Real-Time Perceptual Moving-Horizon Multiple-Description Audio Coding

Østergaard, Jan; Quevedo, Daniel; Jensen, Jesper

Published in:
I E E E Transactions on Signal Processing

DOI (link to publication from Publisher):
10.1109/TSP.2011.2159601

Publication date:
2011

Document Version
Early version, also known as pre-print

Real-Time Perceptual Moving-Horizon Multiple-Description Audio Coding

Jan Østergaard*, *Member, IEEE*, Daniel E. Quevedo, *Member, IEEE*, and Jesper Jensen

***Abstract*—A novel scheme for perceptual coding of audio for robust and real-time communication is designed and analyzed. As an alternative to PCM, DPCM, and more general noise-shaping converters, we propose to use psychoacoustically optimized noise-shaping quantizers based on the moving-horizon principle. In moving-horizon quantization, a few samples look-ahead is allowed at the encoder, which makes it possible to better shape the quantization noise and thereby reduce the resulting distortion over what is possible with conventional noise-shaping techniques. It is first shown that significant gains over linear PCM can be obtained without introducing a delay and without requiring post-processing at the decoder, i.e., the encoded samples can be stored as e.g., 16-bit linear PCM on CD-ROMs, and played out on standards-compliant CD players. We then show that multiple-description coding can be combined with moving-horizon quantization in order to combat possible erasures on the wireless link without introducing additional delays.**

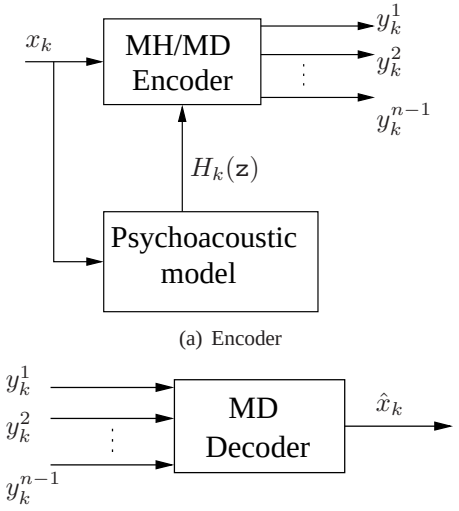
***Index Terms*—Low delay source coding, multiple-description coding, moving horizon quantization, perceptual audio coding**

I. Introduction

The aim of this work is to encode and communicate audio from a remote encoder (e.g., cell phone, ipod, CD player, radio, tv, concert) over a wireless link to a low power listening device e.g., a pair of hearing aids or head phones. Contrary to other applications, it is here essential that the latency is kept low. Low latency is important, primarily in order to avoid distortions due to a direct path acoustic signal reaching the eardrums earlier than the hearing aid output [1], but also to facilitate lip synchronicity in a real-time communication situation. We will assume that the tolerable latency is a few samples or at most up to a few milliseconds.

Due to battery and space considerations, the computational complexity at the decoder should be kept low. Thus, besides the cost of operating the antenna(s) and the demodulators, we only allow simple scaling and table look-up operations in this work.

Since the persons wearing the listening devices are often not spatially stationary, the transmission channel is susceptible to fading. In order to guarantee a certain degree of robustness towards channel impairments without introducing additional delay, we rely on multiple-description (MD) coding [2]. We consider the general case of n channels. For example, a hearing aid may have more than one receiving antennas, and, furthermore, since hearing aids are typically worn pairwise, the hearing aids may communicate with each other. Thus, several channels are available even in the single person situation.



(a) Encoder

(b) Decoder

Fig. 1. The encoder consists of two parts; the moving-horizon multiple-description *MH/MD Encoder* and the *Psychoacoustic model*.

MD coding was recently used for robust perceptual audio coding [3]–[5]. In [3], [4], the case of two descriptions was considered, whereas in [5] it was shown, that even with highly unreliable networks, it is possible to achieve audio streaming of acceptable quality by using more than two descriptions. In [4], [5], perceptual models were employed at the encoder in order to derive masked thresholds. These were used as perceptual weighting filters at the decoder and therefore needed to be encoded and transmitted to the decoder as side information, in addition to the encoded audio data. It turns out that the bit rate required for encoding the perceptual weighting filter is up to 8 kbps for mono audio signals with a sampling frequency of 44.1 kHz [4], [5]. Since the perceptual weighting filters are required in all the descriptions, the bit rate of the side information can be significant. Moreover, it is an open question how to optimally distribute the bit budget between the perceptual model and the actual audio data.

Jan Østergaard is with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark; janoe@ieee.org.

Daniel Quevedo is with the School of Electrical Engineering & Computer Science, The University of Newcastle, NSW 2308, Australia; dquevedo@ieee.org.

Jesper Jensen is with Oticon, Copenhagen, Denmark; jsj@oticon.dk.

This research was partially supported by the Danish Research Council for Technology and Production Sciences, grant no. 274-07-0383 and partially supported under Australian Research Council's Discovery Projects funding scheme (project number DP0988601).

To achieve perceptually efficient encoding without introducing large delays, we employ moving-horizon (MH) quantization techniques at the encoder [6]. MH quantization relies upon online optimization of a finite-horizon cost function and was recently cast in the framework of low delay audio coding [6]. In [6], given a fixed rather than a time-varying perceptual weighting filter, it was shown that, by increasing the optimization horizon, better performance could be achieved at the expense of increased complexity at the encoder. The delay of the design in [6], was dictated by that of the optimization horizon, i.e. was on the order of a few samples.

In the work presented here, we first extend [6] to the case of a *time-varying* perceptual weighting filter. A key feature of our design is that, as in [6], the perceptual weighting filter need not be transmitted as side information to the decoder. Thus, we avoid the issue of having to distribute the bits between the audio data and the perceptual weighting filters. We then provide a rate-distortion analysis of MH quantization. Subsequently, we show how one can combine MD coding and MH quantization in order to achieve robustness towards packet losses. The overall delay of the proposed design, depends upon the choice of perceptual model. For example, if the psychoacoustic model of MPEG1 layer 1 [7] is chosen, then the delay is about 6 ms. at 44.1 kHz. sampling frequency. We also show that significant gains over conventional linear PCM can be achieved with zero delay, by deriving the perceptual weighting filters from an approximation of the threshold in quiet of the human hearing system. Interestingly, if one leaves out entropy coding, the MH encoded samples may be stored as e.g., 16-bit linear PCM on CD-ROMs, and no post-processing is then required at the decoder. Thus, the encoded samples can be directly played out on any typical CD-player. The encoder and decoder of our proposal are presented in Fig. 1(a) and Fig. 1(b), respectively.

This paper is structured as follows: In Section II, we describe the setup, present known results on MH quantization for the case of fixed and time-invariant filters, and finally extend these results to include time-varying filters. Sections III and IV contain the main contributions, i.e., a rate-distortion analysis of single-description MH quantization and the proposed perceptual MH MD audio coding scheme, respectively. In Section V we show how to design the system in practice and provide extensive rate-distortion simulations. Conclusions appear in Section VI.

II. The Perceptual Moving Horizon Coder

In this section, we present background material on MH quantization. In particular, we revise the framework of [6], [8] and extend it to the case of time-varying filters.

A. Perceptual Moving Horizon Quantization

In MH quantization, the current scalar sample $x_k \in \mathbb{R}$ is combined with $N-1$ future samples and quantized using a vector quantizer $\mathcal{Q}_k^N(\cdot)$ [6]. Thus, the input to the quantizer is the N-dimensional vector $\vec{x}_k = (x_k, x_{k+1}, \cdots, x_{k+N-1})^T$ and the output of the quantizer, i.e. the quantized version of $\vec{x}_k$ is the vector $\vec{y}_k = (y_k, y_{k+1}, \cdots, y_{k+N-1})^T$. More precisely, given the current input vector $\vec{x}_k$, the quantizer $\mathcal{Q}_k^N(\cdot)$ minimizes a cost function, $J_k^N(\cdot)$, which includes perceptual weighting. For example, the cost function may be taken to be[1]

$$J_k^N(\vec{x}_k) \triangleq \sum_{i=k}^{k+N-1} \epsilon_i^2 = \|\vec{\epsilon}_k\|^2, \tag{1}$$

where $\epsilon_i \in \mathbb{R}$ is the perceptually weighted error at the ith time-lag, that is

$$\epsilon_i \triangleq \vec{h}_i * (\vec{x} - \vec{y}) \triangleq \sum_{n=0}^{K} h_{i,n}(x_{i-n} - y_{i-n}), \tag{2}$$

where

$$\vec{h}_i = (h_{i,0}, h_{i,1}, \ldots, h_{i,K})^T$$

denotes the set of filter coefficients of the perceptual weighting filter $H_i(\mathtt{z})$ to be used at time i (and $*$ is the linear convolution operator). Thus,

$$\epsilon_i(\mathtt{z}) = H_i(\mathtt{z})(\vec{x}(\mathtt{z}) - \vec{y}(\mathtt{z}))$$

and

$$H_i(\mathtt{z}) = 1 + \sum_{n=1}^{K} h_{i,n} z^{-n} \tag{3}$$

is a causal linear time varying filter of finite order K with a direct feedthrough and thus $\vec{h}_{i,0} = 1, \forall i$.

It follows that, given an input vector $\vec{x}_k$, the (locally) optimal output vector $\vec{y}_k^* = \mathcal{Q}_k^N(\vec{x}_k)$ (locally, for the current time k) is given by

$$\vec{y}_k^* = \underset{\vec{y}_k \in \mathcal{Y}_k, \vec{y}_k = \mathcal{Q}_k^N(\vec{x}_k)}{\arg\min} J_k^N(\vec{x}_k) \tag{4}$$

where $\mathcal{Y}_k$ denotes the alphabet (or codebook) of $\vec{y}_k$.

The output of the MH encoder is then simply taken to be y_k, i.e. the first sample of the quantized vector $\vec{y}_k^*$. Thus, an MH encoder consists of the non-linear map $\mathcal{Q}_k^N(\vec{x}_k) = \vec{y}_k^*$ which is followed by a function that picks out the scalar element y_k. At any time k, the MH encoder therefore takes as input the current sample x_k (as well as $N-1$ future samples) and outputs a single sample y_k.

B. State-Space Interpretation

Since we are working with time varying filters it is convenient to formulate the problem in the state space domain.

