A MULTIDIMENSIONAL COMPANDING SCHEME FOR SOURCE CODING WITH A PERCEPTUALLY RELEVANT DISTORTION MEASURE

J. Crespo, P. Aguiar

Department of Electrical and Computer Engineering
Instituto Superior Técnico
Lisboa, Portugal

R. Heusdens

Department of Mediamatics
Technische Universiteit Delft
Delft, The Netherlands

ABSTRACT

In this paper, we develop a multidimensional companding scheme for the preceptual distortion measure by S. van de Par [1]. The scheme is asymptotically optimal in the sense that it has a vanishing rate-loss with increasing vector dimension. The compressor operates in the frequency domain: in its simplest form, it pointwise multiplies the Discrete Fourier Transform (DFT) of the windowed input signal by the square-root of the inverse of the masking threshold, and then goes back into the time domain with the inverse DFT. The expander is based on numerical methods: we do one iteration in a fixed-point equation, and then we fine-tune the result using Broyden's method. Additionally, we will show simulations which corroborate the approximations and results of the theoretical derivations.

***Index Terms*—** Multidimensional companding, asymptotic optimality, perceptual distortion measure, sinusoidal audio coding

1. INTRODUCTION

In this last decade we have observed an explosive increase in the usage of audio coding schemes and coded audio content, which enable the reduction of the information throughput (bitrate) of an audio signal on the order of 7 to 15 times with respect to the original Pulse Code Modulation (PCM) coded signal, with very reduced penalty in perceptual quality [2]. These schemes make countless applications possible, such as handheld audio decoders with reduced memory capacity which nevertheless can carry hours of audio content, streaming thorugh bandwidth constrained channels such as the Internet with low bandwith usage and high experienced quality, delivery of audio content interactively through the World Wide Web (WWW), digital radio, digital television, recorded digital video, and many more.

This decrease in bitrate yet with surprisingly high transparency is achieved through the exploitation of the perceptual irrelevance and statistical redundancy present in the audio signal [3]. Indeed, contemporary audio coders use *lossy* source coding mechanisms based on quantization, where less perceptually important information is quantized more roughly and vice-versa, and *lossless* coding of the symbols emitted by the quantizer to reduce statistical redundancy. In the quantization process of the (en)coder, it is desirable to quantize the information extracted from the signal in a rate-distortion optimal sense, i.e., in a way that minimizes the perceptual distortion experienced by the user subject to the constraint of a certain available bitrate.

Due to its mathematical tractability, the Mean Square Error (MSE) is the elected choice for the distortion measure in many coding applications. In audio coding, however, the usage of more complex perceptual distortion measures different from the MSE, exploting the frequency selectivity and masking phenomena of the human auditory system, achieves a much higher performance, i.e., a higher perceptual quality for the same bitrate. A way to code with these measures is to multiply the input signal by certain perceptual weights before quantizing on the encoder and do the inverse (divide) on the decoder, so that the perceptual distortion measure gets mapped into an MSE in the normalized domain. These weights are usually related to the masking threshold, a time-frequency function delivering the maximum distortion power insertible in a time-frequency bin so that the distortion is still inaudible. The ease of use of the MSE distortion measure makes this solution attractive, having it been employed in several quantization schemes [4], [5], [6], [7], [8].

Nevertheless, such an approach has the inconvenient that the weights have to be transmitted as side information through the channel so that the receiver can do the inverse normalization, thereby introducing an overhead in the transmission process. In the case of a multiple description coding scenario [9], this overhead may be intolerably high. In such a scenario we have to transmit information through $n > 1$ channels and at the receiver $m \leq n$ channels are received successfully. The channels that were not transmitted successfully do not arrive at all (they are "erased"). As we don't know at the sender which channels will arive, we have to transmit the perceptual weights in all channels. With increasing number of channels n, for a constant total bit budget, the useful information for the audio signal becomes thus smaller and smaller, becoming a multiple description audio coding scheme thus useless for high n.

To solve this problem, we will make use of *multidimensional companding* [10]. As shown in figure 1, in a multidimensional companding quantization scheme the source $x \in \mathbb{R}^N$ is first pre-processed by applying a non-linear vector function F to it, that we call the *compressor*, then the result is vector quantized, transmitted through the channel in an efficient way, and finally at the receiver the inverse function F^{-1}, the *expander*, is applied as a post-processing step, being the reproduction y obtained in this fashion. The set of both functions, the compressor and the expander, is called the *compander*.

$$x \longrightarrow \boxed{F(\cdot)} \longrightarrow \boxed{Q(\cdot)} \xrightarrow{\text{channel}} \boxed{F^{-1}(\cdot)} \longrightarrow y$$

Fig. 1. Source coding with multidimensional companding

In [10], it was shown that general non-difference measures (which satisfy some natural conditions) are equivalent to the MSE measure at high resolution (at low distortion levels) in the compressed domain (between the compressed source signal and the not-yet-expanded received signal) if an "optimal" companding scheme (in the sense described in [10]) was applied, thus making all schemes optimal for the MSE measure also optimal for this type of non-MSE measures upon application of the optimal companding scheme. The advantage of using an optimal companding scheme in audio source coding is then that coding with a perceptually relevant (non-MSE) distortion measure can be done with any MSE-based scheme without the need to do pre-normalization of the input signal with perceptual weights. The transmission of this side information is thereby removed, and multiple description coding can be done with an arbitrary large number of descriptions n without performance loss.

Although the usage of multidimensional companding seems promising, an optimal companding scheme does in general not exist [10]. To work around that problem, a *suboptimal* companding scheme must be built, having an additional rate (entropy), the so called *rate loss*, at the output of the quantizer in relation to the rate that would be theoretically achievable with the (non-existent) optimal companding scheme. In this paper we will sum up the work done in [11], which had as objective the construction and simulation of such a sub-optimal companding scheme, with vanishing rate-loss with increasing vector dimension, for the perceptual distortion measure developed by S. van de Par [1], a distortion measure designed for sinusoidal audio coding [12] applications.

