEEE*, Moritz Diehl, *Member, IEEE*, and Marc Moonen, *Fellow, IEEE*

***Abstract*—The restoration of clipped audio signals, commonly known as *declipping*, is important to achieve an improved level of audio quality in many audio applications. In this paper, a novel declipping algorithm is presented, jointly based on the theory of compressed sensing (CS) and on well-established properties of human auditory perception. Declipping is formulated as a sparse signal recovery problem using the CS framework. By additionally exploiting knowledge of human auditory perception, a novel perceptual compressed sensing (PCS) framework is devised. A PCS-based declipping algorithm is proposed which uses $\ell_1$-norm type reconstruction. Comparative objective and subjective evaluation experiments reveal a significant audio quality increase for the proposed PCS-based declipping algorithm compared to CS-based declipping algorithms.**

***Index Terms*—Compressed sensing, declipping, perception, sparsity.**

## I. Introduction

CLIPPING introduces undesired signal distortion in many audio applications. Clipping can occur both in the analog domain and in the digital domain, and is generally caused by the inability of an audio playback, recording or processing device to deliver the dynamic range required by the audio signal. When a digital audio signal is clipped[1], all its sample values lying beyond a maximum amplitude level (referred to as the *clipping level* $\theta_c$) are mapped onto $\pm\theta_c$, as shown in Fig. 1.

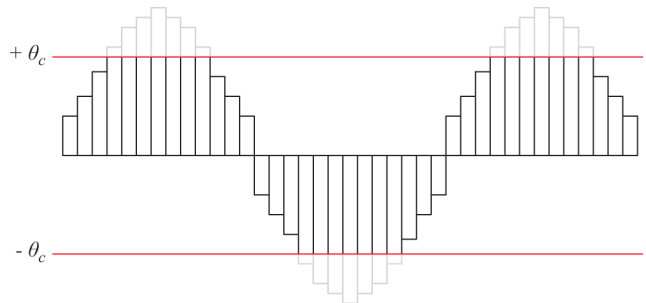



Fig. 1. Clipping of a digital audio signal.

Clipping inevitably introduces nonlinear distortion into the audio signal, consisting of both the *modification* of existing frequency components, and the *introduction* of new harmonic and aliasing frequency components into the signal [1]. In a series of listening experiments [2], it has been concluded that clipping has a significant negative effect on the perceived audio quality. More specifically, clipping is typically associated with the perceptible addition of crackling noises to the original audio signal, often qualified as (very) annoying.

Therefore, performing *declipping*, i.e., the restoration of the clipped audio signal, is necessary to achieve an improved level of audio quality and listener satisfaction. In past research contributions, several approaches to the declipping problem have been proposed. A first approach is based on performing an interpolation/extrapolation step to recover the clipped signal portions based on the knowledge of unclipped signal portions [3]. A second approach consists in adopting a suitable model of both the clean signal (typically autoregressive) and the clipping distortion, and subsequently recovering the clean signal through Bayesian statistical estimation of its model parameters [4]. Other notable declipping approaches include the use of artificial neural networks [5].

Recently, the declipping problem has been adressed in the generic framework of compressed sensing (CS), and encouraging results have been reported [6]–[9]. In the CS framework, declipping is formulated and solved as a sparse signal recovery problem, where one takes advantage of the sparsity of the clean audio signal (in some basis) in order to recover it from a subset of its samples.

[1]In the context of this research, digital *hard* clipping is considered, and it will be simply termed "clipping" throughout this paper. Note that in analogue systems, *soft* clipping is very common.

Manuscript received March 04, 2013; revised June 12, 2013; accepted August 27, 2013. Date of publication September 16, 2013; date of current version October 24, 2013. This work 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, Flemish Government iMinds 2013, 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 Dr. Emmanuel Vincent.

B. Defraene, T. van Waterschoot, M. Diehl, and M. Moonen are with the Department of Electrical Engineering, ESAT-STADIUS, Katholieke Universiteit Leuven, 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).

N. Mansour was with the Department of Electrical Engineering, ESAT-STADIUS, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium. He is now with Alcatel-Lucent, 2018 Antwerp, Belgium (e-mail: naim.mansour@telenet.be).

S. De Hertogh was with the Department of Electrical Engineering, ESAT-STADIUS, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium. He is now with EVS Broadcast Equipment, 4102 Seraing, Belgium (e-mail: stevendehertogh@telenet.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/TASL.2013.2281570

1558-7916 © 2013 IEEE

In this paper, we propose a perceptual compressed sensing (PCS) framework, in which a declipping algorithm is subsequently developed. The PCS framework integrates CS theory and properties of human auditory perception. By incorporating knowledge of human auditory perception in the algorithm design, it is possible to further improve the perceptual reconstruction quality. The advantage of using perceptual knowledge has been recently demonstrated in related audio applications, such as the sparse approximation of audio signals [10], and the limiting of audio signals [11].

The paper is organized as follows. In Section II, the basic principles of CS as well as its fundamental limitations are reviewed, and the declipping problem is subsequently formulated in the CS framework. In Section III, the PCS framework is presented, leading to a PCS-based declipping algorithm, using $\ell_1$-norm type reconstruction. In Section IV, a comparative objective and subjective evaluation of different declipping algorithms is discussed. Finally, in Section V, some concluding remarks are presented.

## II. A CS Framework for Declipping

The theory of compressed sensing has been introduced in [12], [13] [14] in the context of sparse signal sampling and acquisition. Basically, the theory states that any signal that can be considered sparse in some basis, can be uniquely, and in many cases *perfectly* reconstructed based on sub-Nyquist rate sampled measurements. This notion goes against the commonly accepted Nyquist sampling criterion. In the next subsections, we review the basic principles of CS as well as its fundamental limitations, and subsequently formulate the declipping problem in the CS framework.

### A. CS Basic Principles

To be able to outline the basic principles of CS, we first introduce some necessary definitions and notation. Signals are considered to be real-valued vectors in an $N$-dimensional normed Euclidean vector space $\mathbb{R}^N$. For the purposes of CS, mainly the $\ell_0$ and $\ell_1$ norm are of importance. The *sparsity* $k$ of a vector $\mathbf{x}$ is defined as the number of non-zero components, i.e., $\|\mathbf{x}\|_0 = k$. A vector $\mathbf{x}$ with sparsity $k$ is said to be $k$-*sparse*.

Compressed sensing takes advantage of the sparsity of a signal $\mathbf{s} \in \mathbb{R}^N$ in some fixed basis $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ in order to recover it from a reduced measurement $\mathbf{y} \in \mathbb{R}^M$, where $M < N$. Acquiring a signal $\mathbf{s}$ by CS consists of two main steps:

1) *Measurement*: apply a measurement matrix $\mathbf{\Phi}$ to obtain the measurement $\mathbf{y}$,

$$\mathbf{y} = \mathbf{\Phi s} = \mathbf{\Phi\Psi x} = \mathbf{Ax} \tag{1}$$

with $\mathbf{s} \in \mathbb{R}^N$ the original signal, $\mathbf{x} \in \mathbb{R}^N$ its sparse decomposition, $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ the fixed basis[2], $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ the measurement matrix, $\mathbf{A} = \mathbf{\Phi\Psi} \in \mathbb{C}^{M\times N}$ the sensing matrix, and $\mathbf{y} \in \mathbb{R}^M$ the measurement.

[2]In this paper, we restrict the analysis to fixed bases $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$, and do not consider the case of overcomplete dictionaries $\mathbf{\Psi} \in \mathbb{C}^{N\times K_D}$ with $N < K_D$. Overcomplete dictionaries have been used to allow for sparser representations of audio signals [15].

2) *Recovery*: (in)exactly recover $\mathbf{s}'$ from $\mathbf{y}$ using constrained $\ell_p$-norm optimization,

$$\mathbf{x}' = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_p \quad \text{subject to} \quad \mathbf{Az} = \mathbf{y} \tag{2}$$

$$\mathbf{s}' = \mathbf{\Psi x}' \tag{3}$$

with $p \in \{0, 1\}$ in typical CS applications.