An equivalent minimal state-space form for the filter $H_k(\mathtt{z})$ is, see, e.g., [9]

$$H_k(\mathtt{z}) = 1 + C_k(\mathtt{z}I - A)^{-1}B \tag{5}$$

[1]The cost function $J_k^N(\cdot)$ depends upon the current input vector $\vec{x}_k$, the choice of reconstruction alphabet $\mathcal{Y}_k$ containing the candidate output vectors $\vec{y}_k$, and the perceptual weights $\vec{h}_i$. Moreover, in the next section, we will extend the cost function so that it also depends upon a state vector. To keep the notation brief, we will simply write $J_k^N(\vec{x}_k)$ throughout the document.

where $A \in \mathbb{R}^{K\times K}, B \in \mathbb{R}^{K\times 1}$, and $C_k \in \mathbb{R}^{1\times K}$ are given by

$$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, C_k^T = \begin{bmatrix} h_{k,1} \\ h_{k,2} \\ \vdots \\ h_{k,K} \end{bmatrix} \tag{6}$$

and are related to the sequence of filters $\{\vec{h}_k\}$ through [9]

$$h_{k,n} = C_k A^{n-1} B, \quad n = 1, \ldots, K, \quad k = 0, \ldots . \tag{7}$$

With this, we can express the weighted error $\epsilon_k \in \mathbb{R}$ as given by (2) in state-space form, that is

$$\vec{z}_{k+1} = A\vec{z}_k + B(x_k - y_k) \tag{8}$$

$$\epsilon_k = C_k \vec{z}_k + (x_k - y_k) \tag{9}$$

where $\vec{z}_k \in \mathbb{R}^K$ is the current system state vector given by

$$\vec{z}_k = [x_{k-1} - y_{k-1}, x_{k-2} - y_{k-2}, \ldots, x_{k-K} - y_{k-K}]^T. \tag{10}$$

C. Cost Function with Terminal State Weighting

As mentioned in Section II-A, we will make use of a cost function, which includes perception, cf. (1). In the MH quantization literature, it has been suggested to include state-weighting on the final state $\vec{z}_{k+N}$ within the cost function [8], [10].[2] In this work, the cost function will be based on the following expression:

$$J_k^N(\vec{x}_k) \triangleq \|\vec{\epsilon}_k\|^2 + \|\vec{z}_{k+N}\|_P^2, \tag{11}$$

where $\vec{\epsilon}_k = [\epsilon_k, \epsilon_{k+1}, \ldots, \epsilon_{k+N-1}]^T$ and where the latter term provides a final-state weigthing via a positive semidefinite matrix $P \in \mathbb{R}^{N\times N}$, i.e., we have $\|\vec{z}_{k+N}\|_P^2 = \vec{z}_{k+N}^T P \vec{z}_{k+N}$.

We will now express (11) from a state-space point of view. To do so, we iterate (9) (as was done in [6]) in order to obtain

$$\begin{aligned} \epsilon_{k+1} &= C_{k+1}A\vec{z}_k + C_{k+1}B(x_k - y_k) + (x_{k+1} - y_{k+1}) \\ \epsilon_{k+2} &= C_{k+2}A^2\vec{z}_k + C_{k+2}AB(x_k - y_k) \\ &\quad + C_{k+2}B(x_{k+1} - y_{k+1}) + (x_{k+2} - y_{k+2}) \\ &\vdots \end{aligned}$$

From the above, we deduce that the perceptually weighted error can be written as

$$\|\vec{\epsilon}_k\|^2 = \|\Psi_k(\vec{x}_k - \vec{y}_k) + \Gamma_k \vec{z}_k\|^2, \tag{12}$$

where $\Psi_k \in \mathbb{R}^{N\times N}$ is the matrix with unit determinant given by

$$\Psi_k = \begin{bmatrix} h_{k,0} & 0 & \cdots & \cdots & 0 \\ h_{k+1,1} & h_{k+1,0} & 0 & & \vdots \\ h_{k+2,2} & h_{k+2,1} & h_{k+2,0} & 0 & \vdots \\ \vdots & & & \ddots & 0 \\ h_{k+N-1,N-1} & \cdots & \cdots & h_{k+N-1,1} & h_{k+N-1,0} \end{bmatrix} \tag{13}$$

[2]The motivation behind using final state weighting is partly to stabilize the feedback loop by approximating the effect of the infinite-horizon behavior [8], [10]. For example, in certain cases, it may be useful (from a stabilization point of view) to choose P_k so that it satisfies the Lyapunov equation $A^T P_k A + C_k^T C_k = P_k$, cf. [8], [11].

and $\Gamma_k \in \mathbb{R}^{N\times K}$ satisfies

$$\Gamma_k = \left[C_k^T, (C_{k+1}A)^T, \ldots, (C_{k+N-1}A^{N-1})^T\right]^T. \tag{14}$$

Following a similar recursive principle, the final state $\vec{z}_{k+N}$ can be written as

$$\vec{z}_{k+N} = A^N \vec{z}_k + M(\vec{x}_k - \vec{y}_k), \tag{15}$$

where

$$M \triangleq [A^{N-1}B, A^{N-2}B, \ldots, AB, B]. \tag{16}$$

With this, the cost function $J_k^N(\vec{x}_k)$ can be written as

$$\begin{aligned} J_k^N(\vec{x}_k) &= \|A^N \vec{z}_k + M(\vec{x}_k - \vec{y}_k)\|_P^2 \\ &\quad + \|\Psi_k(\vec{x}_k - \vec{y}_k) + \Gamma_k \vec{z}_k\|^2. \end{aligned} \tag{17}$$

D. Nearest Neighbor Euclidean Vector Quantization

In this section, we use ideas of [6], [8] and show that the MH quantizer can be implemented as a nearest neighbor vector quantizer by utilizing appropriate mappings in the state-space domain.

Let us define $\Phi_k \in \mathbb{R}^{N\times N}$ as the positive semidefinite matrix square root in $\Phi_k^T\Phi_k \triangleq \Psi_k^T\Psi_k + M^T P M$ and rewrite the cost function (17) as

$$\begin{aligned} J_k^N(\vec{x}_k) &= \|\vec{y}_k\|_{\Phi_k^T\Phi_k}^2 \\ &\quad - 2\langle \vec{y}_k, \Phi_k^T\Phi_k\vec{x}_k + (\Psi_k^T\Gamma_k + M^T P A^N)\vec{z}_k\rangle + \Xi_k(\vec{x}_k, \vec{z}_k) \end{aligned} \tag{18}$$

$$\begin{aligned} &= \|\vec{y}_k\|_{\Phi_k^T\Phi_k}^2 - 2\langle \Phi_k\vec{y}_k, \Phi_k\vec{x}_k + \Phi_k^{-T}(\Psi_k^T\Gamma_k + M^T P A^N)\vec{z}_k\rangle \\ &\quad + \Xi_k(\vec{x}_k, \vec{z}_k), \end{aligned} \tag{19}$$

where the function $\Xi_k(\vec{x}_k, \vec{z}_k)$ at time k is independent of $\vec{y}_k$ and given by

$$\begin{aligned} \Xi_k(\vec{x}_k, \vec{z}_k) &= \|\vec{x}_k\|_{\Phi_k^T\Phi_k}^2 + 2\langle \vec{x}_k, (\Psi_k^T\Gamma_k + M^T P A^N)\vec{z}_k\rangle \\ &\quad + \|\vec{z}_k\|_{\Gamma_k^T\Gamma_k + (A^N)^T P A^N}^2. \end{aligned} \tag{20}$$

Inspired by (19), we now let $\vec{\xi}_k \triangleq \Phi_k\vec{y}_k$ and introduce the metric $f_k^{\vec{w}} : \mathbb{R}^N \to \mathbb{R}$ defined as

$$f_k^{\vec{w}}(\vec{\xi}_k) \triangleq \|\vec{\xi}_k\|^2 - 2\vec{\xi}_k^T\vec{w}_k, \tag{21}$$

where

$$\vec{w}_k \triangleq \Phi_k\vec{x}_k + \Phi_k^{-T}(\Psi_k^T\Gamma_k + M^T P A^N)\vec{z}_k. \tag{22}$$

With this notation, $J_k^N(\vec{x}_k) = f_k^{\vec{w}}(\vec{\xi}_k) + \Xi_k(\vec{x}_k, \vec{z}_k)$, which implies that the optimal $\vec{y}_k^*$ is given by

$$\vec{y}_k^* = \arg\min_{\vec{y}_k \in \mathcal{Y}_k} J_k^N(\vec{x}_k) = \Phi_k^{-1} \arg\min_{\vec{\xi}_k \in \Phi_k\mathcal{Y}_k} f_k^{\vec{w}}(\vec{\xi}_k). \tag{23}$$

From (21), it may be observed that $f_k^{\vec{w}}(\vec{\xi}_k)$ has isocontours (level sets) that are shifted spheres in $\mathbb{R}^N$ and centered at $\vec{w}_k$. Thus, for any $\vec{\xi}_k', \vec{\xi}_k'' \in \mathcal{S}_c$, where $\mathcal{S}_c \triangleq \{\vec{\xi}_k \in \mathbb{R}^N : f_k^{\vec{w}}(\vec{\xi}_k) = c\}$, for some $c \in \mathbb{R}$, it follows that $\|\vec{\xi}_k' - \vec{w}_k\| = \|\vec{\xi}_k'' - \vec{w}_k\|$. Clearly, the optimal $\vec{\xi}_k$ should therefore be chosen as close as possible to $\vec{w}_k$ and we establish the following relationship:

$$\vec{y}_k^* = \Phi_k^{-1} \arg\min_{\vec{\xi}_k \in \Phi_k\mathcal{Y}_k} f_k^{\vec{w}}(\vec{\xi}_k) = \Phi_k^{-1}\mathcal{Q}_{\Phi_k\mathcal{Y}_k}(\vec{w}_k), \tag{24}$$

where $\mathcal{Q}_{\Phi_k\mathcal{Y}_k}(\cdot) \in \Phi_k\mathcal{Y}_k$ is a conventional nearest-neighbor (Euclidean) vector quantizer with code vectors in the transformed alphabet given by $\Phi_k\mathcal{Y}_k$.

E. Noise Shaping Architecture

In this section, we show that the closed-form expression for the optimizer given in (24) allows us to describe the system by a noise-shaping architecture, which can be implemented efficiently.

As is evident from (17), the optimizing vector $\vec{y}_k^*$ should be chosen such that the filtered error vectors $\Psi_k(\vec{x}_k - \vec{y}_k)$ and $M(\vec{x}_k - \vec{y}_k)$ are close to the mirror images of $A^N\vec{z}_k$ and $\Gamma_k\vec{z}_k$, respectively. Thus, the past decisions contained in $\vec{z}_k$ affect current and future decisions. We will now follow the approach of [6], [8] and show that the MH quantizer has an equivalent noise-shaping architecture.

Let $G_k(\mathtt{z})$ be defined as

$$G_k(\mathtt{z}) \triangleq (\mathtt{z}I - A)^{-1}B, \tag{25}$$

where the square matrix $\mathtt{z}I$ contains the one-step advance (forward) operator $\mathtt{z}$ on its diagonal and let $F_k(\mathtt{z}) \in \mathbb{R}^{N\times N}$ be defined as

$$F_k(\mathtt{z}) \triangleq \Phi_k^T\Phi_k + (\Psi_k^T\Gamma_k + M^TPA^N)G(\mathtt{z})[1\ 0\cdots 0], \tag{26}$$

where the unit row vector $[1\ 0\cdots 0]$ is of dimension $1\times N$ as will become clear below. Then, writing $\vec{z}_{k+1}$ in terms of the forward operator, i.e., $\vec{z}_k\mathtt{z}I = \vec{z}_{k+1}$, inserting into (8) and solving for $\vec{z}_k$ yields

$$\begin{aligned}\vec{z}_k &= (\mathtt{z}I - A)^{-1}B(x_k - y_k) &(27)\\ &= G_k(\mathtt{z})[1\ 0\ \cdots\ 0](\vec{x}_k - \vec{y}_k). &(28)\end{aligned}$$

Moreover, we may express $\vec{w}_k$ given by (22) through

$$\begin{aligned}\vec{w}_k &= (\Phi_k + \Phi_k^{-T}(\Psi_k^T\Gamma_k + M^TPA^N)G(\mathtt{z})[1\ 0\cdots 0])\vec{x}_k\\ &\quad - \Phi_k^{-T}(\Psi_k^T\Gamma_k + M^TPA^N)G(\mathtt{z})[1\ 0\cdots 0]\vec{y}_k\\ &= \Phi_k^{-T}\big(F_k(\mathtt{z})\vec{x}_k - (F_k(\mathtt{z}) - \Phi_k^T\Phi_k)\vec{y}_k\big) &(29)\end{aligned}$$

Recalling that $y_k = [1\ 0\cdots 0]\vec{y}_k$, $\vec{y}_k = \Phi^{-1}\vec{\xi}_k$ and using (29), leads to the noise-shaping nearest-neighbor quantization architecture shown in Fig. 2.