This paper is organizes as follows. In section

2. PRELIMINARIES

We will start overviewing the theme of multidimensional companding, and in particular the condition for optimality that we will need to derive the companding scheme and the additional rate at the output of the quantizer when using a sub-optimal scheme, with respect to the optimal scheme, the so-called *rate-loss*. For an extensive treatment of the subject we refer the reader to [10]. Subsequently, we will deliver the distortion measure which will be used in this paper, described in [1].

2.1. Multidimensional Companding

Linder et al. [10] treated the theme of multidimensional companding (figure 1) assuming that the distortion measure which we are coding with is *locally quadratic*. This type of distortion measures is characterized by being of class $C^3(\mathbb{R}^N)$ (d is 3 times continuously differentiable) with respect to y, by being strictly positive except in its absolute minimum $y = x$, where it should be 0,

$$d(x,y) \geq 0 \quad \text{with equality iff } y = x, \tag{1}$$

and by having a Taylor expansion with respect to y around x given by

$$d(x,y) = (y-x)^{\mathrm{T}} M(x)(y-x) + O(\|y-x\|^3), \tag{2}$$

where $\|\cdot\|$ denotes the l_2 norm and where the matrix $M(x)$, dubbed in [10] as the *sensitivity matrix*, is defined as

$$[M(x)]_{m,l} = \frac{1}{2}\frac{\partial^2 d(x,y)}{\partial y_m \partial y_l}\bigg|_{y=x}, \quad m,l = 0,1,\ldots,N-1, \tag{3}$$

i.e., it is half of the Hessian matrix of d with respect to y calculated at $y = x$. The absence of the 0^{th} and 1^{st} order term in (2) comes directly from the condition (1). Due to the same condition, $M(x)$ is positive definite.

It was defined in [10] that a multidimensional companding scheme is optimal when the Jacobian matrix

$$[F'(x)]_{m,l} = \frac{\partial F_m}{\partial x_l}(x), \quad m,l = 0,1,\ldots,N-1 \tag{4}$$

of the compressor satisfies[1]

$$M(X) = F'(X)^{\mathrm{T}} F'(X) \tag{5}$$

almost everywere, where $M(X)$, is the sensitivity matrix of equation (3), where $^{\mathrm{T}}$ denotes transposition and where we denote random variables by uppercase letters and their realizations by the lowercase correspondences. Due to the positive definiteness of (3), there exists a matrix F' satisfying (5) [13].

[1]In the original paper there was a multiplying constant c which is set to 1 for simplicity. No loss of generality occurs.

We call any such a matrix a square-root of M, denoted here by $\sqrt{M}$. An equivalent condition to (5) is then

$$F'(x) = \sqrt{M(x)}, \tag{6}$$

for some square-root of M. When fixing the matrix $\sqrt{M}$, the solution of (6) is not unique: any solution of the form $U(x)\sqrt{M(x)}$ with $U(x)$ orthogonal is also a solution of (5) and all solutions of (5) are separated by the left-multiplication by an orthogonal matrix. In this paper, we will restrict ourselves to the case $U(x) = \mathrm{I}$, where I is the identity matrix.

Additionally, the rate-loss of a sub-optimal companding scheme at high resolution (low distortion) was derived in [10] as being

$$\begin{aligned} H(Q_{D,\tilde{F}}) - H(Q_{D,F}) &\approx \\ \approx \frac{1}{2N} \mathrm{E} \log_2 &\left(\frac{\det \tilde{M}(X)}{\det M(X)} \right) \\ &+ \frac{1}{2} \log_2 \left[\mathrm{E}\ \mathrm{tr} \left(\frac{\tilde{M}(X)^{-1} M(X)}{N} \right) \right] \end{aligned} \tag{7}$$

with

$$\tilde{M}(X) = \tilde{F}'(X)^{\mathrm{T}} \tilde{F}'(X), \tag{8}$$

2.2. Perceptual Distortion Measure

In [1], S. van de Par defined an auditory, perceptually relevant distortion measure to be used in the context of sinusoidal audio coding, which delivers the distortion detectability of an N-dimensional signal x and its reproduction y as a weighted mean square error of the windowed signals in the frequency domain. More precisely, the distortion measure between x and y is defined as

$$d(x, y) = \sum_{f=0}^{L-1} \hat{a}^2(x, f)\, |\widehat{yw}(f) - \widehat{xw}(f)|^2, \tag{9}$$

where the juxtaposition of two vectors denotes the pointwise product between them, the $\hat{}$ operator denotes the L-point unnormalized discrete Fourier transform (DFT) in which the input signal is zero-padded up to size L, i.e. for any N-dimensional signal u

$$\hat{u}(f) = \sum_{n=0}^{N-1} u(n)\, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{L} f n}, \quad f = 0, 1, \ldots, L-1, \tag{10}$$

and where w is an N-size frequency selective window (e.g. the Hann one) with $w(n) > 0\ \forall n$ and $\hat{a}^2(x, f)$ is selected to be the (signal dependent) inverse of the masking threshold at frequency $f/L \cdot f_s$ (f_s denotes the working sample rate), given by

$$\hat{a}^2(x, f) = N c_1 \sum_{i=1}^{P} \frac{|\hat{h}_i(f)|^2}{\sum_{f'=0}^{L-1} |\hat{h}_i(f')|^2\, |\widehat{xw}(f')|^2 + c_2}, \quad f = 0, 1, \ldots, L-1. \tag{11}$$

In this last equation, $c_1 > 0$ and $c_2 > 0$ are calibration constants, being c_1 independent of N and $h_i(n)$ is the (L-point) impulse response of the cascade of the filter simulating the behavior of the outer- and middle ear with the i^{th} gammatone filter of the filter-bank of size P, simulating the band-pass characteristic of the basilar membrane of the cochlea [1]. These filters h_i are assumed to be absolutely summable.