The actual choice of the $\ell_p$ norm in the recovery step (2), has considerable implications on the resulting solution $\mathbf{x}'$ as well as on its computation. Firstly, using an $\ell_0$ or $\ell_1$ norm in the minimization problem will obviously lead to a sparser solution $\mathbf{x}'$ compared to using $\ell_p$ norms with $p > 1$. This is expected to provide a more accurate approximation for the sparse signal under consideration. Secondly, optimization problem (2) has distinct properties depending on the $\ell_p$ norm used in the minimization, and consequently requires distinct optimization methods. Considering $\ell_0$-norm minimization, the resulting optimization problem is non-convex, implying that one has to rely on greedy methods such as orthogonal matching pursuit (OMP) [16]. On the other hand, $\ell_1$-norm minimization also induces sparsity in the solution (albeit to a lesser extent than $\ell_0$-norm minimization) and has the advantage of leading to a convex optimization problem, which can be solved by convex optimization methods or dedicated algorithms such as Basis Pursuit (BP) [17] or Dantzig selector [18]. It is also possible to consider $\ell_p$ norms with $0 < p < 1$, again leading to non-convex optimization problems.

### B. Perfect Recovery Guarantees

We now review the sufficient conditions under which it is possible to perfectly recover a $k$-sparse signal through CS as introduced in (1)–(3). The sensing matrix $\mathbf{A}$ will be of crucial importance, as the sufficient conditions for perfect recovery will be based on its properties. We remark that the sufficient conditions in compressed sensing are generally not satisfied in practical applications, and that their main interest is theory rather than applicability.

*1) Spark Property:* A first important requirement is that the sensing matrix $\mathbf{A}$ maps all distinct $k$-sparse vectors $\mathbf{x_1} \neq \mathbf{x_2}$ onto distinct vectors $\mathbf{y_1} \neq \mathbf{y_2}$. For an exact $k$-sparse $\mathbf{x}$, this holds if and only if

$$\text{spark}(\mathbf{A}) > 2k \tag{4}$$

where $\text{spark}(\mathbf{A})$ is defined as the smallest number of columns of $\mathbf{A}$ that are linearly dependent.

*2) Restricted Isometry Property (RIP):* This property provides a more subtle recovery guarantee in the case of noisy measurements. Essentially, a matrix $\mathbf{A}$ satisfies the RIP of order $k$ if there exists a $\delta_k \in (0, 1)$ such that [4]

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{5}$$

holds for all $k$-sparse $\mathbf{x}$. If $\mathbf{A}$ satisfies the RIP of order $2k$ with $0 < \delta_{2k} \leq 0.5$, this is a sufficient condition for a variety of CS algorithms to be able to perfectly recover any $k$-sparse signal.

*3) Coherence Property:* In practice, the spark and RIP properties are difficult to compute for a given sensing matrix $\mathbf{A}$. The coherence property provides a more practically feasible way to establish recovery guarantees. The *coherence* $\mu$ of a matrix $\mathbf{A}$ is defined as [20],

$$\mu(\mathbf{A}) = \max_{i,j,i\neq j} \frac{|\mathbf{a}_i^H \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2} \tag{6}$$

where $\mathbf{a}_i$ denotes the $i$-th column of $\mathbf{A}$. The coherence of a matrix is a measure of the decorrelation it provides as a transformation between the original and the analysis domain. The coherence property is related to the spark property through [21],

$$\text{spark}(\mathbf{A}) \geq 1 + \frac{1}{\mu(\mathbf{A})}. \tag{7}$$

Combining (7) and (4), and assuming no noise on the measurements, this leads to the sufficient condition

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{A})}\right) \tag{8}$$

for perfect recovery of a $k$-sparse signal using $\ell_0$ and $\ell_1$- norm minimization to be possible.

### C. CS-Based Declipping

The CS framework is a suitable framework for addressing the declipping problem. Because of its ability to reconstruct a sparse signal from a reduced measurement, CS can theoretically recover the original signal, including the parts that were clipped and deemed "lost." CS-based declipping is based on the following principles:

*1) Measurement Matrix $\mathbf{\Phi}$:* The measurement $\mathbf{y} = \mathbf{\Phi s}$ simply consists of the $M$ unclipped samples in the signal $\mathbf{s}$. The measurement matrix $\mathbf{\Phi}$ is entirely defined by the clipping pattern, i.e., it is a submatrix of an identity matrix, where the rows corresponding to the clipped samples have been removed.

*2) Fixed Basis $\mathbf{\Psi}$:* Appropriate sparsifying time-domain signal decompositions for audio signals include the Discrete Fourier Transform (DFT) and the Discrete Cosine Transform (DCT), which are defined as

$$X_l^{\text{DFT}} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n e^{-j2\pi l/Nn}, \quad l = 0 \ldots N-1 \tag{9}$$

$$X_l^{\text{DCT}} = \alpha(l,N) \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right) l\right] \tag{10}$$

$$\text{with} \quad \alpha(l,N) = \frac{\sqrt{1+\min(l,1)}}{\sqrt{N}},\ l = 0 \ldots N-1 \tag{11}$$

where $x_n$ denotes the $n$-th element of the time-domain vector $\mathbf{x} \in \mathbb{R}^N$, and $X_l$ denotes the $l$-th element of the transform domain vector $\mathbf{X} \in \mathbb{R}^N$. The DFT and DCT bases are suitable candidates to serve as the fixed basis in the declipping problem (i.e., $\mathbf{\Psi}$ can be chosen as the Inverse DFT matrix or the inverse DCT matrix). However, the DCT has an advantage compared to the DFT in that it involves only real-valued calculations.

*3) Perfect Recovery Guarantees:* The RIP is a very restrictive condition, which can only hold for specific classes of sensing matrices. One such class is the class of *random partial orthogonal matrices*, i.e matrices obtained by randomly choosing $M$ rows from a normalized orthogonal matrix [22]. For this class of matrices, the following result applies.

*Theorem 1 [23]:* Given an orthogonal matrix $\mathbf{\Psi}$ with entries of magnitude $O(1/\sqrt{N})$. A sensing matrix $\mathbf{A} = \mathbf{\Phi\Psi}$ consisting of a random subset of $M$ rows of $\mathbf{\Psi}$ satisfies the RIP condition $\delta_{3k} + 3\delta_{4k} \leq 2$ with high probability if and only if

$$M \geq O(k \log N). \tag{12}$$

For the declipping problem under consideration, this is presumably the most relevant perfect recovery guarantee that can be given. However, this perfect recovery guarantee only holds under the assumption of random positions of the clipped samples, which is not expected to be met for most clipped audio signals encountered in real-world audio devices. Moreover, even if this assumption is met, it is very likely that the RIP condition does not hold for dictionaries commonly used and values of sparsity $k$ commonly encountered in audio signal declipping.

*4) Frame-Based Processing:* Because of the short-time stationarity of audio signals, a declipping algorithm should operate on short-time signal frames. In the declipping algorithms proposed in this paper, a clipped audio signal is first split into Hann-windowed half-overlapping frames of $N$ samples each. The frames are sequentially declipped, and synthesized to form the declipped audio signal. Introducing the subscript $m$ as the frame index, the recovery step (2) for the $m$-th frame can be rewritten as

$$\mathbf{x}'_m = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_p \text{s.t. } \mathbf{A}_m\mathbf{z} = \mathbf{\Phi}_m\mathbf{\Psi z} = \mathbf{y}_m. \tag{13}$$

*5) Alternative Recovery Step:* The structure of a clipped audio signal has an inherent advantage, pertaining to the CS recovery step. As mentioned before, when applying a clipping level $\theta_c$ to a signal, all sample values beyond $\theta_c$ are mapped onto $\theta_c$. This means that the original value of any clipped sample is in absolute value *larger than or equal to* $\theta_c$. This forms an additional constraint on the eligible solution space for the sparse recovery [7], which leads to the following recovery step,

$$\mathbf{x}'_m = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_p \text{s.t. } \begin{cases} \mathbf{A}_m\mathbf{z} = \mathbf{\Phi}_m\mathbf{\Psi z} = \mathbf{y}_m \\ \mathbf{C}_m^+\mathbf{\Psi z} \geq \theta_c \\ \mathbf{C}_m^-\mathbf{\Psi z} \leq -\theta_c \end{cases} \tag{14}$$

where $\mathbf{C}_m^+$ and $\mathbf{C}_m^-$ are submatrices of an identity matrix where the rows corresponding to positively and negatively clipped samples, are respectively selected. Adding these constraints to the optimization problem is expected to improve the signal recovery quality.