The analysis provided in the previous sections showed that the MH quantizer can be implemented as a nearest neighbor (Euclidean) vector quantizer. It is important to see that the mapping Φ_k^{-1} is applied upon the quantized vector $\vec{\xi}_k$ to obtain $\vec{y}_k$ whereafter the first element y_k of $\vec{y}_k$ is transmitted to the decoder. In general, it may not be possible to apply an abitrary transform Φ_k^{-1} on a quantized vector $\vec{\xi}_k$ and yet be within a desired quantization space, e.g., within $\mathcal{Y}_k$.

If the nearest neighbor quantizer $Q_{\Phi_k\mathcal{Y}_k}(\cdot)$ is obtained as the transformation $\Phi_k\mathcal{Y}_k$ of the codebook $\mathcal{Y}_k$ of the MH quantizer $Q_k^N(\cdot)$, then clearly $\Phi_k^{-1}\vec{\xi}_k \in \mathcal{Y}_k$. In this case, if $\mathcal{Y}_k$ defines a lattice codebook, then $\Phi_k\mathcal{Y}_k$ will be a *shaped* lattice w.r.t. $\mathcal{Y}_k$, where Φ_k is called the shaping operator [12]. Thus, $\Phi_k\mathcal{Y}_k$ will also be a lattice.

If the final state weighting matrix P is taken to be the all-zero matrix, then the cost function simplifies to the one given in (1). In this case, $\Phi_k = \Psi_k$ and it follows from (13) that Φ_k will then be lower unitriangular. But since the inverse of finite-dimensional and lower unitriangular is also lower unitriangular, it follows that the first row of Ψ_k^{-1} will be the unit vector $[1\ 0\ \cdots\ 0]$. This implies that $y_k = \xi_k$, i.e., the first element of the quantized vector $\vec{\xi}_k$ will be equal to the first element of $\vec{y}_k$.

In the general case where P is not the all-zero matrix and the codebook for the nearest neighbor quantizer $\mathcal{Q}_{\Phi_k\mathcal{Y}_k}$ is arbitrarily designed (e.g., the codebook could be a fixed lattice) the resulting ouput variable $y_k = [1\ 0\ \cdots\ 0]\Phi_k^{-1}\vec{\xi}_k$ generally lies in a time-varying domain, since Φ_k is time-varying. In this case, care must be taken, since Φ_k is not known at the decoder and hence the resulting codebook is not known at the decoder. One possible approach is to make Φ_k^{-1} all integers (in which case the transpose Φ_k^{-T} is also all integers). If now the codebook $\mathcal{Y}_k$ is chosen to consist of all integer coordinates, as is the case if e.g., appropriately scaled scalar quantizers are used, then $y_k \in \mathbb{Z}$ as desired. This approach where the quantization operation is applied before the transformation has been studied in e.g., the Wavelet literature where it is known as *lifting* [13] and in the source coding literature where it is commonly referred to as *integer-to-integer* transformations [14] or *reversible integer* mappings [15].

III. Rate-Distortion Analysis of Perceptual Moving Horizon Quantization

In this section we perform a rate-distortion analysis of the perceptual MH quantizer. We use bold faced symbols for stochastic variables, e.g., $\vec{\boldsymbol{\xi}}_k$ denotes the kth vector of the vector process $\vec{\boldsymbol{\xi}}$ and $\vec{\xi}_k$ denotes a realization. For sequences e.g., $\{\vec{\boldsymbol{\xi}}_i, \vec{\boldsymbol{\xi}}_{i+1}, \ldots, \vec{\boldsymbol{\xi}}_k\}$ we use the notation $\{\vec{\boldsymbol{\xi}}_j\}_{j=i}^k$. Also recall the slight abuse of notation that ξ_k denotes the first sample of the kth vector $\vec{\xi}_k$ so that ξ_{k+1} is the first sample of the vector $\vec{\xi}_{k+1}$. We will make use of the following information theoretic quantities, $\mathbb{E}[\cdot], H(\cdot), h(\cdot), I(\cdot;\cdot)$, and $D(\cdot;\cdot)$, which denote statistical expectation, discrete entropy, continuous entropy, mutual information, and Divergence (or Kullback Leibler distance), respectively, see [16] for details.

A. Without Final State Weighting

We first consider the case of MH quantization without final state weighting (i.e., where $P = 0I$). In this case, it follows from the previous section that $\Psi_k = \Phi_k$ is lower triangular and hence $y_k = \xi_k$. To make the proposed system amenable to a rigorous analysis, we will be using entropy coded and substractively dithered lattice quantizers, which are hereafter abbreviated ECDQs and denoted by the symbol $\mathcal{Q}_\Lambda$, where $\Lambda \in \mathbb{R}^N$ refers to the underlying N-dimensional lattice. An ECDQ uses a dither signal $\vec{\nu}_k \in \mathbb{R}^N$ which is i.i.d. and uniformly distributed over a Voronoi cell of the lattice Λ [17]. Given an input $\vec{w}_k$ the output of the ECDQ is given by

$$\vec{\xi}_k = \mathcal{Q}_\Lambda(\vec{w}_k + \vec{\nu}_k). \tag{30}$$

Thus, $\vec{\xi}_k$ belongs to a discrete alphabet (i.e., Λ). The first sample of $\vec{\xi}_k$, i.e., ξ_k, is further entropy coded in order to be represented by a sequence of bits $\bar{b}_k = \mathcal{E}(\xi_k, \nu_k)$, where $\mathcal{E}(\cdot,\cdot)$ denotes the entropy coder.[3] The reconstruction $\vec{\xi}_k'$ at the *encoder* is obtained by subtracting the dither signal, i.e.

[3]We emphasize that the entropy coding is conditioned upon the dither signal ν_k [17].

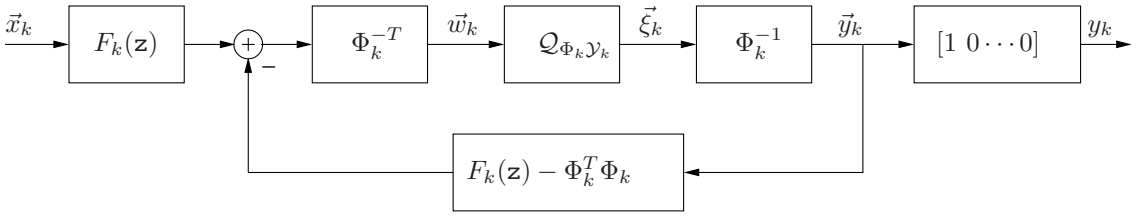


Fig. 2. MH quantization implemented using a noise-shaping architecture.

$\vec{\xi}'_k = \vec{\xi}_k - \vec{\nu}_k$, and $\vec{y}_k$ is then given by $\vec{y}_k = \Psi_k^{-1}\vec{\xi}'_k$. Notice that due to dithering, $\vec{\xi}'_k$ and $\vec{y}_k$ belong to continuous alphabets. It was shown in [17], that the quantization error $\vec{q}_k$, where

$$\vec{q}_k = \vec{\xi}'_k - \vec{w}_k = \vec{\xi}_k - \vec{w}_k - \vec{\nu}_k, \tag{31}$$

is i.i.d., independent of $\vec{w}_k$, and distributed as $-\vec{\nu}_k$. The reconstruction ξ'_k at the *decoder* follows by first obtaining $\xi_k = \mathcal{D}(\bar{b}_k, \nu_k)$, where $\mathcal{D}(\cdot,\cdot)$ denotes entropy decoding, and then subtracting the dither ν_k, that is

$$\xi'_k = \mathcal{D}(\bar{b}_k, \nu_k) - \nu_k \tag{32}$$
$$= \xi_k - \nu_k \tag{33}$$
$$= w_k + q_k, \tag{34}$$

where it is assumed that the first sample ν_k of the dither vector $\vec{\nu_k}$ is known at the decoder.[4] To accomodate dithering, we redraw the schematics of Fig. 2 into the form shown in Fig. 3.

Let $\bar{R}$ denote the average conditional entropy of the ECDQ, when the quantized variables $\{\xi_k\}$ are independent (sample-by-sample) entropy coded, i.e.,

$$\bar{R} \triangleq \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} H(\boldsymbol{\xi}_i|\boldsymbol{\nu}_i). \tag{35}$$

It is known that this conditional entropy provides a lower bound on the per sample average (operational) coding rate $\bar{R}^*$ of the ECDQ. Moreover, the conditional entropy $H(\boldsymbol{\xi}_k|\boldsymbol{\nu}_k)$ is equal to the mutual information over the additive noise channel $\boldsymbol{\xi}'_k = \boldsymbol{w}_k + \boldsymbol{q}_k$ [17]. Thus, the operational coding rate is lower bounded by

$$\bar{R}^* \geq \bar{R} = \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} I(\boldsymbol{w}_k; \boldsymbol{\xi}'_k). \tag{36}$$

We are interested in designing the quantizer codebook Λ so as to minimize the time-averaged expected perceptual distortion $\bar{D}$ (per dimension), for a fixed horizon length N, subject to a target entropy constraint $\bar{R} \leq R_T$ on the average conditional entropy $\bar{R}$ when independently encoding the sequence of "first coordinates" of the quantized outputs, i.e., $\{\boldsymbol{\xi}_j\}_{j=0}^{\infty}$.[5] Specifically, by use of (1), we can express $\bar{D}$ as

$$\bar{D} = \lim_{k\to\infty} \frac{1}{kN}\sum_{i=0}^{k-1} \mathbb{E}\|\vec{\boldsymbol{\epsilon}}_i\|^2 \tag{37}$$

Lemma 1: The average perceptual distortion $\bar{D}$ of Fig. 3 is given by

$$\bar{D} = \lim_{k\to\infty} \frac{1}{kN}\sum_{i=0}^{k-1} \mathbb{E}\|\vec{\boldsymbol{\nu}}_i\|^2. \tag{38}$$