In accordance to what was explained in the introduction (section 1), we would like to point out that this distortion measure can be transformed into a mean-square-error (MSE) by normalization of the input signal. Indeed, looking at the form (9) of the distortion measure, we can define

$$\hat{x}'(f) = \hat{a}(f)\, \widehat{xw}(f) \qquad \hat{y}'(f) = \hat{a}(f)\, \widehat{yw}(f) \tag{12}$$

and work on the normalized domain x', y', i.e. quantize and transmit x' and recover y from $\hat{y}'$. Nevertheless, if we use the normalization directly, the square-root of the inverse of the masking threshold $\hat{a}$ has to be transmitted to the receiver to perform the inverse normalization

$$\widehat{yw}(f) = \frac{\hat{y}'(f)}{\hat{a}(f)}, \tag{13}$$

with the consequence of an intorable overhead in a multiple description coding scenario with high number of channels.

3. A SUBOPTIMAL COMPANDING SCHEME

In section 2, we have shown the basic elements necessary to construct a companding scheme for the distortion measure in question. Based on those elements, we will construct a sub-optimal companding scheme in this section.

3.1. Sensitivity Matrix

We will first work out the optimality condition (5), check whether optimality can be achieved, and then, due to arriving at a negative result, we will construct a sub-optimal compressor based on the same condition. As we can see through the condition equation, to express it explicitely we have to calculate the sensitivity matrix (3). Those calculations were done in [11], where the cases $L = N$ and $L \neq N$ were treated separately, being the result

$$M(x) = \Lambda_w M_c(x) \Lambda_w. \tag{14}$$

with

$$M_c(x) = D_N^{\mathrm{H}} \Lambda_{N \hat{a}(x)^2} D_N, \tag{15}$$

where Λ_u denotes a diagonal matrix with the elements of the vector u in the diagonal, $^{\mathrm{H}}$ denotes hermitian transposition and D_N denotes the normalized DFT matrix [14]. M_c is a circulant matrix due to being diagonalized by the DFT matrix [15]. In turn, for $L \neq N$, the result is

$$M(x) = \Lambda_w M_t(x) \Lambda_w, \tag{16}$$

with

$$[M_t(x)]_{ml} = [M_c(x)]_{ml}, \quad m,l = 0,1,\ldots,N-1 \quad (17)$$

where in this case M_c is a larger L-by-L matrix, given by

$$M_c(x) = D_L^{\mathrm{H}} \Lambda_{L\hat{a}(x)^2} D_L. \quad (18)$$

The matrix M_t is a cropped version of the circulant M_c, being consequently a Toeplitz matrix. Due to the difficulty in dealing with Toeplitz matrices, namely in calculating their eigenvalues, determinant, inverse and other functions, M_t was approximated by a circulant matrix $M_{t,a}$ on basis of asymptotic equivalence results of [15]. The approximation $M_t \approx M_{t,a}$ will therefore deliver exact results as $N \to \infty$ and good approximations for the sizes of N tipically dealt with in audio coding $N \sim 1024$. After some calculations, the approximation resulted in

$$M_{t,a}(x) = D_N^{\mathrm{H}} \Lambda_{N\hat{\hat{a}}^2} D_N, \quad (19)$$

where

$$\hat{\hat{a}}^2(x,f) = \frac{L}{N}\, \hat{a}^2\left(x, \frac{L}{N} f\right), \quad f = 0,1,\ldots,N-1 \quad (20)$$

is a decimated, energy corrected version of the masking threshold. It was also shown, for N not too small, that we can decimate the band-pass filters instead of the masking threshold, i.e. define

$$\hat{\hat{h}}_i(f) \stackrel{\text{def}}{=}; \sqrt{\frac{L}{N}}\, \hat{h}_i\left(\frac{L}{N} f\right), \quad f = 0,1,\ldots,N-1 \quad (21)$$

and use

$$\hat{\hat{a}}^2(x,f) = Nc_1 \sum_{i=1}^{P} \frac{|\hat{\hat{h}}_i(f)|^2}{\sum_{f'=0}^{N-1} |\hat{\hat{h}}_i(f')|^2\, |\widehat{xw}_N(f')|^2 + c_2}, \quad f = 0,1,\ldots,N-1, \quad (22)$$

where $\hat{u}_N$ stands for the N-point DFT of u.

3.2. Compressor

Having delivered the form of the sensitivity matrix in section 3.1, we will work now more deeply out what condition the optimal compressor should satisfy and develop a compressor using the derived optimality condition.

From equations (5), (6) and (16) with the approximation (19) for $L \neq N$) or (14) and (15) for $L = N$, we can see that a possible square-root is

$$F'(x) = \sqrt{M(x)} = D_N^{\mathrm{H}} \Lambda_{\sqrt{N}\hat{\hat{a}}(x)} D_N \Lambda_w, \quad (23)$$

where for $L = N$, $\hat{a}$ sould be used instead of $\hat{\hat{a}}$ (in this case $\hat{a} = \hat{\hat{a}}$ anyway, but the origin of the equations is different). In [11], it was proven that no optimal companding scheme exists for the specific sensitivity matrix given by (23) and for all sensitivity matrices given by the left-multiplication of equation (23) by a signal-independent orthogonal matrix U. The signal-dependent case was left for future work.

A suboptimal companding scheme was thus developed, and $U(x)$ was set to I. The proposed scheme windows the input signal, goes into the frequency domain using the DFT, applies a non-linear function G and goes back into the time domain with the inverse DFT. In a block diagram, the scheme can be synthesized as depicted in figure 2. In mathematical terms, the compressor is given by

$$F(x) = \frac{D_N^H}{\sqrt{N}}\, G(\sqrt{N} D_N \Lambda_w x). \quad (24)$$

It is easy to see through the application of the chain rule [16] that the optimality condition (23) degenerates into

$$G'(z) = \sqrt{N} \Lambda_{\hat{\hat{a}}(x(z))} \equiv \sqrt{N} \Lambda_{\hat{\hat{a}}(z)}, \quad (25)$$

where $x(z) = \Lambda_w^{-1} D_N^H z / \sqrt{N}$ and we use a slight abuse of notation, saying that now $\hat{\hat{a}}$ is a function of z directly.