*6) Relaxation of Equality Constraints:* When the signal $\mathbf{s}$ is not exactly sparse but only compressible, as it is the case for most real-world audio signals, a certain relaxation to the

equality constraints $\mathbf{A}_m\mathbf{z} = \mathbf{y}_m$ is desirable [24]. For $\ell_0$-norm minimization, a possible relaxation is to solve the following optimization problem,

$$\mathbf{x}'_m = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_0 \text{ s.t. } \begin{cases} \|\mathbf{y}_m - \mathbf{A}_m\mathbf{z}\|_2^2 \leq \epsilon_m \\ \mathbf{C}_m^+\mathbf{\Psi}\mathbf{z} \geq \theta_c \\ \mathbf{C}_m^-\mathbf{\Psi}\mathbf{z} \leq -\theta_c \end{cases} \quad (15)$$

where the value of the parameter $\epsilon_m > 0$ should be carefully selected [7]. As opposed to the $\ell_0$-norm minimization based declipping algorithms presented in [7], we will focus on the use of $\ell_1$-norm minimization for the declipping of audio signals. When using $\ell_1$-norm minimization, the relaxation proposed here is to solve the following optimization problem,

$$\mathbf{x}'_m = \arg\min_{\mathbf{z}} \|\mathbf{z}\|_1 + \frac{1}{2\gamma_m}\|\mathbf{y}_m - \mathbf{A}_m\mathbf{z}\|_2^2 \text{ s.t. } \begin{cases} \mathbf{C}_m^+\mathbf{\Psi}\mathbf{z} \geq \theta_c \\ \mathbf{C}_m^-\mathbf{\Psi}\mathbf{z} \leq -\theta_c \end{cases} \quad (16)$$

where $\gamma_m > 0$ is a regularization parameter. It is interesting to note that this optimization problem is similar to the well-known Basis Pursuit Denoising problem [17], except for the additional inequality constraints.

## III. A PCS Framework for Declipping

In Section II, the CS framework has been shown to be a suitable framework for addressing the declipping problem. However, of crucial importance for audio applications is the resulting *perceived audio quality* of the declipped signal, which does not necessarily coincide with the physical signal reconstruction quality. By additionally incorporating knowledge of human auditory perception, a novel perceptual compressed sensing (PCS) framework is presented in Sections III-A and III-B. A PCS-based declipping algorithm, using $\ell_1$-norm type reconstruction, is presented in Subsection III-C.

### A. Perceptual CS Framework

It is known that audio signal components at certain frequencies are more perceptible than components at other frequencies, and that the relative perceptibility is partly signal-dependent. Two phenomena of human auditory perception are responsible for this:

- The *absolute threshold of hearing* 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 [25].
- *Simultaneous masking* is a phenomenon where the presence of certain 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* $\mathbf{t}$ of an audio signal gives the amount of signal energy (dB) at each frequency that can be masked by the rest of the signal. As such, the masking threshold $\mathbf{t}$ gives an indication of the relative perceptibility of signal components at different frequencies, and can be used in the CS framework in order to focus on the recovery of perceptually important signal components, while at the same time avoiding audible recovery errors.

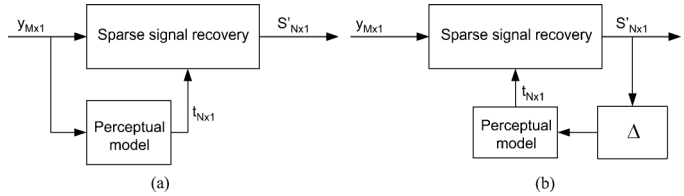



Fig. 2. Schematic overview of perceptual compressed sensing: (a) Feedforward mode (b) Feedback mode. The symbol $\Delta$ denotes a one-frame delay.

Fig. 2(a) gives a schematic overview of perceptual compressed sensing in *feedforward mode*. First, from the measurement $\mathbf{y}$, the masking threshold $\mathbf{t}$ is estimated through the use of a perceptual model. Second, the sparse signal recovery step uses the measurement $\mathbf{y}$ in conjunction with the masking threshold $\mathbf{t}$ in order to recover the signal $\mathbf{s}'$.

In practice, calculating the masking threshold $\mathbf{t}$ based on the measurement $\mathbf{y}$, may not yield an accurate estimate of the masking threshold of the original signal $\mathbf{s}$. Therefore, an alternative *feedback mode* of perceptual compressed sensing is proposed and illustrated in Fig. 2(b). In feedback mode, the masking threshold $\mathbf{t}_m$ for a current frame $\mathbf{y}_m$ is computed from the declipped signal $\mathbf{s}'_{m-1}$ of the previous frame. In the declipping algorithm proposed in Subsection III-C, perceptual compressed sensing will be used in feedback mode. The implications of this feedback masking threshold estimation procedure on the declipping performance will be evaluated in Subsection IV-C.

### B. Masking Threshold Calculation

The instantaneous global masking threshold $\mathbf{t}$ 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 [26] and [27]). We outline the relevant steps in the computation of the instantaneous global masking threshold and illustrate the result of each step on an example audio signal (see Fig. 3):

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. 3(a) shows the time-domain input signal, Fig. 3(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 [28]. 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