Proof: Whether or not we use dithering, the output $\vec{\xi}_k$ of the quantizer can always be written as $\vec{\xi}_k = \vec{w}_k + \vec{q}_k$, where $\vec{q}_k$ at this point can be arbitrarily distributed. The cost metric (1) can be rewritten as

$$\|\vec{\epsilon}_k\|^2 = f_k^{\vec{w}}(\vec{\xi}_k) + \Xi_k(\vec{x}_k, \vec{z}_k) \tag{39}$$
$$= \|\vec{\xi}_k\|^2 - 2\vec{\xi}_k^T\vec{w}_k + \Xi_k(\vec{x}_k, \vec{z}_k) \tag{40}$$
$$= \|\vec{w}_k + \vec{q}_k\|^2 - 2(\vec{w}_k + \vec{q}_k)^T\vec{w}_k + \Xi_k(\vec{x}_k, \vec{z}_k) \tag{41}$$
$$= \|\vec{q}_k\|^2, \tag{42}$$

where the last equality follows since $P = 0I$ so that $\vec{w}_k = \Psi_k\vec{x}_k + \Gamma_k\vec{z}_k$ and $\Xi_k(\vec{x}_k, \vec{z}_k) = \|\vec{x}_k\|^2_{\Psi_k^T\Psi_k} + 2\langle\vec{x}_k, \Psi_k^T\Gamma_k\vec{z}_k\rangle + \|\vec{z}_k\|^2_{\Gamma_k^T\Gamma_k}$. When the quantizer is an ECDQ, we note that the reconstruction is $\vec{\xi}'_k = \vec{\xi}_k - \vec{\nu}_k = \vec{w}_k + \vec{q}_k$ and the perceptual distortion satisfies $\|\vec{\epsilon}_k\|^2 = f_k^{\vec{w}}(\vec{\xi}'_k) + \Xi_k(\vec{x}_k, \vec{z}_k)$. Inserting this into (39) and using that $\|\vec{q}_k\|^2 = \|\vec{\nu}_k\|^2$ yields (38). ■

Corollary 1: If the lattice Λ in Fig. 3 is fixed, i.e., $\vec{\boldsymbol{\nu}}$ is i.i.d., and $\|\vec{\boldsymbol{\nu}}_k\|^2 = \sigma^2, \forall k$, then

$$\bar{D} = \frac{1}{N}\sigma^2. \tag{43}$$

Proof: Follows immediately from Lemma 1 by using the fact that $\vec{\boldsymbol{\nu}}$ is zero-mean and identically distributed for all k if Λ is fixed. ■

Theorem 1: Let $\boldsymbol{x}$ be stationary, having finite differential entropy rate, but otherwise arbitrarily distributed. Then the coding rate for the scheme in Fig. 3 is bounded between:

$$\bar{R} \leq \bar{R}^* < \bar{R} + 1.2547, \tag{44}$$

where $\bar{R} = \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{\xi}'_i) - h(\boldsymbol{\nu}_i)$.

Proof: We first prove the lower bound. As explained in Section II-E, for $P = 0I$, the first element of $\vec{y}_k$ is identical to the first element of $\vec{\xi}_k$. The marginal distribution $p_{\boldsymbol{\xi}_k}$ of $\boldsymbol{\xi}_k$ is therefore identical to the marginal distribution $p_{\boldsymbol{y}_k}$ of $\boldsymbol{y}_k$. We may therefore proceed by considering $\boldsymbol{\xi}_k$ instead of $\boldsymbol{y}_k$.

[4]This kind of *common* randomness can be obtained by e.g., guaranteeing that the encoder and decoder are synchronized w.r.t., to their random generators, or e.g., by transmitting (or agreeing upon) a common seed.

[5]We are interested in the situation where the elements of the sequence of output samples $\{\boldsymbol{\xi}_j\}_{j=0}^{\infty}$ are encoded separately for two reasons. First, it leads to a simple low delay design. Second, it guarantees that the encoder and decoder remain "synchronized" also in the case of packet dropouts.

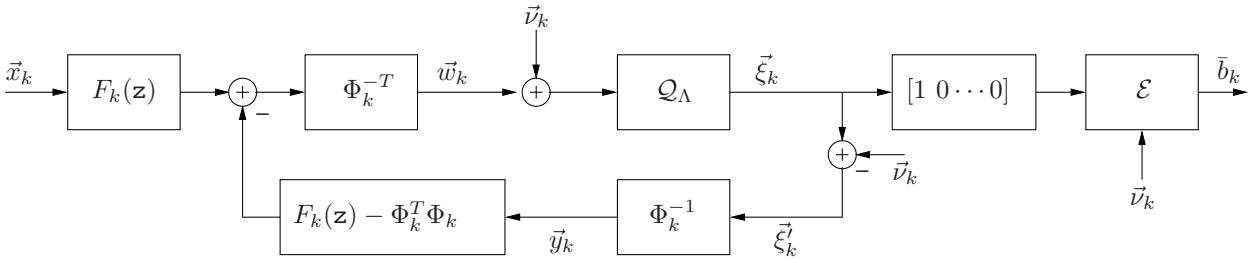


Fig. 3. MH quantization (with subtractive dithering) implemented using a noise-shaping architecture.

The operational coding rate is lower bounded by the average scalar mutual information between $\boldsymbol{w}_k$ and $\boldsymbol{\xi}'_k$. Specifically,

$$\bar{R} \geq \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} I(\boldsymbol{w}_i;\boldsymbol{\xi}'_i) \tag{45}$$

$$= \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} \big(h(\boldsymbol{\xi}'_i) - h(\boldsymbol{\xi}'_i|\boldsymbol{w}_i)\big) \tag{46}$$

$$= \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} \big(h(\boldsymbol{\xi}'_i) - h(\boldsymbol{q}_i)\big) \tag{47}$$

$$= \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} \big(h(\boldsymbol{\xi}'_i) - h(\boldsymbol{\nu}_i)\big)\,. \tag{48}$$

We will now prove the upper bound by using [18, Lemma 2] in order to show that $I(\boldsymbol{w}_i;\boldsymbol{\xi}'_i)$ can be upper bounded by replacing the variables $\boldsymbol{w}_i$ and $\boldsymbol{q}_i$ by Gaussian variables $\boldsymbol{w}_i^g$ and $\boldsymbol{q}_i^g$, having the same second-order statistics. With this, we have that

$$I(\boldsymbol{w}_i;\boldsymbol{\xi}'_i) = I(\boldsymbol{w}_i;\boldsymbol{w}_i+\boldsymbol{q}_i) \tag{49}$$

$$\leq I(\boldsymbol{w}_i^g;\boldsymbol{w}_i^g+\boldsymbol{q}_i^g) + D(\boldsymbol{q}_k||\boldsymbol{q}_k^g). \tag{50}$$

It follows that the Divergence term $D(\boldsymbol{q}_k||\boldsymbol{q}_k^g)$, only depends upon the marginal distribution of the first sample of the quantization noise vector. Since we are using lattice vector quantizers, where the quantization noise (due to dithering) is uniformly distributed over a Voronoi cell, the resulting marginal distribution depends only upon the shape of the Voronoi cells. In general, the more spherically shaped Voronoi cells, the more "Gaussian"-like quantization noise [12]. The worst lattice vector quantizer is obtained by using a sequence of scalar uniform quantizers individually along each dimension of the source vectors. In this case, $D(\boldsymbol{q}_k||\boldsymbol{q}_k^g) < 0.2547$ [12]. The proof is completed by using the well known fact, that there exists entropy coders with an average rate, which is strictly less than the output entropy plus 1 bit/dimension [19]. ■

Remark 1: Theorem 1 provides a sandwich on the operational coding rate. The upper bound is due to using non-asymptotic quantizers and non-asymptotic entropy coders. As is well known, the 1 bit/sample "loss" of the entropy coder tends to zero at high coding rates or at high vector dimensions [19]. The remaining gap, i.e., the 0.2547 bits/sample, is the loss due to not using optimal vector quantizers. Thus, in the limit where $N \to \infty$ and if optimal vector quantization is used, it can be shown that also this gap vanishes. Thus, at high rates and large vector quantizer dimension, the lower bound is achievable. In the simulations section, we show that even without dithering, simple scalar quantization gets very close to the lower bound.

Remark 2: If the quantizer codebook is designed in the original domain, i.e., if $\mathcal{Y}_k$ is designed and then $\Psi_k\mathcal{Y}_k$ is used, then the performance is inferior as to when the codebook is designed in the transform domain. However, in this case, using larger horizon lengths can be expected to give additional gains over that of the space-filling gain of the vector quantizers. Such situations were examined in [8], [11].

We have so far considered the case where the encoder separately encodes the sequence of quantized variables. It would be interesting to compare this to the gain by allowing the encoder and decoder to exploit all the memory within the system, when encoding the first sample of the vector outputs. In this case, we have the following lower bound on the average entropy $\bar{R}$ of the process $\{\boldsymbol{\xi}'_j\}_{j=0}^{\infty}$:

Lemma 2: Let $\boldsymbol{x}$ have finite differential entropy rate but otherwise arbitrarily distributed. Moreover, fix the lattice in the ECDQ such that $\boldsymbol{\nu}$ is i.i.d., and independent of x. If all memory within the system is exploited, then the average entropy is lower bound by

$$\bar{R} \geq \bar{h}(\{\boldsymbol{\xi}'_j\}_{j=0}^{\infty}) - h(\boldsymbol{\nu}), \tag{51}$$

where $\bar{h}(\cdot)$ denotes the differential entropy rate [16].

Proof: Since we have a source coding system within a feedback loop, there is memory in the system and it follows from [20] (see also Theorem 1 in [21]), that the average entropy is lower bounded by Massey's notion of directed mutual information [22]. Thus,

$$\bar{R} \geq I(\{\boldsymbol{w}_j\}_{j=0}^{\infty} \to \{\boldsymbol{y}_j\}_{j=0}^{\infty}) \tag{52}$$

$$\triangleq \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} I(\boldsymbol{y}_i;\{\boldsymbol{w}_j\}_{j=0}^{i}|\{\boldsymbol{y}_j\}_{j=0}^{i-1}). \tag{53}$$

We now use that $\vec{\xi}'_k = \vec{w}_k + \vec{q}_k$ and recall that $y_k = \xi'_k$ since $P = 0I$. This allows us to further lower bound the rate as follows:

$$\bar{R} \geq I(\{\boldsymbol{w}_j\}_{j=0}^{\infty} \to \{\boldsymbol{\xi}'_j\}_{j=0}^{\infty}) \tag{54}$$

$$\triangleq \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} I(\boldsymbol{\xi}'_i;\{\boldsymbol{w}_j\}_{j=0}^{i}|\{\boldsymbol{\xi}'_j\}_{j=0}^{i-1}) \tag{55}$$

$$= \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{\xi}_i'|\{\boldsymbol{\xi}_j'\}_{j=0}^{i-1}) - h(\boldsymbol{\xi}_i'|\{\boldsymbol{\xi}_j'\}_{j=0}^{i-1},\{\boldsymbol{w}_j\}_{j=0}^{i}) \tag{56}$$

$$= \bar{h}(\{\boldsymbol{\xi}_j'\}_{j=0}^{\infty}) - \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{w}_i+\boldsymbol{q}_i|\{\boldsymbol{\xi}_j'\}_{j=0}^{i-1},\{\boldsymbol{w}_j\}_{j=0}^{i}) \tag{57}$$

$$= \bar{h}(\{\boldsymbol{\xi}_j'\}_{j=0}^{\infty}) - \lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{q}_i|\{\boldsymbol{\xi}_j'\}_{j=0}^{i-1},\{\boldsymbol{w}_j\}_{j=0}^{i}) \tag{58}$$

$$= \bar{h}(\{\boldsymbol{\xi}_j'\}_{j=0}^{\infty}) - h(\boldsymbol{\nu}), \tag{59}$$

where the last equality follows since q_k is independent of past and current input and quantization error samples due to the use of independent dithering. Moreover, q_k is distributed as $-\nu_k$ but negation does not affect the differential entropy. ■

Remark 3: As expected, the reduction in the lower bound on the average entropy when memory is utilized is solely given by the difference between the average differential entropy and the differential entropy rate of $\{\boldsymbol{\xi}_k'\}$.