There are N^2 equations and N unknowns in equation (25), meaning that the equation system is overdetermined. To build the compressor, we discard the $N^2 - N$ equations outside the diagonal elements, and choose to satisfy only the N equations on the diagonal. This negligence of equation results in a sub-optimal compressor, since with the resulting compressor the off-diagonal elements of its Jacobian matrix will not be 0. Calculations show that the resulting compressor is of the form

$$\tilde{G}_m(z) = \sqrt{N}\, \Gamma_m(z)\, z(m), \quad m = 0,1,\ldots,N-1, \quad (26)$$

where the expression for the component m was given, where the tilde stands for sub-optimality, and where the signal-dependent gain Γ has a component m given by

$$\Gamma_m(z) = \int_0^1 \gamma_m(z,t)\, \mathrm{d}t \quad (27)$$

with

$$\gamma_m(z,t) = \sqrt{Nc_1 \sum_{i=1}^{P} \frac{|\hat{\hat{h}}_i(m)|^2}{\|\Lambda_{\hat{\hat{h}}_i} z\|^2 + 2|\hat{\hat{h}}_i(m)|^2\, |z(m)|^2\, (t-1) + c_2}} \quad (28)$$

for $m \neq 0$ and $m \neq N/2$, and

$$\gamma_m(z,t) = \sqrt{Nc_1 \sum_{i=1}^{P} \frac{|\hat{\hat{h}}_i(m)|^2}{\|\Lambda_{\hat{\hat{h}}_i} z\|^2 + |\hat{\hat{h}}_i(m)|^2 z^2(m)\, (t^2-1) + c_2}} \quad (29)$$

for $m = 0$ or $m = N/2$.

$$x \longrightarrow \boxed{\times\ \text{window}} \longrightarrow \boxed{\text{DFT}} \longrightarrow \boxed{G(\cdot)} \longrightarrow \boxed{\text{iDFT}} \longrightarrow F(x)$$

Fig. 2. Block Diagram of the compressor $F(x)$

3.3. Taylor Expansion

Equations (28) and (28) show that the integrand of (27) is given in terms of a sum of $\|\Lambda_{\hat{\tilde{h}}_i} z\|^2 + c_2$ with a term proportional to the m^{th} component of the same squared norm, $|\hat{\tilde{h}}_i(m)|^2\,|z(m)|^2$. The proportionality constant (take $n \neq 0, N/2$ as an example) $2(t-1)$ is on the order of the units (note that $t \in [0,1]$), and as consequence, the dependence on t gets weaker and weaker with increasing vector dimension N – the m^{th} component of the squared norm gets more and more neglegible with respect to the whole squared-norm. This motivates a Taylor expansion of Γ_m around, say, around $t = 1$ with order M (for the full motivation, see [11]).

The Taylor expansion results in

$$\Gamma_{m,M}(z) = \sum_{k=0}^{M} \frac{(-1)^k}{(k+1)!} \frac{\partial^k \gamma_m}{\partial t^k}(z,1), \tag{30}$$

where the derivatives of γ_m with respect to t are obtained by a recursive equation, which computes a certain order k using all previously calculated orders of the same quantity (orders smaller than k) and an "inhomogeneous" function sequence (for each m), with the property that the first element of the sequence is proportional to γ_m^2 and the derivative of the k^{th} element is the $(k+1)^{\text{st}}$ element.

Note that if we use $M = 0$ we get the very simple expression

$$\Gamma_{m,0}(z) = \gamma_m(z,1) = \hat{\tilde{a}}(z,m), \tag{31}$$

i.e., in its most simple form, the compressor multiplies the (windowed) input signal by the square-root of the inverse of the masking threshold in the frequency domain. If you compare the compressor (26) with the normalization step (12), you will notice that asymptotically, the compressor does exactly the thing we wanted to avoid: normalize the input signal by perceptual weights. Nevertheless, instead of transmitting perceptual weights through the channel, at the receiver we now only have to apply the inverse of the compressor (26) (the expander); it is not necessary to use the weights at the receiver. How to calculate the inverse of the compressor will be the subject of subsection 3.5.

3.4. Analysis of the Compressor

In [11], the compressor was analysed in terms of the optimality condition (25). Its effective Jacobian matrix was calculated, having been the result given in terms of recursive equations as well. Note that the Jacobian matrix of $\tilde{G}$ is different from the one in (25) due to the discarding of the diagonal equations on this last equation; even the diagonal elements are not exactly the same due to the execution of the Taylor expansion with finite M. Furthermore, it was noted that the Jacobian matrix is given by a low-rank update of a matrix defined there as being *cross diagonal*, meaning that it is of the form

$$V(z) = \Lambda_{v_f(z)} + \Lambda_{v_b(z)} D_N^2, \tag{32}$$

where, in this particular case, v_f is a real symmetric ($v_f(m) = v_f(N-m)$) vector and v_b is a hermitian symmetric ($v_b(m)^* = v_b(N-m)$) complex vector. The expression of the Jacobian matrix is then given by the update

$$\tilde{G}'(z) = \sqrt{N}(V(z) + A(z)H(z)^{\mathrm{H}}), \tag{33}$$

where $A(z)$ and $H(z)$ are N-by-P tall matrices ($P \ll N$). For a higher detail, the results for the Jacobian matrix can be consulted in [11].

Additionally, the asymptotic ($N \to \infty$) optimality of the compressor (26) was established in [11]. Intuitively, as the term $\|\Lambda_{\hat{\tilde{h}}_i} z\|^2 + c_2$ gets dominant with respect to the term $|\hat{\tilde{h}}_i(m)|^2\,|z(m)|^2$ in the Taylor expansion of γ_m (equations (28) and (29)), asymptotically, the compressor for an arbitrary $M = 0, 1, \ldots, \infty$ behaves like the compressor for $M = 0$. The Jacobian matrix of the compressor for $M = 0$ can be calculated, being the result the expression

$$\tilde{G}'(z)\Big|_{M=0} = \sqrt{N}\Lambda_{\hat{\tilde{a}}(z)}\,(E + \mathrm{I}) \tag{34}$$

with E having as components

$$[E]_{ml} = \frac{-\sum_{i=1}^{P} \frac{|\hat{\tilde{h}}_i(m)|^2 |\hat{\tilde{h}}_i(l)|^2\, z(m)z(l)^*}{\left[\|\Lambda_{\hat{\tilde{h}}_i} z\|^2 + c_2\right]^2}}{\sum_{i=1}^{P} \frac{|\hat{\tilde{h}}_i(m)|^2}{\|\Lambda_{\hat{\tilde{h}}_i} z\|^2 + c_2}}. \tag{35}$$

As intuition indicates, the term $1/(\|\Lambda_{\hat{\tilde{h}}_i} z\|^2 + c_2)$ in the denominator of (35) decays with N, but its square, present in the numerator, decays faster. It is thus also intuitive that $E \to 0$ with increasing N and that $\tilde{G}'$ fulfills the optimality condition (25) asymptotically, making the rate-loss (7) vanish asymptotically. A formal derivation can be found in [11].