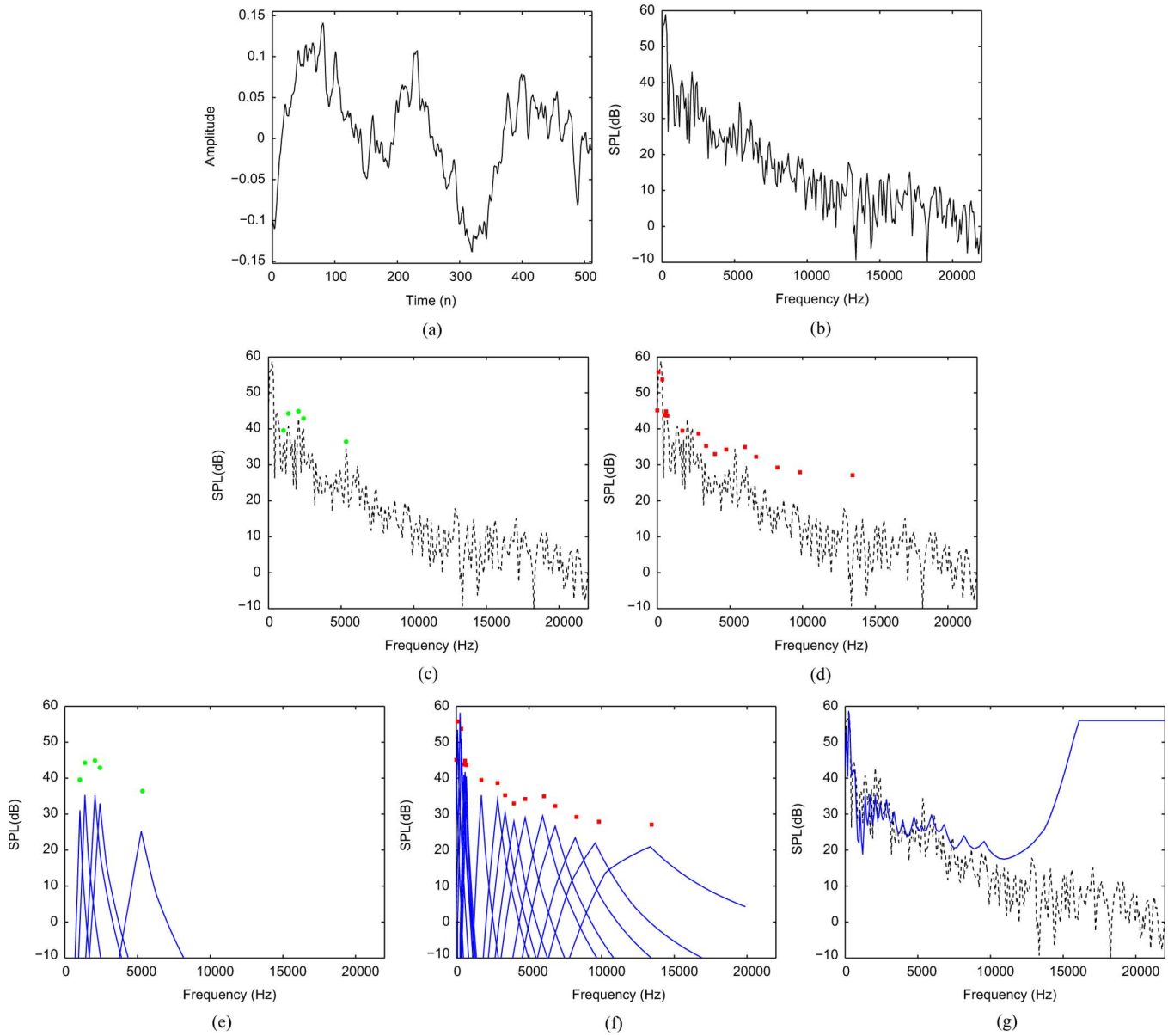



Fig. 3. Different steps in the computation of the global masking threshold using the ISO/IEC 11172–3 MPEG-1 Layer 1 psychoacoustic model 1: (a)-(b) Time domain and normalized frequency domain representations of the input audio signal (c)-(d) Tonal maskers (circles), non-tonal maskers (squares) and input 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) [11].

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. 3(c) and 3(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 level and tonality of the masker. Fig. 3(e) and 3(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. 3(g).

### C. PCS-Based Declipping Using $\ell_1$-norm Optimization

In order to focus CS-based declipping on the recovery of perceptually relevant signal components while at the same time avoiding audible recovery errors, the masking threshold will be incorporated into the CS recovery step. The proposed approach

is to introduce in the optimization problem a diagonal *perceptual weighting matrix* where the diagonal elements are the reciprocal of the masking threshold, hence indicating the relative perceptual importance of the different signal components. The perceptual weighting matrix $\mathbf{P}_m \in \mathbb{R}^{N\times N}$ is defined as

$$\mathbf{P}_m = \begin{bmatrix} t_{m,0}^{-1} & 0 & 0 & \dots & 0 \\ 0 & t_{m,1}^{-1} & 0 & \dots & 0 \\ 0 & 0 & t_{m,2}^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & t_{m,N-1}^{-1} \end{bmatrix} \tag{17}$$

where $t_{m,l}$ denotes the $l$-th component of the $m$-th frame's global masking threshold $\mathbf{t}_m$.

Different ways of incorporating the perceptual weighting matrix $\mathbf{P}_m$ into the declipping optimization problem can be envisaged. The proposed way of incorporating $\mathbf{P}_m$ in the declipping of the $m$-th frame using $\ell_1$-norm minimization, is to solve

$$\mathbf{x}'_m = \arg\min_{\mathbf{z}} \|\mathbf{P}_m\mathbf{z}\|_1 + \frac{1}{2\gamma_m}\|\mathbf{y}_m - \mathbf{A}_m\mathbf{z}\|_2^2 \quad \text{s.t.} \begin{cases} \mathbf{C}_m^+\boldsymbol{\Psi}\mathbf{z} \geq \theta_c \\ \mathbf{C}_m^-\boldsymbol{\Psi}\mathbf{z} \leq -\theta_c. \end{cases} \tag{18}$$

The declipped signal $\mathbf{s}'_m$ is constructed using

$$\begin{cases} \mathbf{C}_m^+\mathbf{s}'_m = \mathbf{C}_m^+\boldsymbol{\Psi}\mathbf{x}'_m \\ \mathbf{C}_m^-\mathbf{s}'_m = \mathbf{C}_m^-\boldsymbol{\Psi}\mathbf{x}'_m \\ \boldsymbol{\Phi}_m\mathbf{s}'_m = \mathbf{y}_m. \end{cases} \tag{19}$$

In formulation (18), the objective function term $\|\mathbf{P}_m\mathbf{z}\|_1$ is introduced in order to obtain a perceptually meaningful reconstruction. The perceptual weighting matrix $\mathbf{P}_m$ favors the use of those frequency components $l$ that have a high masking threshold $t_{m,l}$. This approach is perceptually desirable:

- The introduction of distinctively audible new signal components (low masking threshold) that are not present in the original signal, is discouraged. The introduction of less audible or inaudible additional signal components (high masking threshold) is tolerated to a greater extent. As mentioned before, the masking threshold for a given frequency component indeed quantifies the signal energy for that frequency that can be masked by the original signal, or equivalently, the required signal energy for that frequency component to become audible in the simultaneous presence of the original signal.
- The recovery of perceptually important signal components present in the original signal, is encouraged. These salient signal components will, by their relatively large signal energy, indeed possess high corresponding masking thresholds.

The perceptual weighting that is applied to the components of the sparse decomposition can be alternatively interpreted in the framework of Bayesian Compressive Sensing [29]. Omitting the constraints, the optimization problem in (18) is seen to be equivalent to a maximum a posteriori (MAP) formulation using independent Laplace priors [30] for each basis coefficient in $\mathbf{z}$, with mean prior values scaled by the corresponding diagonal elements of the perceptual weighting matrix $\mathbf{P}_m$.

Algorithm 1 describes the different steps of the proposed PCSL1 algorithm for declipping the $m$-th frame of a clipped audio signal, using $\ell_1$-norm type reconstruction[3].

**Algorithm 1** (PCSL1)

**Input** $\mathbf{y}_m \in \mathbb{R}^M$, $\mathbf{s}'_{m-1} \in \mathbb{R}^N$, $\boldsymbol{\Psi} \in \mathbb{C}^{N\times N}$, $\theta_c \in \mathbb{R}_+$, $\gamma_m \in \mathbb{R}_+$

**Output:** $\mathbf{s}'_\mathbf{m} \in \mathbb{R}^N$

1: $\mathbb{S}_m^r = \{l \| y_{m,l}| < \theta_c\}$
2: $\mathbb{S}_m^+ = \{l | y_{m,l} = \theta_c\}$
3: $\mathbb{S}_m^- = \{l | y_{m,l} = -\theta_c\}$
4: $\boldsymbol{\Phi}_m = I_N(\mathbb{S}_m^r)$
5: $\mathbf{C}_m^+ = I_N(\mathbb{S}_m^+)$
6: $\mathbf{C}_m^- = I_N(\mathbb{S}_m^-)$
7: $\mathbf{A}_m = \boldsymbol{\Phi}_m\boldsymbol{\Psi}$
8: Calculate $\mathbf{t}_m$ based on $\mathbf{s}'_{m-1}$ [using MPEG-1 Layer 1 psychoacoustic model 1]
9: Determine $\mathbf{P}_m$ using **(17)**
10: Recover $\mathbf{s}'_m$ by solving **(18)** and evaluating **(19)**

## IV. Evaluation

In order to comparatively evaluate the designed declipping algorithm with respect to existing declipping algorithms, objective tests as well as subjective listening tests have been conducted. The set-up, results and interpretation of the conducted objective and subjective experiments will be discussed in this section.

### A. Objective Evaluation

To evaluate the declipping algorithms, two objective measures are used. A first measure $\Delta\text{SNR}$ indicates the physical declipping improvement and is defined as the SNR improvement,

$$\Delta\text{SNR} = \text{SNR}(\mathbf{s}, \mathbf{s}') - \text{SNR}(\mathbf{s}, \mathbf{s_c}) \tag{20}$$

where

$$\text{SNR}(\mathbf{u}, \mathbf{v}) = 10\log_{10}\left(\frac{\|\mathbf{u}\|_2^2}{\|\mathbf{u}-\mathbf{v}\|_2^2}\right) \tag{21}$$

and $\mathbf{s}$ is the original signal, $\mathbf{s_c}$ is the clipped signal and $\mathbf{s}'$ is the declipped signal. A second measure $\Delta\text{ODG}$ indicates the perceptual declipping improvement and is defined as the Objective Difference Grade (ODG) improvement,

$$\Delta\text{ODG} = \text{ODG}(\mathbf{s}, \mathbf{s}') - \text{ODG}(\mathbf{s}, \mathbf{s_c}) \tag{22}$$

where $\text{ODG}(\mathbf{u}, \mathbf{v})$ is an objective measure of audio quality, which is calculated using the Basic Version of the PEAQ (*Perceptual Evaluation of Audio Quality*) recommendation [31]. The $\text{ODG}(\mathbf{u}, \mathbf{v})$ predicts the basic audio quality of a signal under test $\mathbf{v}$ with respect to a reference signal $\mathbf{u}$, and has a range between 0 and $-4$, corresponding to the ITU-R five grade impairment scale depicted in Fig. 4. The suitability of the PEAQ ODG

[3]In Algorithm 1, the notation $I_N(\mathbb{S})$ is introduced to denote a matrix consisting of those rows of an $N \times N$ identity matrix $I_N$ corresponding to the row indices in the index set $\mathbb{S}$.