B. With Final State Weighting

We will now examine MH quantization with non-zero final state weighting, i.e., where $P \neq 0I$. We will assume that the codebook $\mathcal{Y}_k$ is given and a nearest neighbor vector quantizer is using the transformed codebook $\Phi_k\mathcal{Y}_k$. The average expected distortion is now based on (11), that is

$$\bar{D} = \lim_{k\to\infty} \frac{1}{kN}\sum_{i=0}^{k-1} \mathbb{E}\left\{\|\vec{\boldsymbol{\epsilon}}_i\|^2 + \|\vec{\boldsymbol{z}}_{i+N}\|_P^2\right\}. \tag{60}$$

Lemma 3: Let $\vec{\boldsymbol{\xi}}_k = \vec{\boldsymbol{w}}_k + \vec{\boldsymbol{q}}_k$, where $\vec{\boldsymbol{q}}_k \in \mathbb{R}^N$ is an arbitrarily distributed quantization error vector. Then,

$$\bar{D} = \lim_{k\to\infty} \frac{1}{kN}\sum_{i=0}^{k-1} \mathbb{E}\left\{\|\vec{\boldsymbol{q}}_i\|^2 + \|\vec{\boldsymbol{z}}_i\|_{\boldsymbol{\varphi}_i^T\boldsymbol{\varphi}_i}\right\}. \tag{61}$$

where

$$\boldsymbol{\varphi}_i^T\boldsymbol{\varphi}_i \triangleq \boldsymbol{\Gamma}_i^T\boldsymbol{\Gamma}_i + (A^N)^T P A^N - (\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i + M^T P A^N)^T \boldsymbol{\Phi}_i^{-1}\boldsymbol{\Phi}_i^{-T}(\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i + M^T P A^N).$$

Proof: The proof follows along the lines of the proof of Lemma 1. ■

Unfortunately, we have not been able to obtain non-trivial rate bounds for this harder situation where $P \neq 0I$. The main reason is that, even if one fixes the lattice $\mathcal{Y}_k$ in the original $\vec{\boldsymbol{y}}_k$-domain, the resulting lattice after the transformation $\boldsymbol{\Phi}_k\mathcal{Y}_k$ is random since $\boldsymbol{\Phi}_k$ is a random matrix. Moreover, after quantization the quantized vector $\vec{\boldsymbol{\xi}}_k$ is mapped to the original $\vec{\boldsymbol{y}}_k$-domain where the first sample, i.e., $\boldsymbol{y}_k$, is to be transmitted. Thus, the inverse mapping $\boldsymbol{\Phi}_k^{-1}$ affects the marginal distribution of $\boldsymbol{y}_k$ so that it is not identical to the marginal distribution of $\boldsymbol{\xi}_k$.

IV. Multiple-Description Perceptual Moving Horizon Quantization

With the results of Sections II and III as background, in this section we will present our main proposal, namely, the use of MH quantization together with MD coding.

A. Multiple-Description Moving Horizon Quantization

In MD coding a single source vector $\vec{x}_k$ is mapped to multiple output vectors $(\vec{y}_k^0, \vec{y}_k^1, \ldots, \vec{y}_k^{n-1})$, which are usually referred to as descriptions [2]. In the general case, we have $n \geq 1$ descriptions, see e.g., [23]. Hence, we have n encoders

$$f_j : \vec{x}_k \mapsto \vec{y}_k^j \in \mathbb{R}^N, j = 0,\ldots,n-1, \tag{62}$$

and 2^n decoders

$$g_\ell : \{\vec{y}_k^j : j \in \ell\} \mapsto \hat{\vec{y}}_k^\ell \in \mathbb{R}^N, \ell \subseteq \{0,\ldots,n-1\}. \tag{63}$$

For every time instance k, the first sample of each of the n current descriptions, i.e. $\{y_k^0, y_k^1, \ldots, y_k^{n-1}\}$, are transmitted over n channels so that description j, i.e. y_k^j, is transmitted on the jth channel. At any time k, an arbitrary subset of the channels may break down. Which of the channels are currently working is not known to the encoder, but it is known to the decoder.[6] The problem is then to construct the n descriptions, so that they provide a certain degree of redundancy, which can be exploited at the decoder during channel failures. Generally, the descriptions are able to refine each other, and the distortion achieved therefore depends upon which subset of descriptions was received.

To successfully combine MD coding and MH quantization, one needs to carefully consider several issues. Firstly, an MD encoder outputs multiple descriptions, whereas the MH quantizer $\mathcal{Q}_k^N(\cdot)$ studied in Sections II and III gives only a single output. Furthermore, there is a feedback loop at the encoder, since past decisions affect the current decision through the system state vector $\vec{z}_k$, see e.g., (10) and Fig. 2. In order for this feedback loop to to be well defined at the encoder, we need to form a single output based on the n descriptions. Towards that end, for some fixed set of scalar weights $\{\gamma_\ell \in \mathbb{R}\}_{\ell\subset\{0,\ldots,n-1\}}$, we define[7]

$$\tilde{\vec{y}}_k \triangleq \sum_{\ell\subseteq\{0,\ldots,n-1\}} \gamma_\ell \hat{\vec{y}}_k^\ell \tag{64}$$

and update the state vector $\vec{z}_k$ (which was previously given by (10)) by the following rule:

$$\vec{z}_k = [x_{k-1} - \tilde{y}_{k-1}, x_{k-2} - \tilde{y}_{k-2}, \ldots, x_{k-K} - \tilde{y}_{k-K}]^T, \tag{65}$$

where $\tilde{y}_k$ denotes the first sample of the vector $\tilde{\vec{y}}_k$ given in (64). The weights $\{\gamma_\ell\}$ in (64) may, for example, reflect successful decoding probabilities, i.e. the probability of receiving only the descriptions, which are indexed by ℓ.

Another issue which should be taken into account when designing an MD coder for MH quantization is that the cost function J_k^N introduced in (1) and extended to include state cost in (17) does not explicitly take into account the distortion as observed by the decoder. While this might be somewhat curious from a coding point of view, it is the de facto standard in moving horizon optimization methods (see e.g., [11]) and model predictive control [10]. However, motivated by (64),

[6] It is assumed that the decoder can deduce which channels are working e.g. based on the set of received descriptions.

[7] We note that how to form the vector to be fed back at encoder is a non-trivial problem. This is partly due to the fact that the encoder does not know in advance which descriptions will be received at the decoder.

we propose to rewrite (1) as a weighted sum over the possible outcomes due to packet dropouts. Specifically, in the case of n descriptions, we propose the following cost function:

$$J_k^N(\vec{x}_k) \triangleq \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell \|\vec{\epsilon}_k^{\,\ell}\|^2, \tag{66}$$

where, for $\ell \subseteq \{0,\ldots,n-1\}$, the perceptually filtered error sample is given by

$$\epsilon_k^\ell \triangleq C_k \vec{z}_k + (x_k - y_k^\ell). \tag{67}$$

Lemma 4: Let the cost function be given by (66). Moreover, let the weights $0 \leq \gamma_\ell \in \mathbb{R}, \ell \subseteq \{0,\ldots,n-1\}$ be given. Then, the optimal set of reconstruction vectors $\{\vec{y}_k^\ell \in \mathcal{Y}_k^\ell\}_{\ell \subseteq \{0,\ldots,n-1\}}$, where $\mathcal{Y}_k^\ell$ denotes the codebook for $\vec{y}_k^\ell$, can be found as

$$\arg \min_{\{\vec{\xi}_k^\ell \in \Psi_k \mathcal{Y}_k^\ell\}_{\ell \subseteq \{0,\ldots,n-1\}}} \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell f_k^{\vec{w}}(\vec{\xi}_k^\ell), \tag{68}$$

where $f_k^{\vec{w}}(\vec{\xi}_k^\ell)$ is given by (21), by using the relationship $\vec{y}_k^\ell = \Psi_k^{-1}\vec{\xi}_k^\ell, \ell \subseteq \{0,\ldots,n-1\}$.

Proof: By adopting a similar approach as when forming (12), it is easy to show that $\|\vec{\epsilon}_k^{\,\ell}\|^2 = \|\Psi_k(\vec{x}_k - \vec{y}_k^\ell) + \Gamma_k \vec{z}_k\|^2$, where $\vec{z}_k$ is given by (65). Moreover, using (18) – (22) it follows that

$$\|\Psi_k(\vec{x}_k - \vec{y}_k^\ell) + \Gamma_k \vec{z}_k\|^2 = f_k^{\vec{w}}(\vec{\xi}_k^\ell) + \Xi_k(\vec{x}_k, \vec{z}_k), \tag{69}$$

where $\Xi_k(\vec{x}_k, \vec{z}_k)$ is independent of $\vec{y}_k^\ell$ (at time k). We therefore establish that $J_k^N(\vec{x}_k)$ given by (66) can be rewritten as

$$J_k^N(\vec{x}_k) = \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell \left(f_k^{\vec{w}}(\vec{\xi}_k^\ell) + \Xi_k(\vec{x}_k, \vec{z}_k) \right). \tag{70}$$

The lemma is now proved by recognizing that minimizing (70) is equivalent to solving (68). ∎

Remark 4: Lemma 4 shows that minimizing the perceptually weighted cost function (66) is equivalent to solving the weighted MSE minimization problem (68), i.e., solving $\arg\min \sum_\ell \gamma_\ell \|\vec{w}_k - \vec{\xi}_k^\ell\|^2$. Since this defines a (weighted-Euclidean) nearest-neighbor MD vector quantization problem, we may use conventional MD quantization techniques. In this work, we will apply the n-description index-assignment based lattice vector quantization construction of [23], [24].

B. Rate-Distortion Analysis of Perceptual MD MH quantization

The optimum rate-distortion performances of MD problems are generally not known. In fact, it is only completely solved for two descriptions in the case of MSE distortions and white Gaussian sources [2], [25] or colored Gaussian sources [26], [27]. In the case of more than two descriptions, even less is known.