3.5. Expander

After having derived a suboptimal compressor and having analysed it in terms of its asymptotic behaviour with increasing vector dimension, it is now natural to complete the whole chain of the companding scheme depicted in 1 by building an expander which implements the inverse function of the suboptimal compressor.

It can be proven [11] that, at least for $M = 0$, the compressor is invertible. We are thus sure that if we use an adequate numerical method to solve the equation

$$\tilde{G}(z) = \hat{\xi} \tag{36}$$

for a certain $\hat{x}i$ with a sufficiently close initial estimate $z^{(0)}$, the solution exists, $z = \tilde{G}^{-1}(\hat{\xi})$ and the method will converge to it. For the inversion of the complete compressor function $F(x)$ we only have to use the equivalence relation

$$\xi = \tilde{F}(y) = \frac{D_N^{\mathrm{H}}}{\sqrt{N}}\,\tilde{G}(\sqrt{N}D_N\Lambda_w y) \iff$$
$$\iff y = \tilde{F}^{-1}(\xi) = \Lambda_w^{-1}\frac{D_N^{\mathrm{H}}}{\sqrt{N}}\tilde{G}^{-1}(\sqrt{N}D_N\xi), \tag{37}$$

which is directly deductible from the definition of G (or the equivalent for the suboptimal compressor) (24).

For the first iteration of the method, we propose to rearrange equation (26) as

$$z(m) = \frac{\hat{\xi}}{\sqrt{N}\,\Gamma_m(z)} \tag{38}$$

and perform a fixed-point iteration of it, i.e., build a second estimate $z^{(1)}$ with

$$z^{(1)}(m) = \frac{\hat{\xi}}{\sqrt{N}\,\Gamma_m(z^{(0)})}, \quad m = 0, 1, \ldots, N-1. \tag{39}$$

This first iteration can be motivated as follows. In the first place, it can be shown that $\Gamma(z)$ does not dependent on the phase of the components of z, but only on their magnitude. The execution of iteration (39) has thus the advantage that it wipes out phase differences between the initial estimate $z^{(0)}$ and the final desired value z of (36). In other words, if $z^{(0)}$ is very close to z up to phase differences, the second estimate $z^{(1)}$ will be an excellent initial estimate for the next numerical method, which will be used to "fine-tune" the obtained vector $z^{(1)}$. The same argument can be applied if $z^{(0)}$ has a similar masking threshold $\hat{\tilde{a}}^{-2}$ than the one of z. Remember that the 0^{th} order term of Γ_m is exactly the square root of the inverse of the masking threshold and that for high vector dimension only this term matters (cf. equation (31) and the considerations on the asymptotical compressor in subsection 3.4) so that for similar masking thresholds we get similar $\Gamma(z)$.

For obtaining similar masking thresholds, we propose to use the vector z obtained from running the numerical methods described in this section on the last audio frame as an initial estimate $z^{(0)}$ for this frame; for this choice the masking threshold will not have changed much between these two frames due to the stationarity of typical audio signals in the order of the dozens of miliseconds. By using equation (39) in this way, we thereby get a good initial estimate for fine-tuning.

For the fine-tuning process we propose to use the Broyden's method [17], due to its supralinear convergence [17] but lower complexity than the Newton's method. The equations defining this method are

$$z^{(n+1)} = z^{(n)} - \tilde{J}_n^{-1}\left(\tilde{G}(z^{(n)}) - \hat{\xi}\right), \quad n = 1, 2, \ldots \tag{40}$$

with

$$\tilde{J}_n = \tilde{J}_{n-1} + \frac{\left(\Delta\tilde{G}^{(n)} - \tilde{J}_{n-1}\Delta z^{(n)}\right)\Delta z^{(n)\mathrm{H}}}{\|\Delta z^{(n)}\|^2}, \quad n = 2, 3, \ldots \tag{41}$$

and the initial matrix equal to the Jacobian matrix of the compressor at the initial guess

$$\tilde{J}_1 = \tilde{G}'(z^{(1)}). \tag{42}$$

A memory optimization taking advantage of the form (33) of the Jacobian matrix of the compressor and of the one-rank updates of (41) (and of its inverse) was done in [11] to enable the execution of the expander for large N without calculating the N-by-N matrices J_n explicitely. The largest matrix size stored in memory after the optimization was N-by-P (remember that $P \ll N$).

4. SIMULATIONS

4.1. Distortion-rate performance

In this section we will simulate the companding scheme, showing simulation figures and tables that corroborate the theoretical results. After calculating the rate-distortion function for the distortion measure in question, using results of [18], and the rate-loss of equation (7), simulations of the distortion-rate performance of the compander with the parameters of table 1 were done. We used calculations for the non-approximated version of the sensitivity matrix for $N \le 1024$ (i.e. M_t instead of $M_{t,a}$) and the calculations for the approximated version for $1024 \le N < 8192$ ($M_{t,a}$ instead of M_t). From $N = 8192$ on (inclusive) we used the matrix M_c (in this case $M_t = M_{t,a} = M_c$. Furthermore, the expected values in the equations were replaced by statistical averages (several realizations of the signal X were emitted), with lower number of realizations for higher vector size N and vice-versa. The reason for decreasing the number of realizations with higher N is that the quantities for which we should estimate the expected value are averages themselves (normalized traces and normalized sums of logarithms of eigenvalues). Assuming that these inner quantities that we are averaging (the eigenvalues and their logarithms) are well behaved, the inner average is consistent, so that these quantities have a low variance for high N. The outer averages (the ones that replace the expectation) reduce the variance