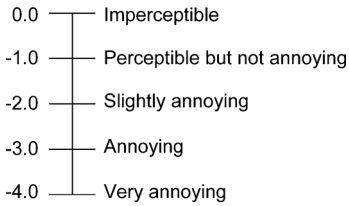



Fig. 4. The ITU-R five-grade impairment scale.

TABLE I
AUDIO EXCERPTS USED FOR THE COMPARATIVE EVALUATION OF DECLIPPING ALGORITHMS.

| Audio excerpt | Dynamics | $k'$ | SNR$(\mathbf{s}, \mathbf{s_c})$ [dB] ODG=-1 | ODG=-2 | ODG= -3 | Origin |
|---|---|---|---|---|---|---|
| BachPartita | high | 63 | 32.0 | 24.4 | 22.5 | [32] |
| FolkMusic | low | 178 | 18.7 | 15.2 | 12.2 | [33] |
| Mariana | low | 111 | 45.9 | 42.0 | 37.9 | [34] |
| Vivaldi | low | 235 | 22.0 | 17.2 | 12.8 | [35] |
| Whispers | high | 291 | 16.7 | 12.6 | 9.6 | [36] |

as an objective measure of perceived clipping degradation will be subjectively evaluated in Subsection IV-E.

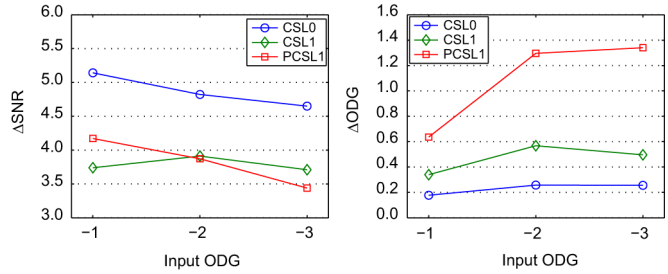
A test database consisting of five audio excerpts was composed (16 bit mono, 44.1 kHz). The audio signals were selected so as to cover different music styles, signal dynamics and signal sparsities. The length of each audio excerpt was 10 seconds. All audio signals were normalized to the same average loudness level. Table I gives details for the different audio excerpts, including their approximate sparsity[4] $k'$ for the DFT basis.

Each audio signal in the test database was first clipped at three distinct clipping levels $\theta_c$, corresponding to fixed input ODGs $\{-1, -2, -3\}$. Each clipped audio signal was subsequently processed by three different declipping algorithms, all using the IDFT matrix as fixed basis $\mathbf{\Psi}$ and operating on frames of $N = 512$ samples:

- CSL0 algorithm: CS-based declipping using $\ell_0$-norm optimization, all optimization problems (15) solved by Orthogonal Matching Pursuit [16], $\epsilon_m = 10^{-5}$. Although there are some differences (regarding the choice of the fixed basis and the choice of the stopping criterion for the OMP method), this declipping algorithm is in essence similar to the declipping algorithm proposed in [7], and can therefore adequately represent a baseline declipping algorithm in our comparative evaluation.
- CSL1 algorithm: CS-based declipping using $\ell_1$-norm optimization, all optimization problems (16) solved by the Basis Pursuit Denoising technique proposed in [37], $\gamma_m = 10^{-2}$.
- PCSL1 algorithm: PCS-based declipping using $\ell_1$-norm optimization, all optimization problems (18) solved by cvx[38], $\gamma_m = 10^{-2}$.

For each of the resulting total of $5 \times 3 \times 3 = 45$ declipped audio signals, the objective measures $\Delta$SNR and $\Delta$ODG were calculated. In Fig. 5, the mean $\Delta$SNR and $\Delta$ODG scores over all five audio signals are plotted as a function of the input ODG, for all considered declipping algorithms. The detailed evaluation results per audio signal are shown in Table II, from which

[4]The approximate sparsity $k'$ is defined as the average per-frame number of DFT signal components exceeding 0.001 times the maximal per-frame DFT signal component in absolute value, for frame length $N = 512$.




(a) (b)

Fig. 5. Comparative evaluation of different declipping algorithms in terms of objective audio quality: (a) mean $\Delta$SNR (b) mean $\Delta$ODG scores.

TABLE II
OBJECTIVE EVALUATION RESULTS FOR DIFFERENT DECLIPPING ALGORITHMS.

| Audio excerpt | Declipping algorithm | ΔSNR/ΔODG ODG=-1 | | ODG=-2 | | ODG=-3 | |
|---|---|---|---|---|---|---|---|
| BachPartita | CSL0 | 6.79 | 0.34 | 5.24 | 1.09 | 5.60 | 0.43 |
| | CSL1 | 6.35 | -0.09 | 7.73 | -0.03 | 7.92 | 0.16 |
| | PCSL1 | 7.21 | 0.24 | 6.81 | 1.03 | 5.85 | 1.83 |
| Vivaldi | CSL0 | 6.55 | 0.06 | 6.79 | 0.06 | 6.42 | 0.28 |
| | CSL1 | 3.27 | 0.38 | 2.91 | 0.53 | 2.47 | 0.39 |
| | PCSL1 | 2.98 | 0.81 | 2.63 | 1.33 | 2.26 | 1.32 |
| Mariana | CSL0 | 4.88 | 0.42 | 4.97 | 0.61 | 4.36 | 0.36 |
| | CSL1 | 3.65 | 0.63 | 3.70 | 1.14 | 3.31 | 1.03 |
| | PCSL1 | 5.58 | 0.93 | 5.07 | 1.69 | 4.48 | 1.64 |
| FolkMusic | CSL0 | 4.94 | 0.61 | 4.71 | 0.33 | 4.65 | 0.38 |
| | CSL1 | 3.32 | 0.46 | 3.10 | 0.62 | 2.86 | 0.50 |
| | PCSL1 | 3.14 | 0.41 | 2.95 | 0.63 | 2.76 | 0.68 |
| Whispers | CSL0 | 2.55 | -0.54 | 2.40 | -0.80 | 2.22 | -0.17 |
| | CSL1 | 2.11 | 0.32 | 2.12 | 0.58 | 1.99 | 0.40 |
| | PCSL1 | 1.95 | 0.47 | 1.90 | 1.23 | 1.85 | 0.95 |

the influence of the signal sparsity on the declipping performance can be evaluated.

From Fig. 5(a), we observe a positive average SNR improvement ranging between 3 and 6 dB for all considered declipping algorithms. This improvement appears to remain relatively constant over the input ODG range. The proposed PCSL1 algorithm does not outperform the CS-based algorithms in terms of SNR improvement. From Fig. 5(b), the average ODG improvement is significant for all declipping algorithms. The proposed PCSL1 algorithm significantly outperforms the CS-based algorithms in terms of ODG improvement, and this for all considered input ODGs. Also note that from Table II, it is observed that the best declipping performance is obtained for the sparser audio signals, such as BachPartita and Mariana.

Moreover, it is interesting to note from Fig. 5(a) the superior performance in terms of SNR improvement of the CSL0 algorithm compared to the CSL1 algorithm. This seems to confirm an earlier suggestion [7] that, in terms of SNR improvement, the use of $\ell_0$-norm optimization for declipping may be preferable over the use of $\ell_1$-norm optimization. However, we observe from Fig. 5(b) a superior performance in terms of ODG improvement of the CSL1 algorithm compared to the CSL0 algorithm, which would plead in favor of using $\ell_1$-norm optimization for declipping, as far as audio quality is concerned.