In this work, we let the MD quantizer be the simple index-assignment based lattice vector quantizer presented in [23]. With this MD quantizer, the reconstruction rule is given as the average of the received descriptions, or in the case all descriptions are received, it is given by the inverse of a fixed mapping function. The MD quantizer consists of a single high-quality quantizer, referred to as a *central* quantizer, and n coarser quantizers referred to as *side* quantizers. The central quantizer has Voronoi cells of volume ν_c and the side quantizers have Voronoi cells of volume $\nu = \rho^N \nu_c$, where $\rho > 1$ denotes the nesting factor, which is inversely proportional to the amount of redundancy within the system. Thus, a large nesting factor yields poor side performance but very good central performance, whereas a small nesting factor yields good side performance and only slightly better central performance, see [23] for details. Under high-resolution assumptions, the coding rate per description is given by [23]

$$\bar{R}^* \dot{\approx} \frac{1}{k} \sum_{i=0}^{k-1} h(\boldsymbol{w}_i) - \frac{1}{N} \log_2(\nu), \tag{71}$$

where $\dot{\approx}$ means that the approximation is exact in the limit where the rate diverges to infinity and the distortion tends to zero. In the case of an n-description system, the average distortion $\bar{D}_{0,\ldots,n-1}$ when receiving all n descriptions is given by [23]

$$\bar{D}_{0,\ldots,n-1} \dot{\approx} G(\Lambda) \nu_c^{2/N}, \tag{72}$$

where $G(\Lambda)$ denotes the dimensionless normalized second-moment of inertia [28] of the N-dimensional lattice quantizer Λ being used. On the other hand, since we are here referring to a *symmetric*[8] setup, the distortion $\bar{D}_\ell$ where $\ell \subseteq \{0,\ldots,n-1\}$ and $|\ell| = \kappa$, when receiving any $1 < \kappa < n$ descriptions, is given by [23]

$$\bar{D}_\ell \dot{\approx} \left(\frac{n-\kappa}{2n\kappa} \right) G(S_N) \mathfrak{B}_{n,N}^2 2^{2(h(\boldsymbol{\xi}) - \bar{R}_c)^*} 2^{\frac{2n}{n-1}(\bar{R}_c^* - \bar{R}^*)}, \tag{73}$$

where $\bar{R}^*$ is given by (71), $\bar{R}_c^* = \frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{w}_i) - \frac{1}{N}\log_2(\nu_c)$, $G(S_N)$ denotes the dimensionsless normalized second-moment of inertia of an N-dimensional hypersphere [28], and $\mathfrak{B}_{n,N}$ is an expansion factor. The latter is a function of the number of descriptions n and the vector dimension N, see [23] for details. With this, the average perceptual cost function (66) can be written as

$$J_k^N(\vec{x}_k) \triangleq \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell \|\vec{\epsilon}_k^{\,\ell}\|^2 \dot{\approx} \sum_{\ell \subseteq \{0,\ldots,n-1\}} \gamma_\ell \bar{D}_\ell, \tag{74}$$

where $\bar{D}_\ell$ is given by (72) and (73).

V. Design Study

In this section, we design and simulate the proposed coding architecture. We first show how one may obtain the perceptual weighting filter. We then motivate the use of MH quantization by considering a single-description setup, and show that by using a simple fixed perceptual weighting filter, significant gains over linear PCM can be achieved. We finally construct an MD MH quantization based scheme, and consider a scenario with three descriptions.

[8] *Symmetric* MD coding refers to the case where: 1) all side descriptions are encoded at the same descriptions rate. 2) the distortion observed at the decoder depends only upon the number of received descriptions and as such not upon which descriptions that are received.

A. Obtaining the Perceptual Weighting Filter

Most psychoacoustic models are defined in the frequency domain and are based on a block of M time-domain samples. We therefore need to introduce a certain amount of delay in order to achieve sufficient accuracy of the frequency response. The specific choice of psychoacoustic model is not essential for our design. We could, for example, choose the model from the MPEG1 layer 1 standard [7], which is based on a block of $M = 512$ samples, at a sample rate of 44.1 kHz or one could use the model presented in [29], which is based on $M = 128$ time-domain samples. Alternatively, one could simply use a fixed perceptual weighting filter in which case there is no need for a delay.

In order to obtain the perceptual filter $\vec{h}_k$ of order K, we use an idea suggested by Schuller et al. [30]. Let $|\theta_k(f)|^2$ be the masked threshold as computed by the perceptual model, for the kth block, and notice that we would like to find a perceptual weighting filter with a transfer function that satifies $|H_k(f)|^2 \approx |\theta_k(f)|^{-2}$. If we use $|\theta_k(f)|^2$ as a short-term power spectrum, then the symmetric autocorrelation sequence $\{r_{k,i}\}, i = 0, \ldots, \frac{M}{2}$, is found simply as the inverse DFT of $|\theta_k(f)|^2$. The filter coefficients $h_{k,1}, \ldots, h_{k,K}$ are now easily found from $\{r_{k,i}\}$ by use of the Yule-Walker equations [31].

In the simulations that follow in the next sequel, we will use a simple fixed third-order perceptual weighting filter. In particular, we use a filter which mimics the threshold in quiet [32]. Let f denote frequency (in Hz), then the threshold in quiet $\mathcal{T}_q(f)$ can be approximated by the following expression [32], [33]:

$$\mathcal{T}_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5\exp\left(-0.6\left(\frac{f}{1000} - 3.3\right)^2\right) + 10^{-3}\left(\frac{f}{1000}\right)^4. \tag{75}$$

Using the technique described above, we obtain $\bar{h}_k$ from (75), i.e., for $K = 3$, we get

$$\bar{H}_k(z) = 1 + 0.4367\mathtt{z}^{-1} - 0.6407\mathtt{z}^{-2} - 0.5839\mathtt{z}^{-3}, \quad \forall k. \tag{76}$$

Recall that $\bar{h}_k$ is used in a *noise-shaping* process, and that this operation does not introduce a delay.

B. Real-Time Single-Description Perceptual MH quantization

To avoid delay, we will in this first simulation use simple uniform scalar quantization. Thus, $N = 1$ and the current sample is encoded and decoded independently of future samples. However, the current sample is encoded by taking into account previous samples and coded values, as summarized by the current state vector. With $N = 1$, and using only a single description, the proposed scheme is akin to noise-shaping coders.

As a baseline, we first directly quantize the music signal, using a uniform scalar quantizer, which corresponds to conventional linear PCM. Under high-resolution assumptions, the resulting discrete entropy $H(\hat{\boldsymbol{y}})$ of the quantized signal can be approximated by [34]:

$$H(\hat{\boldsymbol{y}}) \approx h(\boldsymbol{x}) - \log_2(\Delta), \tag{77}$$

where Δ denotes the step-size of the quantizer. Moreover, it is well known that the (MSE) distortion D is approximately $D \approx \Delta^2/12$. It follows from (77), that for a given coding rate $\bar{R}^* \approx H(\hat{\boldsymbol{y}})$, the step-size of the quantizer is given by $\Delta = 2^{h(\boldsymbol{x})-\bar{R}^*}$, and knowledge of the differential entropy of the source signal is required in order to obtain the optimal scaling Δ of the quantizer. If the signals are Gaussian distributed, $h(\boldsymbol{x}) = \frac{1}{2}\log_2(2\pi e\sigma^2)$, where σ^2 denotes the variance of the signal. Thus, in this case, only knowledge of the source variance is required.

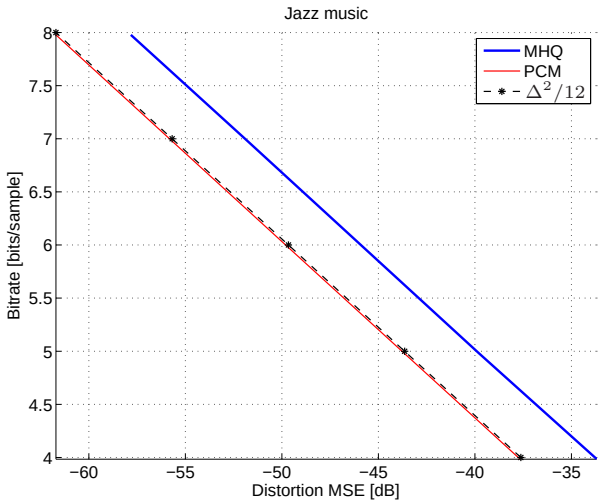


Fig. 4. Operational rate-distortion curves.

We use three different fragments of music; *Jazz*, *Pop*, and *Rock*, all having a sampling rate of 48 kHz and a duration of 15.0, 6.8, and 13.5 seconds, respectively. We measure their variances, and use the Gaussian approximation given above, in order to derive the scaling factor Δ. As can be observed from Fig. 4, the approximation is quite accurate, i.e., the operational rate-distortion function of the linear PCM encoded signal approximately coincides with that obtained from a truly Gaussian signal.[9] In Fig. 4, the x-axis describes the MSE distortion in dB, and the y-axis describes the discrete entropy in bits per sample.[10]

Also shown in Fig. 4, is the operational MSE rate-distortion performance obtained with the proposed perceptual MH quantizer, in the simplest case where $N = 1, P = 0I$, and only a single description is used. It may be noticed that the performance of MH quantization appears to be up to 5 dB worse than that of linear PCM. However, it is important to keep in mind, that the MH quantizer is optimized for a perceptual measure and not for the MSE. To further stress this point, we have shown the objective difference grades (ODGs) for the linear PCM signal as well as for the MH quantized

[9]We note that in general such behavior cannot be expected, and one would then need to use an alternative estimate of the differential entropy in order to obtain Δ.

[10]The discrete entropy lower bounds the resulting coding rate that one would obtain when using entropy coding on the quantized signal. At high-resolutions, e.g., at least 2 – 3 bits/sample, the resulting coding rate will be very close to the discrete entropy.

signal in Table II.[11] The ODGs provide an indication of the perceived quality of the coded audio signals and are related to the standard ITU-R 5-grade impairment scale as shown in Table I.

TABLE I
RELATIONSHIP BETWEEN THE ITU-R 5-GRADE IMPAIRMENT SCALE AND ODGS [33].

Impairment	ITU-R 5-grade scale	ODG
Imperceptible	5.0	0.0
Perceptible, but not annoying	4.0	-1.0
Slightly annoying	3.0	-2.0
Annoying	2.0	-3.0
Very annoying	1.0	-4.0

As can be seen from Table II, the quality of the MH quantized audio is significantly better than conventional linear PCM, when ODG rather than MSE is the preferred figure of merit. We have also performed simulations where we replaced the scalar quantizer in the PCM setup by a $\log$-quantizer, i.e., the signal is first *compressed* by the $\log$-function and quantized using a uniform scalar quantizer. Then, at the decoder, the inverse operation is required, i.e., the $\exp$-function is applied in order to map the reconstruction from the perceptual domain and back into the Euclidean domain.[12] However, this companding approach did not give better ODGs than that achieved by standard linear PCM encoding.[13]

TABLE II
OBJECTIVE DIFFERENCE GRADES FOR THREE FRAGMENTS OF MUSIC; *Jazz*, *Pop*, AND *Rock*.