Parameter	Value
Input signal x	Gaussian i.i.d. samples with zero mean, variance σ^2
Power of the input signal σ^2	$0{,}01^2$
Vector dimension N	Powers of 2 from 256 to 65536
DFT dimension L	8192 for $N \leq 8192$, N otherwise
Window w	Hamming, $w(n) = 0{,}54 - 0{,}46\cos\left(2\pi\frac{n}{N}\right)$
Sample frequency f_s	48 kHz
Quantizer	$\mathbb{Z}^N$ (componentwise uniform scalar quantization)
Inverse of the quantization step size, $1/s$	5 scales varying exponentially from 10^3 to 10^5
Order of Taylor expansion M	3

Table 1. Values for the parameters used in the simulations

furthermore, and thus the lower the N, the heavier that reduction needs to be performed (the higher the needed number of realizations) and vice-versa. In practice, we used on the order of 100 realizations for $N = 256$, of 10 realizations for $N = 1024$ and 1 realization for $N = 8192$ or higher.

As output, graphs like the one in 3 were delivered. This graph shows the distortion-rate performance of the companding scheme (for $N = 1024$, in this particular case). The rate-

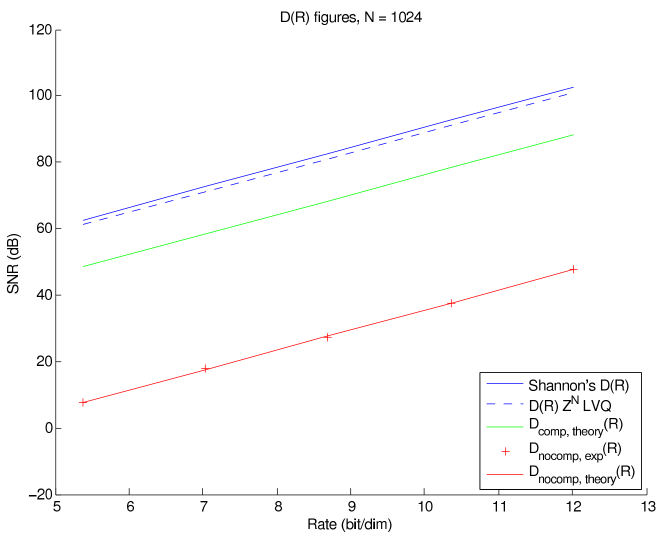


Fig. 3. Distortion-rate performance of the companding scheme for $N = 1024$. Distortions are in terms of the SNR in dB ($10\log_1 0(\sigma^2/D)$). The blue line represents the Shannon's distortion-rate function. The blue dotted line gives the best achievable $\mathbb{Z}^N$ lattice vector quantizer (LVQ) distortion-rate performance. The green and red lines give the performance of the companding scheme and of the identity compander $F(x) = F^{-1}(x) = x$. The red crosses are results obtained from quantizing the source directly (i.e. using the identity compander), and calculating the rate and distortion values directly.

loss (the horizontal distance between the blue dotted line and the green line in the graph) was extracted for several values of N, being the results for the *distortion-loss*, the correspondent quantity in the vertical axis of the graph, shown in table 2.

As we can see from figure 3, the scheme performs well with $N = 1024$, and from table 2 we can confirm that asymptotic optimality is achieved. Note that $N = 1024$ is the vector size which corresponds to the frame-size used in practice ($N/f_s = 21{,}3\,\mathrm{ms}$, where f_s is the sample frequency); (much) lower or higher values of N don't correspond to the frequency and time resolution of the human ear, respectively. Furthermore, for $N/f_s \gg 20\,\mathrm{ms}$ the audio signal can't be considered stationary.

Additionally, we computed the distortion-rate figures for $N = 1024$ with and without approximation of M_t by $M_{t,a}$, and overlaid the resulting curves. The resultant graph can be seen in figure 4. As we can see, the lines which correspont to each other are practically perfectly overlaid, which indicates that for $N = 1024$ we can approximate M_t by $M_{t,a}$ when $L \neq N$.

4.2. Taylor expansion

To test how good the approximation of the original compressor of section 3.2 by the Taylor expansion of section 3.3 is, and what order of M we must use to get a good approximation, we computed the contribution of the 0^{th} to the 6^{th} order terms of the signal dependent gain Γ of equation (30) (remember that the compressor essentially multiplies pointwise the windowed input in the frequency domain by the gain, equation (26)), for the vector size most useful in practice, $N = 1024$. The contributions can be seen in figure 5, depicted in double-logarithmic scale. The figure shows that already for $N = 1024$ we can use $M = 0$. If we want a finer detail we can use higher M, but the order of units will be more than enough. Indeed, the distance between the 0^{th} and 1^{st} order terms is at least $15{,}7$ dB, which means that the 1^{st} order term has values that are attenuated at least $6{,}1\times$ with respect to the values of the 0^{th} order term. For the 2^{nd} order term this distance is $23{,}9$ dB ($15{,}7\times$), and for the 6^{th} order $42{,}8$ dB ($138\times$).

N	256	512	1024	2048	4096	8192	16384	32768	65536
Distortion Loss (dB)	14,6	14,0	12,6	11,0	8,65	6,27	4,28	2,77	1,57

Table 2. Distortion Loss of the companding scheme for varying N

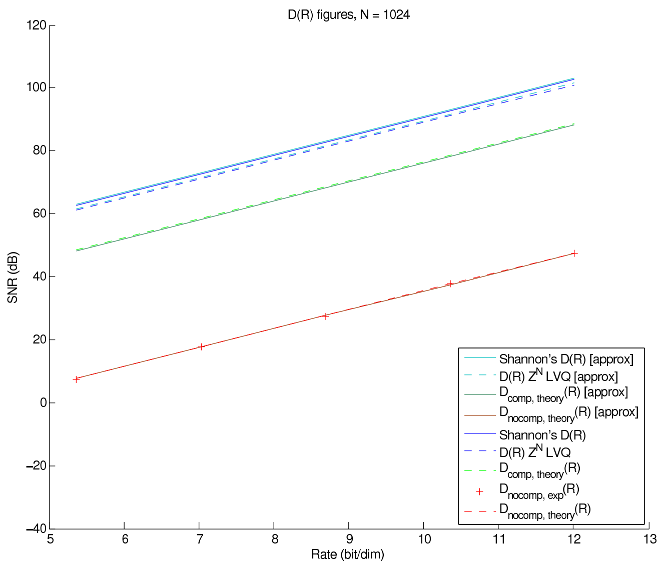


Fig. 4. Distortion-rate performance of the companding scheme for $N = 1024$. Comparison of the circulant matrix approximation with the real values. See 3 for the meaning of the colors (the dotted red and green here are exactly the same as the solid red and green in figure 3, respectively). The common colors between the two figures refer to the non-approximated version. The light-blue, dotted light-blue, dark green and brown lines correspond to the blue, dotted blue, green and red lines, respectively, but for the approximated version.