### *B. Impact of Regularization Parameter $\gamma_m$*

In [7], it was shown that the parameter $\epsilon_m$ in the CSL0 optimization formulation (15) should be carefully selected, as

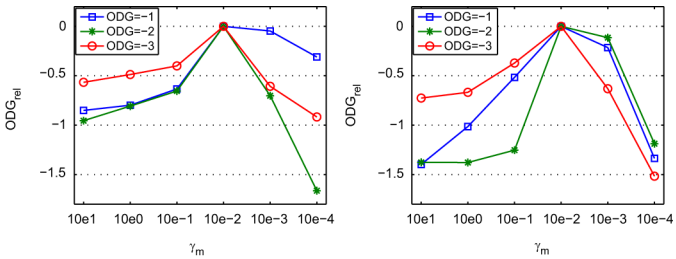



Fig. 6. Impact of the regularization parameter $\gamma_m$ in the PCSL1 declipping algorithm on the resulting relative (to the maximum) ODG score, for different input ODGs: (a) "Vivaldi" (b) "Whispers."

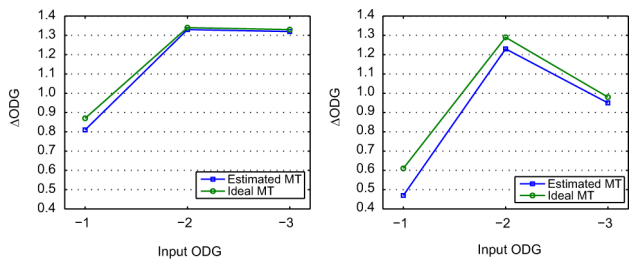



Fig. 7. Impact of the masking threshold estimation procedure in the PCSL1 declipping algorithm on the $\Delta$ODG score, for different input ODGs: (a) "Vivaldi" (b) "Whispers."

the resulting SNR improvement was seen to be very sensitive to the value of $\epsilon_m$. Based upon these findings and some additional experiments, we have selected $\epsilon_m = 10^{-5}$ as an appropriate parameter setting for the experiments discussed in Subsection IV-A.

In order to study the impact of the regularization parameter $\gamma_m$ in the PCSL1 optimization formulation (18) on the resulting audio quality, the following experiment was conducted. Each of the five audio signals in the test database detailed in Table I were clipped at three distinct clipping levels, corresponding to fixed input ODGs $\{-1, -2, -3\}$. These signals were subsequently processed by the PCSL1 declipping algorithm using optimization formulation (18), in which the value of the regularization parameter $\gamma_m$ was fixed over all frames within one audio signal, and this processing was repeated for six different values of $\gamma_m = \{10^1, 10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. For each of the resulting total of $5 \times 3 \times 6 = 90$ declipped audio signals, the ODG between the original signal and the declipped signal was calculated.

The results of this experiment are partly visualized in Fig. 6, which shows the results for the audio excerpts "Vivaldi" and "Whispers." In these figures, the objective measure $\mathrm{ODG}_{\mathrm{rel}}$, defined as

$$\mathrm{ODG}_{\mathrm{rel}}(\gamma_m) = \mathrm{ODG}(\gamma_m) - \max_{\gamma_m} \mathrm{ODG}(\gamma_m) \qquad (23)$$

is plotted as a function of $\gamma_m$, and this for different input ODGs. We observe that the choice of $\gamma_m$ has a significant impact on the resulting audio quality scores after declipping and this regardless of the input ODG, so care should be taken when selecting the value of $\gamma_m$. Moreover, we observe that the different curves reach a joint maximum for $\gamma_m = 10^{-2}$. We have selected $\gamma_m = 10^{-2}$ as an experimentally validated setting for our experiments discussed in Subsection IV-A.

### C. Impact of Masking Threshold Estimation Procedure

In Section III, the estimation of the masking threshold was discussed. It was proposed to estimate the masking threshold $\mathbf{t}_m$ for a current signal frame $\mathbf{y}_m$ by computing it from the previously declipped signal frame $\mathbf{s}'_{m-1}$. The impact of this feedback masking threshold estimation procedure on the PCSL1 declipping performance has been evaluated as follows.

The audio signals "Vivaldi" and "Whispers" (see test database detailed in Table I) were clipped at three distinct clipping levels, corresponding to fixed input ODGs $\{-1, -2, -3\}$. These signals were subsequently processed by the PCSL1 declipping algorithm, in which the per-frame masking thresholds were established using two different procedures:

- Ideal masking threshold: the per-frame masking threshold was calculated using the clean signal frame $\mathbf{s}_m$. Using this *ideal* masking threshold provides an upper bound for the declipping performance of the PCSL1 algorithm.
- Estimated masking treshold: the per-frame masking threshold $\mathbf{t}_m$ was computed using the previously declipped signal frame $\mathbf{s}'_{m-1}$, as detailed in Algorithm 1.

The results of this experiment are shown in Fig. 7. It can be observed that for both audio excerpts, the resulting $\Delta$ODG scores using the estimated masking threshold in the declipping algorithm, are very close to the upper bounds provided by using the ideal masking threshold. These results indicate that the use of the proposed masking threshold estimation procedure does not have a significant negative impact on the resulting objective audio quality scores.

### D. Subjective Evaluation

*1) Research Question and Hypothesis:* The research question to be answered through performing a formal subjective listening test [39] is the following: "how does the perceived audio quality improvement of audio signals declipped by the proposed PCSL1 algorithm compare to that of audio signals declipped by the CSL0 and CSL1 algorithms?." The research hypothesis, that may or may not be rejected, is that the perceived audio quality improvement is identical for all three declipping algorithms.

*2) Test Subjects:* A representative sample of 16 test subjects having considerable musical listening and performance experience was selected to perform the listening test. All subjects were remunerated for their participation.

*3) Experimental Design and Set-Up:* 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[5] connected to a soundcard-equipped laptop[6]. Self-developed software was used to automate stimulus presentation and response collection. The playback level was fixed at a comfortable level.

[5]Sennheiser HD 439: dynamic, closed transducer, frequency response 17–22500Hz, Sound Pressure Level 112 dB, Total Harmonic Distortion $< 0.1\%$.

[6]Sony Vaio VGN-CR41: Intel Core 2 duo T5550 processor @1.83Ghz, 3GB RAM, Realtek sound card, Intel GMA X3100 Graphics Processor.

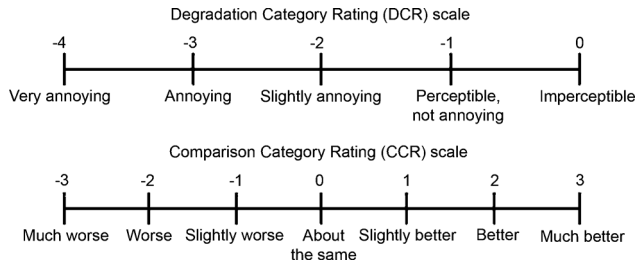



Fig. 8. ITU-T Degradation Category Rating (DCR) and Comparison Category Rating (CCR) scales (adapted from [39]).

The stimuli presented to the test subjects were the same as described in Subsection IV-A, i.e., they consisted of the five audio excerpts detailed in Table I, clipped at three distinct clipping levels, and subsequently declipped by the three declipping algorithms under study. This resulted in a total of $5 \times 3 \times 3 = 45$ pairs of stimuli (each consisting of a clipped signal and the corresponding declipped signal) that were presented to the test subjects. For each pair of stimuli, the test subjects were asked to provide the following responses:

- Rate the perceived audio quality degradation of the presented clipped signal using the ITU-T Degradation Category Rating (DCR) scale [40] (see Fig. 8).
- Rate the perceived audio quality difference of the presented declipped signal relative to the clipped signal, using the ITU-T Comparison Category Rating (CCR) scale [40] (see Fig. 8).

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 effect of clipping on audio signals, by successively listening to an original sample audio signal and its clipped version. The presentation order of the pairs of stimuli was randomized using an altered Latin square scheme [39], thus eliminating possible bias effects due to order effects and sequential dependencies.