Entropy [bits]	MHQ (*Jazz*)	PCM (*Jazz*)	MHQ (*Pop*)	PCM (*Pop*)	MHQ (*Rock*)	PCM (*Rock*)
4	-3.191	-3.733	-3.808	-3.854	-3.864	-3.882
5	-2.890	-3.440	-3.460	-3.752	-3.779	-3.810
6	-1.568	-2.830	-2.570	-3.299	-3.440	-3.512
7	-0.803	-1.837	-1.200	-1.995	-2.382	-2.654
8	-0.473	-1.206	-0.418	-0.855	-1.027	-1.719

We now compare the numerical performance obtained in this section to the analytical expressions provided in Section III-A. We will consider the case of $\bar{R}^* = 6$ bits/sample and use the *Rock* music signal. First, the variance of the music signal is measured to be 0.0385, which results in a differential entropy of $h(\boldsymbol{x}) = -0.3030$ bits/sample, when using the Gaussian approximation. From this, the scaling factor is obtained as $\Delta = 2^{-0.3030-6} = 0.0127$. The average distortion given by (12) is measured to be $\bar{D} = 1.3381 \cdot 10^{-5}$, which is close to $\mathbb{E}[\|\boldsymbol{q}\|^2] = \Delta^2/12 = 1.3366 \cdot 10^{-5}$ (where the first equality is valid under the assumption of uniformly distributed quantization noise $\boldsymbol{q}$) as follows from Lemma 1. It is important to note that $\mathbb{E}[\|\boldsymbol{x} - \boldsymbol{y}\|^2] \neq \Delta^2/12$, since we are not optimizing for the MSE and as such $\Psi_k \neq [1\ 0\ \cdots\ 0]$. Thus, this is not a trivial result, which follows from high-resolution quantization theory. The discrete entropy of the quantized signal is measured to be $\bar{R}^* = 5.9677$ bits/sample, which is close to the desired target rate of 6 bits/sample. At this point, we replace the scalar quantizer by an additive white noise, which is uniformly distributed in the interval $[-\Delta/2; -\Delta/2]$. It follows that $\boldsymbol{\xi}$ is continuous valued and therefore has a density (instead of being discrete due to quantization). Using a nearest-neighbor entropy estimation approach [38], we numerically measure the average differential entropy of $\boldsymbol{\xi}$ to be $h(\boldsymbol{\xi}) = \frac{1}{k}\sum_{i=0}^{k-1} h(\boldsymbol{\xi}_i) = -0.3492$ bits/sample. Since $\boldsymbol{q}$ is uniformly distributed, it is easy to show that $h(\boldsymbol{q}) = \log_2(\Delta)$ and that $\bar{R}^* = h(\boldsymbol{\xi}) - \log_2(\Delta) = 5.9538$ bits/sample which is close to the above measured $\bar{R}^* = 5.9677$ bits/sample obtained using a scalar quantizer.

The lossless coding operation is the same for our scheme as for the schemes used for comparison. Thus, the particular construction is not of importance. In the simulations we tested two different settings. First, an optimal Huffman lossless coder was designed on the empirical statistics of the quantized output. This is an ideal situation. Second, a Gaussian codebook was designed using only knowledge of the variance of the input signal. This is a worst case situation for two reasons: 1) The distribution is not matched to the source distribution. 2) The Gaussian source is the hardest to code under MSE distortion. Thus, if the source variance is fixed, the rate when using a Gaussian codebook is greater than or equal to the rate when using the true distribution. This is also interesting from a practical perspective, since by designing the lossless codebook for a Gaussian distribution (of a fixed variance), one makes sure that the operational coding rate will never exceed that what it would be if the distribution was truly Gaussian. When using the *Jazz* signal, we have measured the average empirical discrete entropy of $\{y_k\}$, as well as the the coding rate obtained after entropy coding using an optimal codebook (i.e., designed using the empirical distribution of the actual sequence $\{y_k\}$). For comparison, we have designed a Gaussian codebook, i.e., by using the variance of $\{y_k\}$ and randomly generating Gaussian samples, which are then used to train a Huffman codebook. Then, we used this unmatched codebook to encode $\{y_k\}$. The obtained rates (in bits/sample) are illustrated in Table III. Notice that in both cases, the operational bit rates are close to the desired discrete entropy of the output, which again is close to the desired target rate.

TABLE III
OPERATIONAL BIT RATES [BITS/SAMPLE] AFTER LOSSLESS CODING USING HUFFMAN CODING.

Target rate	Discrete entropy	Rate: optimal CB	Rate: Gaussian CB
4	3.985	4.021	4.061
5	4.981	5.013	5.053
6	5.979	6.011	6.034
7	6.979	7.010	7.023
8	7.978	8.009	8.054

We next consider an application where final state weighting and vector quantization is used. Note that when vector quantization is utilized, it is important that the first sample of the vector can be decoded independently of the remaining

[11]The ODGs scores are obtained by using the Matlab implementation provided by Kabal et al. [35] of the PEAQ standard [36].

[12]Interestingly, no such operation is required at the decoder for the MH quantization approach, since the encoded symbols are already representing the signal in the original domain.

[13]The log-companding approach is particular useful for fixed-rate coding and when the distortion is the input-weighted mean squared error, where the weight is given by the reciprocal of the square of the input, cf. [37]. However, we are here using entropy-constrained coding and a distortion measure which is different from the input-weighted.

subvector. For example, the whole vector may be quantized using entropy-constrained vector quantization and then the first coordinate of the vector is separately entropy coded and transmitted to the decoder. Alternatively, the first coordinate may be quantized using a scalar quantizer and the remaining coordinates may be quantized using either a vector quantizer or a sequence of scalar quantizers applied individually along the remaining dimensions of the input vector.

Let $N = 4$ so that three future samples are required, and thus there is an inherent delay of three samples. Furthermore, we use the $K = 3$ order perceptual weighting filter from (76). We use the D_4 (four-dimensional) lattice vector quantizer at a bit-rate of $\bar{R}^* = 6$ bits/sample [28].[14] We use a simple state-weighting $P = I$, i.e., the $K \times K$ identity matrix, which results in an ODG of -2.5037 on the *Pop* music signal. On the other hand, when no state-weighting is used, i.e., $P = 0I$, the performance is in this case -2.7391, which is only slightly worse. It is an interesting topic for further study, to examine the impact of final state-weighting on the subjective audio quality and to find optimal weighting matrices.

C. Real-Time Multiple-Description Perceptual MH quantization

We now propose a design for the MD case and where $P = 0I$. Recall from Lemma 4 that a (Euclidean) nearest-neighbor MD quantizer may be used and that we use the index-assignment construction presented in [23]. This results in n descriptions, which are combined at the encoder as described by (64). In the following simulations, we will consider $N = 1$ and $n = 3$ descriptions. Thus, each sample is encoded into three descriptions, which are each treated as a separate packet. Let the weights in (64) be given as $\gamma_0 = \gamma_1 = \gamma_2 = (1-p)p^2, \gamma_{01} = \gamma_{02} = \gamma_{12} = (1-p)^2 p$, and $\gamma_{012} = (1-p)^3$, where $p = 0.1, 0.2, 0.3$. Moreover, let the nesting factor be $\rho = 9$ and let the rates of the side descriptions be identical. Table IV shows the ODGs for different subsets of descriptions when the *Jazz* music signal is encoded. It may be observed that the performance of the individual descriptions as well as the performance when using any two descriptions is largely unaffected by the choice of weights. However, the central reconstruction, i.e., when all descriptions are used, (last column) is highly affected. In fact, at relatively high bit rates (relatively low bit rates), the central reconstruction improves (becomes worse) with increasing packet loss rates. The relationship between weights (and how to form the feedback at the encoder), bit rates, and performance is unfortunately a non-trivial and open problem, see also Footnote 7. Figs. 5(a) – 5(c) show the ODGs for all three fragments at different bit rates. Here the weights are based on $p = 0.1$. It may be noticed that, the more descriptions used in the reconstruction, the better the performance. Moreover, as expected, increasing the bit-rate also leads to better performance.

In Table V, we compare the average run-time perceptual distortion given by (74) to the performance observed at the receiver and obtained by simulations. At this point, we let the weight p denote the the packet-loss rate, which in the simulations is the range $p \in [5\%; 30\%]$ and is incremented in steps of 5%. For each packet-loss rate, the numeric results are averaged over 10 different randomly chosen packet-loss realizations. The shown results are for the *Jazz* fragment using three descriptions and 5 bits/sample per description. Also shown is the theoretic performance obtained if one would use single-description (SD) MH quantization at 15 bits/sample (last row in Table V). From Table V, it is clear that the performance obtained from simulations is close to that described by theory. As expected, the performance decreases as the packet-loss rate increases. Interestingly, a three-description system operating at 5 bits/sample per description and at a packet-loss rate of $p = 30\%$, performs better than a single-description system operating at 15 bits/sample and at a packet-loss rate of $p = 5\%$. The latter observation strengthens the relevance of the scheme proposed in the present work. A suboptimal approach to multiple description coding is repetition coding, i.e., where the same description in a single-description setup is simply repeated a number of times. Table V illustrates the situation when allowing one and two repetitions. When allowing one repetition, the bitrate per description is 7.5 bits/sample, whereas when allowing two repetitions, the bitrate is 5 bits/sample. Thus, the total rate is 15 bits/sample as in the other simulations presented in the table. At high loss rates, it is often the case that only a single description is received and the performance of repetition coding becomes close to that of MD coding.

[14]Only the first sample of the vector is entropy coded and transmitted.

TABLE IV
ODGS FOR DIFFERENT SUBSETS OF DESCRIPTIONS AS A FUNCTION OF THE WEIGHTS. TOP THREE ROWS: $\bar{R}^* = 7$ AND BOTTOM THREE ROWS: $\bar{R}^* = 4$ BITS/SAMPLE PER DESCRIPTION.

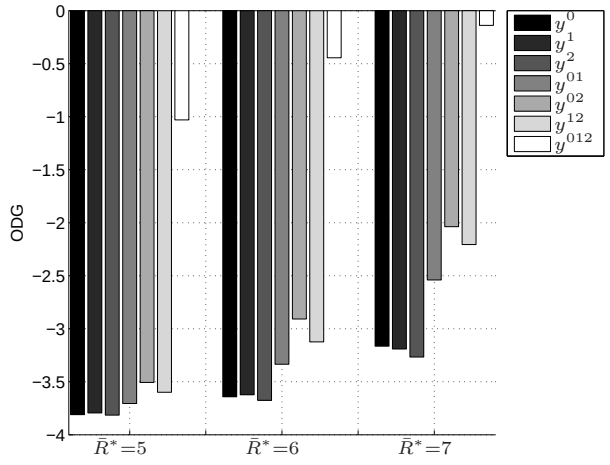
p	y^0	y^1	y^2	y^{01}	y^{02}	y^{12}	y^{012}
0.1	-3.164	-3.191	-3.267	-2.540	-2.038	-2.206	-0.139
0.2	-3.164	-3.212	-3.282	-2.540	-2.068	-2.222	-0.099
0.3	-3.138	-3.204	-3.287	-2.535	-2.044	-2.183	-0.052
0.1	-3.863	-3.803	-3.868	-3.828	-3.751	-3.796	-1.622
0.2	-3.860	-3.806	-3.865	-3.826	-3.745	-3.795	-1.993
0.3	-3.861	-3.794	-3.870	-3.834	-3.741	-3.812	-2.458

TABLE V
AVERAGE PERCEPTUAL DISTORTION $\bar{D}$ GIVEN BY (74) AND BY SIMULATIONS, FOR THE *Jazz* FRAGMENT.