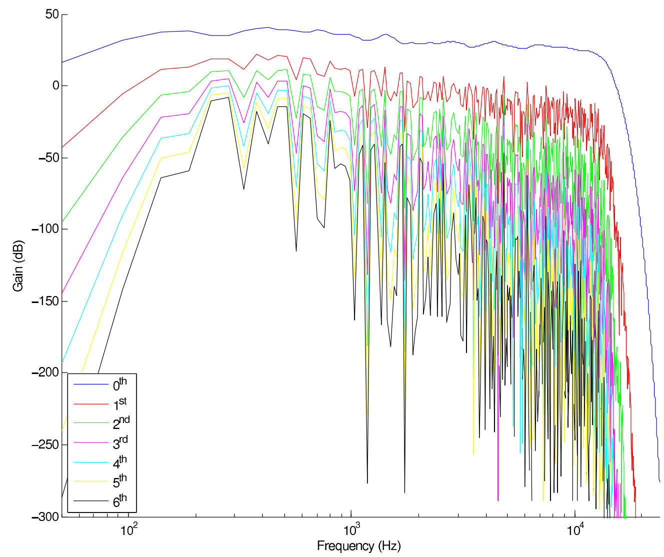


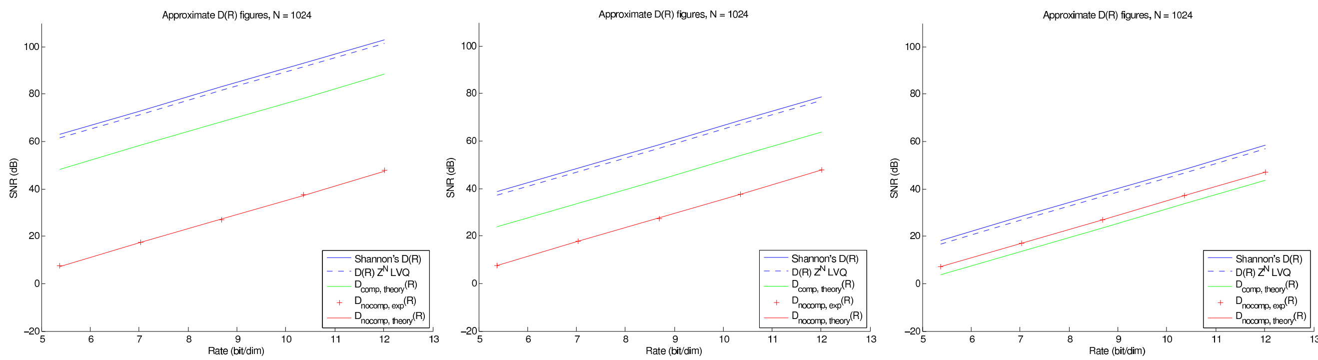
Fig. 5. Contribution of the j^{th} order term of the Taylor expansion of the signal-dependent Γ, $j = 0, 1, \ldots, 6$

4.3. Limitations

Figure 5 shows an important limitation that the developed multidimensional companding scheme has: as the gain Γ decreases abruptly with increasing frequency, the inversion of the compressor, although mathematically possible (at least for $M = 0$; see subsection 3.5), is a numerically badly conditioned problem. Using the rough rationale based on equation (38) that the expander multiplies the incoming signal pointwise by the inverse of Γ, we can see through figure 5 that the quantization noise present in the input of the expander will be heavily amplified in the high frequencies (imagine the characteristics upside-down, since the inverse in the linear domain corresponds to the symmetric in the log domain). The same happens for the very low frequencies. In practice, this means that two very different signals in the non-compressed domain will have very near compressed signals, up to numerical error.