*4) Results:* The raw data resulting from the listening test consists of a categorical DCR and CCR response by each of the 16 test subjects, for each of the 45 presented pairs of stimuli. The categorical DCR and CCR responses were first converted to integers according to the scales in Fig. 8. The analysis here will be focused on the CCR responses, the DCR responses will be used in the analysis in Subsection IV-E. Let us denominate the averaged CCR responses over all 16 test subjects as $\overline{\mathrm{CCR}}$ responses, and the averaged $\overline{\mathrm{CCR}}$ responses over all five audio excerpts as $\overline{\overline{\mathrm{CCR}}}$ responses. In Fig. 9, the $\overline{\overline{\mathrm{CCR}}}$ responses are plotted as a function of the input ODG level, and this for the three different declipping algorithms under study. We observe that the ranking of the algorithms is identical to the one observed in the objective evaluation results in Fig. 5(b). The detailed $\overline{\mathrm{CCR}}$ responses per audio excerpt are given in Table III.

The following statistical analysis was performed on the obtained numerical set of CCR responses. Let us denote the population CCR responses corresponding to audio signals declipped by the CSL0, CSL1 and PCSL1 algorithms by random variables $R_{\mathrm{CSL0}}$, $R_{\mathrm{CSL1}}$, and $R_{\mathrm{PCSL1}}$, respectively. Based on the

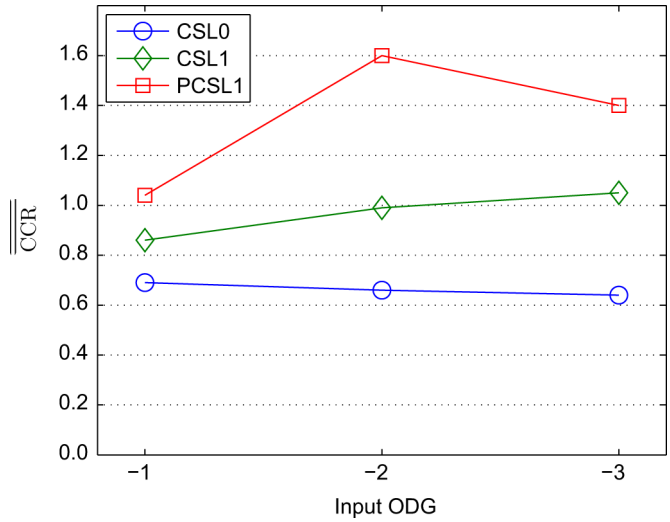



Fig. 9. Comparative evaluation of different declipping algorithms in terms of the $\overline{\overline{\mathrm{CCR}}}$ responses.

TABLE III
SUBJECTIVE EVALUATION RESULTS FOR DIFFERENT DECLIPPING ALGORITHMS.

| Audio excerpt | Declipping algorithm | $\overline{\mathrm{CCR}}$ ODG=-1 | $\overline{\mathrm{CCR}}$ ODG=-2 | $\overline{\mathrm{CCR}}$ ODG=-3 |
|---|---|---|---|---|
| BachPartita | CSL0 | 0.44 | 0.38 | 0.63 |
| | CSL1 | 0.38 | 0.38 | 0.75 |
| | PCSL1 | 1.63 | 2.56 | 2.31 |
| Vivaldi | CSL0 | 0.25 | 0.75 | 0.44 |
| | CSL1 | 0.38 | 0.69 | 0.94 |
| | PCSL1 | 0.38 | 0.88 | 1.56 |
| Mariana | CSL0 | 0.81 | 0.63 | 0.31 |
| | CSL1 | 1.38 | 1.81 | 1.19 |
| | PCSL1 | 1.69 | 2.19 | 2.06 |
| FolkMusic | CSL0 | 0.63 | 0.44 | 0.75 |
| | CSL1 | 0.50 | 1.13 | 1.25 |
| | PCSL1 | 0.44 | 1.19 | 1.06 |
| Whispers | CSL0 | 1.31 | 1.13 | 1.06 |
| | CSL1 | 1.69 | 0.94 | 1.13 |
| | PCSL1 | 1.06 | 1.19 | 0.88 |

sample CCR responses, we tested the following three statistical hypotheses $H_0$ against their alternatives $H_a$,

$$H_0^1 : \tilde{R}_{\mathrm{CSL0}} = \tilde{R}_{\mathrm{PCSL1}} \quad (24)$$
$$H_a^1 : \tilde{R}_{\mathrm{CSL0}} \leq \tilde{R}_{\mathrm{PCSL1}} \quad (25)$$
$$H_0^2 : \tilde{R}_{\mathrm{CSL1}} = \tilde{R}_{\mathrm{PCSL1}} \quad (26)$$
$$H_a^2 : \tilde{R}_{\mathrm{CSL1}} \leq \tilde{R}_{\mathrm{PCSL1}} \quad (27)$$
$$H_0^3 : \tilde{R}_{\mathrm{CSL0}} = \tilde{R}_{\mathrm{CSL1}} \quad (28)$$
$$H_a^3 : \tilde{R}_{\mathrm{CSL0}} \leq \tilde{R}_{\mathrm{CSL1}} \quad (29)$$

where $\tilde{R}$ is the population median of the random variable $R$. These three statistical hypotheses were tested for all three considered input ODGs. All statistical hypotheses were tested using one-tailed Wilcoxon-Mann-Whitney tests [41] with significance level $\alpha = 0.05$. The resulting one-sided P-values are synthesized in Table IV.

The first null hypothesis (24) can be rejected in favor of the alternative(25) for all considered input ODGs. The second and third null hypotheses(26) and(28) can be rejected in favor of their respective alternatives(27) and(29) for input ODGs of $-2$ and $-3$.

TABLE IV
P-VALUES FROM ONE-TAILED WILCOXON-MANN-WHITNEY TESTS ON SAMPLE CCR RESPONSES. SIGNIFICANT P-VALUES WITH RESPECT TO $\alpha = 0.05$ IN BOLD.

| Null hypothesis $\rightarrow$ | $H_0^1$ | $H_0^2$ | $H_0^3$ |
|---|---|---|---|
| ODG=-1 | **0.0202** | 0.2006 | 0.1313 |
| ODG=-2 | **<0.0001** | **0.0003** | **0.0178** |
| ODG=-3 | **<0.0001** | **0.0004** | **0.0035** |

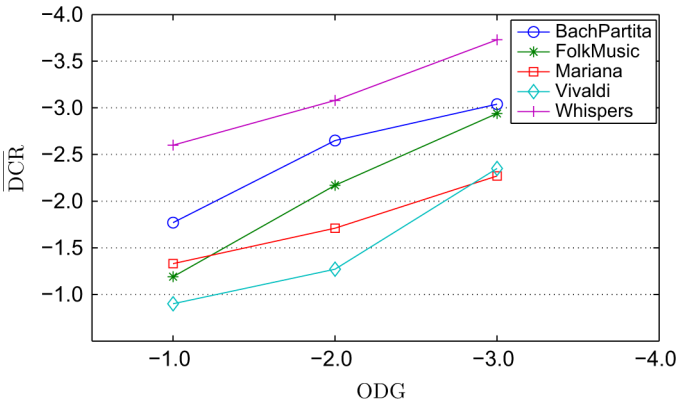



Fig. 10. Subjective $\overline{\mathrm{DCR}}$ responses as a function of the objective ODG score.

*5) Conclusions:* The research hypothesis can be confidently rejected, i.e., the perceived audio quality improvement is not identical for all three declipping algorithms. The statistical analysis has shown that the PCSL1 algorithm delivers a significantly better audio quality improvement than both the CSL0 algorithm (for all input ODGs) and the CSL1 algorithm (for input ODGs in $\{-2, -3\}$). Moreover, the CSL1 algorithm delivers a significantly better perceived audio quality improvement than the CSL0 algorithm (for input ODGs in $\{-2, -3\}$).