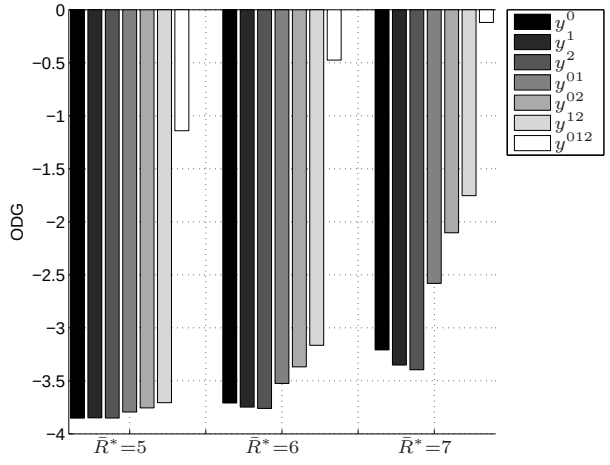
	$p = 0.05$	$p = 0.10$	$p = 0.15$	$p = 0.20$	$p = 0.25$	$p = 0.30$
MD (74) $n = 3$	-49.420	-43.311	-38.886	-35.494	-32.768	-30.499
MD sim. $n = 3$	-48.872	-43.052	-38.869	-35.563	-33.013	-30.982
Rep. sim $n = 3$	-43.314	-41.292	-38.348	-35.344	-32.762	-30.545
Rep. sim $n = 2$	-41.002	-35.080	-31.557	-29.0633	-27.126	-25.554
SD $n = 1$	-28.083	-25.073	-23.312	-22.063	-21.093	-20.302

VI. CONCLUSIONS

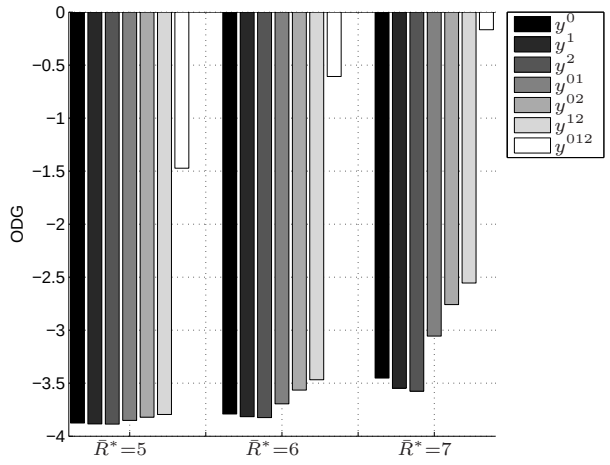
In this work, we have proposed a real-time audio coder which uses elements of multiple-description coding and moving-horizon quantization. In particular, it was shown that MH optimization could be mapped into a domain which allowed the use of existing (Euclidean) nearest-neighbor MD



(a) Jazz



(b) Pop



(c) Rock

Fig. 5. ODGs when using different subsets of descriptions for reconstructions.

quantization techniques. The moving-horizon construction allowed us to efficiently incorporate perceptual weighting. In the single-description case and without packet losses, it was shown that significant gains over linear PCM could be achieved without introducing delay and without having to change the decoding architecture of existing systems. By introducing a few samples delay, with the proposed coder the noise shaping could be improved over what was possible with conventional noise-shaping techniques. It was also shown that the inclusion of multiple descriptions provided a certain degree of robustness towards packet losses.

ACKNOWLEDGMENT

The authors would like to thank the referees for their comments and suggestions, which helped improve the quality and presentation of the paper.

REFERENCES

[1] L. Bramsløw, "Preferred signal path delay and high-pass cut-off in open fittings," *International Journal of Audiology*, vol. 49, pp. 634 – 644, 2010.

[2] A. A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 851 – 857, Nov. 1982.

[3] R. Arean, J. Kovačević, and V. K. Goyal, "Multiple description perceptual audio coding with correlating transform," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 140 – 145, March 2000.

[4] G. Schuller, J. Kovačević, F. Masson, and V. K. Goyal, "Robust low-delay audio coding using multiple descriptions," *IEEE Trans. Speech Audio Processing*, vol. 13, Sep. 2005.

[5] J. Østergaard, O. A. Niamut, J. Jensen, and R. Heusdens, "Perceptual audio coding using n-channel lattice vector quantization," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, pp. 197 – 200, May 2006.

[6] G. C. Goodwin and D. E. Quevedo, "Moving-horizon optimal quantizer for audio signals," *J. Audio Eng. Soc.*, vol. 51, pp. 138 – 149, March 2003.

[7] International Standard ISO/IEC 11172-3 (MPEG), "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.

[8] D. E. Quevedo and G. C. Goodwin, "Multistep optimal analog-to-digital conversion," *IEEE Trans. Circuits Syst. I*, vol. 52, pp. 503 – 515, March 2005.

[9] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Prentice-Hall, 1984.

[10] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory And Design*. Nob Hill Publishing, 2009.

[11] D. E. Quevedo, H. Bölcskei, and G. C. Goodwin, "Quantization of filter bank frame expansions through moving horizon optimization," *IEEE Trans. on Signal Processing*, vol. 57, pp. 503 – 515, February 2009. February.

[12] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1152 – 1159, July 1996.

[13] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmonic. Anal.*, vol. 5, pp. 332 – 269, July 1998.

[14] V. K. Goyal, "Transform coding with integer-to-integer transforms," *IEEE Trans. Inf. Theory*, vol. 46, pp. 465 – 473, march 2000.

[15] P. Hao and Q. Shi, "Matrix factorization for reversible integer mapping," *IEEE Trans. Signal Proc.*, vol. 49, pp. 2314 – 2324, October 2001.

[16] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley, 1991.

[17] R. Zamir and M. Feder, "Information rates of pre/post-filtered dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1340 – 1353, September 1996.

[18] M. S. Derpich, J. Østergaard, and G. C. Goodwin, "The quadratic Gaussian rate-distortion function for source uncorrelated distortions," in *Proc. Data Compression Conf.*, 2008.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379 – 423; 623 – 656, July and October 1948.

[20] R. Zamir, Y. Kochman, and U. Erez, "Achieving the Gaussian rate-distortion function by prediction," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3354 – 3364, July 2008.

[21] E. Silva, M. Derpich, and J. Østergaard, "A framework for control system design subject to average data-rate constraints," *IEEE Transactions on Automatic Control*, 2010. Accepted for publication.

[22] J. Massey, "Causality, feedback and directed information," in *Proceedings of the International Symposium on Information Theory and its Applications*, (Hawaii, USA), 1990.

[23] J. Østergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1956 – 1973, 2006.
[24] J. Østergaard, R. Heusdens, and J. Jensen, "n-channel asymmetric entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inf. Theory*, vol. 56, pp. 6354 – 6375, December 2010.
[25] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. Jour.*, vol. 59, pp. 1909 – 1921, December 1980.
[26] J. Chen, C. Tian, and D. Diggavi, "Multiple description coding for stationary Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2868 – 2881, June 2009.
[27] J. Østergaard, Y. Kochman, and R. Zamir, "Colored Gaussian multiple descriptions: Spectral-domain characterization and time-domain design," *IEEE Trans. Inf. Theory*, 2010. Submitted. Electronically available at: http://arxiv.org/abs/1006.2002.
[28] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*. Springer, 3rd ed., 1999.
[29] H. F. F. Baumgarte and C. Ferekidis, "A nonlinear psychoacoustic model applied to the ISO mpeg layer 3 coder," in *Proc. 99th AES Symp.*, Oct. 1995.
[30] G. D. T. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *Trans. Speech and audio Proc.*, vol. 10, p. 379, Sep. 2002.
[31] J. D. Markel and A. H. Gray, *Linear prediction of speech*. Prentice Hall, 1976.
[32] E. Zwicker and H. Fastl, *Psychoacoustics: facts and models*. Springer series in information sciences, Springer, Berlin, 2nd ed., 1999.
[33] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*. Kluwer Academic Publisher, 2003.
[34] R. Gray, *Source Coding Theory*. Kluwer Academic Press, 1990.
[35] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality." Technical Report, McGill University, Version 2: 2003-12-08, 2003.
[36] International Telecommunication Union, "ITU-R recommendation BS.1387: Method for objective measurements of perceived audio quality (PEAQ)," 2001.
[37] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: multidimensional companding," *IEEE Trans. Inf. Theory*, vol. 45, pp. 548 – 561, March 1999.
[38] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley-Interscience, 2nd ed., 2001.

Jan Østergaard (S'98 – M'99) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1999 and the Ph.D. degree (*cum laude*) in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 2007. From 1999 to 2002, he worked as an R&D engineer at ETI A/S, Aalborg, Denmark, and from 2002 to 2003, he worked as an R&D engineer at ETI Inc., Virginia, United States. Between September 2007 and June 2008, he worked as a post-doctoral researcher in the Centre for Complex Dynamic Systems and Control, School of Electrical Engineering and Computer Science, The University of Newcastle, NSW, Australia. From June 2008 to March 2011, he worked as a post-doctoral researcher at Aalborg University, Aalborg, Denmark. He has also been a visiting researcher at Tel Aviv University, Tel Aviv, Israel, and at Universidad Técnica Federico Santa María, Valparaíso, Chile. He has received a Danish Independent Research Councils Young Researchers Award and a fellowship from the Danish Research Council for Technology and Production Sciences. Dr. Østergaard is currently an Associate Professor at Aalborg University, Aalborg, Denmark.

Daniel E. Quevedo (S'97 – M'05) received Ingeniero Civil Electrónico and Magister en Ingeniería Electrónica degrees from the Universidad Técnica Federico Santa María, Valparaíso, Chile in 2000. In 2005, he received the Ph.D. degree from The University of Newcastle, Australia, where he is currently a research academic. He has been a visiting researcher at ETH Zürich, Switzerland, at Uppsala University, Sweden, at The University of Melbourne, Australia, at Aalborg University, Denmark, at KTH, Stockholm, and at Kyoto University, Japan.

Dr. Quevedo was supported by a full scholarship from the alumni association during his time at the Universidad Técnica Federico Santa María and received several university-wide prizes upon graduating. He received the IEEE Conference on Decision and Control Best Student Paper Award in 2003 and was also a finalist in 2002. In 2009, he was awarded an Australian Research Fellowship. His research interests cover several areas of automatic control, signal processing, communications, and power electronics.

Jesper Jensen received the M.Sc degree in electrical engineering and the Ph.D degree in signal processing from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively.

From 1996 to 2000 he was with the Center for Person Kommunikation (CPK), Aalborg University, as a Ph.D student and assistant research professor. From 2000 to 2007 he was a post-doctoral researcher and assistant professor with Delft University of Technology, The Netherlands, and an external associate professor with Aalborg University, Denmark. Currently, he is with Oticon A/S, Denmark. His main research interests are in the area of acoustical signal processing, including signal retrieval from noisy observations, coding, speech and audio modification and synthesis, intelligibility enhancement of speech signals, and perceptual aspects of signal processing.