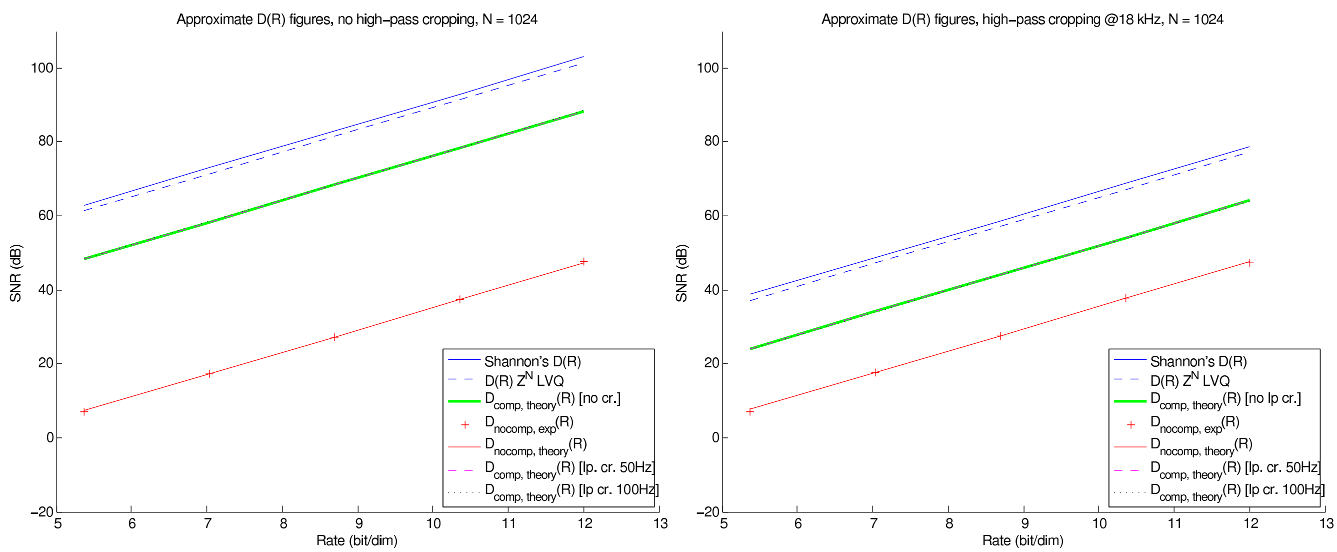
To deal with this limitation we can crop the gain Γ and the masking threshold $\hat{\tilde{a}}$ of the distortion measure, i.e., if the frequency is below a certain threshold in the low-frequency range, we use the boundary value (the value on the threshold), and if the frequency is above another threshold, in the high-frequency range, we do the same. Between the two thresholds, the original gain/masking threshold are conserved. For $N = 1024$, the distortion-rate performance of the compressor was simulated again cropping Γ and $\hat{\tilde{a}}$ in the high frequency range at $\{24, 18, 14\}$ kHz (the first case corresponds to not cropping as $f_s/2 = 24$ kHz), and then cropping in the low frequency range at $\{50, 100\}$ Hz for no high-frequency cropping and for high-frequency cropping at $18\text{k } Hz$. The results are shown in figures 6 and 7, respectively.

We observe a degradation of performance for decreasing threshold of the high-frequency crop: the distortion-rate function, the $\mathbb{Z}^N$ best distortion-rate performance and the distortion-rate performance of the compander all sink in the graphs, i.e., for the same rate, the best achievable and the actual distortions increase (the SNR decreases). At a certain point (in the figure at 14 kHz) the performance of the compander sinks below the performance of not doing anything, i.e., of using the identity compander $F(x) = x$. For the low-frequency cropping case no performance loss is obtained, as the compander curves lie on top of each other. For practical numerical good condition, a cropping on the low frequencies at 50 Hz and on the high frequencies at 18 kHz suffices. The graph 7(b) is thus the one that depicts the performance of the scheme in a practial situation.



(a) 24 kHz (no cropping) (b) 18 kHz (c) 14 kHz

Fig. 6. Distortion-rate performances for cropping in the high-frequency range at 24, 18 and 14 kHz



(a) no cropping (b) high-frequency cropping at 18 kHz

Fig. 7. Distortion-rate performances for cropping in the low-frequency range at 50 and 100 Hz, for no high-frequency cropping and for cropping at 18 kHz

5. CONCLUSIONS

A multidimensional companding scheme was developed for the perceptual distortion measure [1]. Although the developed scheme is suboptimal, it was proven (in [11]) that no optimal scheme exists under a wide range of solutions for the optimality condition (5), and that nevertheless the conceived scheme reached optimality asymptotically, i.e., with increasing vector dimension N. The scheme was developed in the frequency domain; in its most simple form, the compressor windows the input signal, applies a Discrete Fourier Transform (DFT), multiplies the input signal by the square-root of the inverse of its masking threshold, the square-root of equation (11), and then goes back into the time domain with the inverse DFT. This process is exactly the same as the process of normalization by perceptual weights of equation (12), but with the advantage that no perceptual weights have to be transmitted through the channel. At the receiver, an algorithm based on numerical methods, the expander, has to be run. The expander does not depend on the perceptual weights and it can be run using the previous audio frame, already available at the decoder, as its initial condition. The theoretical results and assumptions were corroborated with simulations.

6. REFERENCES

[1] S. van de Par, A. Kohlrausch, R. Heusdens, J. Jensen, and S. Jensen, “A perceptual model for sinusoidal audio coding based on spectral integration,” *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 9, pp. 1292–1304, June 2005.

[2] G. A. Soulodre, T. Grusec, M. Lavoie, and L. Thibault,

"Subjective evaluation of state-of-the-art 2-channel audio codecs," *J. Audio Eng. Soc.*, vol. 46, no. 3, pp. 164 – 176, March 1998.

[3] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451 – 515, April 2000.

[4] ISO/IEC 14496-3:2001, *Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio Edition 2001)*, 2001, ISO/IEC International Standard.

[5] B. Edler and G. Schuller, "Audio coding using a psychoacoustic pre- and post-filter," in *in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, June 2000, vol. 2, pp. 881 – 884.

[6] J. Østergaard, *Multiple-Description Lattice Vector Quantization*, Ph.D. thesis, Technical University of Delft, 2007.

[7] R. Vafin and W. Kleijn, "Entropy-constrained polar quantization and its application to audio coding," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 2, pp. 220 – 232, March 2005.

[8] P. Korten, J. Jensen, and R. Heusdens, "High resolution spherical quantization of sinusoidal parameters using a perceptual distortion measure," in *in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2005, vol. 3, pp. 177 – 180.

[9] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74 – 94, Sept. 2001.

[10] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: Multidimensional companding," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 548–561, March 1999.

[11] J. Crespo, "A multidimensional companding scheme for source coding with a perceptually relevant distortion measure," 2009.

[12] H. Purnhagen, "Advances in parametric audio coding," in *in Proc. 1999 IEEE Workshop on Applications of Signal Proc. to Audio and Acoustics, New Paltz, New York, USA*, 1999, pp. W99–1 – W99–4.

[13] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[14] P. C. Yip and K. Ramamohan Rao, *The Transform and Data Compression Handbook*, CRC, 2000.

[15] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*, Now Publishers, Norwell, Massachusetts, 2006.

[16] T. M. Apostol, *Calculus, Vol. 2: Multi-Variable Calculus and Linear Algebra with Applications*, vol. 2, Wiley, second edition, 1969.

[17] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, 1999.

[18] T. Linder and R. Zamir, "High-resolution source coding for non-difference distortion measures: The rate-distortion function," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 533–547, March 1999.