### *E. Suitability of PEAQ ODG as A Measure of Perceived Clipping Degradation*

The PEAQ ODG measure has been developed in the context of quality evaluation of low-bit-rate encoded audio. As the nature of signal distortions introduced by clipping can be rather different as compared to signal distortions introduced by low-bit-rate codecs, the use of PEAQ ODG as an objective measure of perceived clipping-induced degradations should be well-founded. Therefore, we have investigated the appropriateness of the PEAQ ODG measure for quantifying the perceived clipping-induced audio quality degradation by analyzing the DCR data collected in the listening test described in Subsection IV-D.

Let us denominate the averaged DCR responses over all 16 test subjects as $\overline{\mathrm{DCR}}$ responses. In Fig. 10, the $\overline{\mathrm{DCR}}$ responses are plotted as a function of the corresponding ODG score, and this for the five different audio excerpts. We observe a strong positive correlation between $\overline{\mathrm{DCR}}$ responses and ODG scores, and this for all audio excerpts. Moreover, the different curves are seen to be monotonously increasing and they do not deviate excessively from linear curves. The different curves do have a noticeably different vertical offset, but have a nearly equal slope. These results seem to indicate that the ODG measure can be confidently used to compare the perceived clipping-induced audio quality degradation for the *same* audio excerpt in different processing scenarios (as was done in Subsection IV-A). However, it might not be advisable to use the ODG measure to compare the perceived clipping-induced audio quality degradation for *different* audio excerpts.

## V. CONCLUSION

In this paper, a novel perceptual compressed sensing (PCS) framework has been presented for declipping audio signals, in which the theory of compressed sensing (CS) was combined with properties of human auditory perception. A declipping algorithm using $\ell_1$-norm type reconstruction has been developed in the PCS framework. Comparative evaluation experiments consisting of objective and subjective tests have revealed a significant audio quality increase of the proposed PCS-based declipping algorithm compared to CS-based declipping algorithms for reasonably sparse signals.

## REFERENCES

[1] F. Foti, "Aliasing distortion in digital dynamics processing, the cause, effect, and method for measuring it: The story of 'digital grunge!," presented at the AES 106th Conv., Munich, Germany, May 1999, Preprint no. 4971.
[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] J. S. Abel en and J. O. Smith, "Restoring a clipped signal," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 1991, vol. 3, pp. 1745–1748.
[4] S. J. Godsill, P. J. Wolfe, and W. N. Fong, "Statistical model-based approaches to audio restoration and analysis," *J. New Music Res*, vol. 30, no. 4, pp. 323–338, 2001.
[5] D. Navakauskas, "Artificial neural networks for the restoration of noise distorted songs and audio records," Ph.D. dissertation, Vilnius Gedimina Tech. Univ., Vilnius, Lithuania, 1999.
[6] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting: Problem statement, relation with sparse representations and some experiments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '11)*, 2011.
[7] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 922–932, Mar. 2012.
[8] A. J. Weinstein and M. B. Waking, "Recovering a clipped signal in sparseland," *Sampling Theory in Signal Image Process.*, Oct. 2013, to be published.
[9] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.
[10] M. Christensen and B. Sturm, "A perceptually reweighted mixed-norm method for sparse approximation of audio signals," in *Conf. Rec. 45th Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, Nov. 2011, pp. 575–579.
[11] 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.
[12] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[13] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[14] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[15] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, "Sparse representations in audio and music: From coding to source separation," *Proc. IEEE*, vol. 98, no. 6, pp. 995–1005, Jun. 2010.
[16] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orhogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Annu. Asilomar Conf. Signals, Syst., Comput.*, 1993.
[17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[18] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.
[19] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[20] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[22] L. Gan, C. Lingy, T. Doz, and T. Tranz, "Analysis of the statistical restricted isometry property for deterministic sensing matrices using Stein's method," 2009 [Online]. Available: http://dsp.rice.edu/?les/cs/GanStatRIP.pdf
[23] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.
[24] J. Yang and Y. Zhang, "Alternating direction algorithms for $\ell_1$ -problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.
[25] E. Terhardt, "Calculating virtual pitch," *Hearing Res.*, vol. 1, no. 2, pp. 155–182, 1979.
[26] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.
[27] "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.
[28] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, 2nd ed ed. New York, NY, USA: Springer, 1999.
[29] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
[30] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
[31] "Method for objective measurements of perceived audio quality," Int. Telecomm. Union Rec. bs. 1387, 1998.
[32] J. S. Bach, "Partita in a minor(1)," bWV 826.
[33] "Naragonia, Martin-Pecheur & Pink Molly,". Tandem.
[34] August Burns Red, "Marianas Trench," Constellations, Solid State Records.
[35] A. Vivaldi, "The Four Seasons—Spring," rV 269.
[36] All That Remains, "Whispers (I Hear Your)," The Fall of Ideals, Prosthetic Records.
[37] K. M. Cheman, "Optimization techniques for solving basis pursuit problems, dissertation," Ph.D. dissertation, North Carolina State Univ., Raleigh, NC, USA, 2006.
[38] M. Grant, S. Boyd, and Y. Ye, "Cvx: Matlab software for disciplined convex programming," [Online]. Available: http://stanford.edu/~boyd/cvx
[39] S. Bech and N. Zacharov, *Perceptual audio evaluation-Theory, method and application*. New York, NY, USA: Wiley, 2007.
[40] "Methods for subjective determination of transmission quality," Int. Telecomm. Union, Rec.P. 800.
[41] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.



**Bruno Defraene** (S'10) was born in Halle, Belgium, in 1986. He received the M.Sc. degree in electrical engineering from KU Leuven, Belgium, in 2009.

He is currently pursuing the PhD degree at the Electrical Engineering Department of KU Leuven, and is a member of KU Leuven's Optimization in Engineering Center (OPTEC).

His research interests include audio signal processing, acoustical signal enhancement, audio quality assessment, and optimization for signal processing applications.



**Naim Mansour** was born in Wilrijk, Belgium, in 1989. He received the dual M.Sc. degree in electrical engineering from KU Leuven, Belgium, and Danmarks Tekniske Universitet, Denmark, in 2012.

He wrote his M.Sc. thesis on the topic of perceptual declipping of audio signals through compressed sensing, focusing on the algorithm design and evaluation.

He is currently starting a professional career as an IP Quality Assurance Engineer at Alcatel-Lucent in Antwerp, Belgium.

**Steven De Hertogh** received the M.Sc. degree in electrical engineering from KU Leuven, Belgium, in 2012.

He wrote his M.Sc. thesis on the topic of perceptual declipping of audio signals through compressed sensing, focusing on the perceptual models and optimization methods.

He is now with EVS Broadcast Equipment, Seraing, Belgium.



**Toon van Waterschoot** (S'04–M'12) received the MSc degree (2001) and the PhD 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 (UniversitĹ della Svizzera italiana), Switzerland. He has been teaching courses related to digital signal processing and control theory. His research interests are in adaptive and distributed signal processing and parameter estimation, with application to acoustic signal enhancement, speech and audio processing, computational acoustics, and sensor networks.

Dr. van Waterschoot is serving as an Associate Editor for the Journal of the Audio Engineering Society and for the EURASIP Journal on Audio, Music, and Speech Processing, and as a Nominated Officer for the European Association for Signal Processing (EURASIP). He has been serving as an Area Chair for Speech Processing at the European Signal Processing Conference (EUSIPCO 2010 and 2013), and has been a Technical Program Committee member for several other international conferences. He is a member of the Audio Engineering Society, the Acoustical Society of America, EURASIP, and IEEE.



**Moritz Diehl** (M'11) studied Mathematics and Physics in Heidelberg and Cambridge University and did his PhD in 2001 at Heidelberg's Interdisciplinary Center for Scientific Computing (IWR).

Since 2006 he is a professor at the Electrical Engineering Department of KU Leuven University and the principal investigator of KU Leuven's Optimization in Engineering Center (OPTEC).

In 2013 he accepted a full professorship at the Technical Faculty of Freiburg University. His research focuses on optimization and control, spanning from numerical methods and algorithm development to applications.

One of his application areas is airborne wind energy, which he investigates within an ERC grant running from 2011 to 2016.



**Marc Moonen** (M'94–SM'06–F'07) received the electrical engineering degree and the PhD degree in applied sciences from KU Leuven, Belgium, in 1986 and 1990 respectively. Since 2004 he 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), and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and is currently President of EURASIP (European Association for Signal Processing).

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*.