Multiple Description Spherical Quantization of Sinusoidal Parameters with Repetition Coding of the Amplitudes

Jensen, Jesper Rindom; Christensen, Mads Græsbøll; Larsen, Morten Holm; Jensen, Søren Holdt; Larsen, Torben

Published in:
Proc. of Asilomar Conference on Signals, Systems, and Computers

Publication date:
2009

Document Version
Accepted author manuscript, peer reviewed version

MULTIPLE DESCRIPTION SPHERICAL QUANTIZATION OF SINUSOIDAL PARAMETERS WITH REPETITION CODING OF THE AMPLITUDES

Jesper Rindom Jensen[1], *Mads Græsbøll Christensen*[2], *Morten Holm Jensen*[3], *Søren Holdt Jensen*[1] *and Torben Larsen*[1]

[1] Dept. of Electronic Systems, Aalborg University, Denmark, {jrj,shj,tl}@es.aau.dk

[2] Dept. of Media Technology Aalborg University, Denmark, mgc@imi.aau.dk

[3] Widex A/S Værløse, Denmark, holm@ieee.org

ABSTRACT

Recently, multiple description spherical trellis-coded quantization (MDSTCQ) for quantization of sinusoidal parameters was proposed, which suffered from a suboptimal implementation. Therefore, we propose the multiple description spherical quantization with repetition coding of the amplitudes (MDSQRA) scheme. Here, we apply repetition coding to the amplitudes, whereas the phases and the frequencies are coded using multiple description quantization. We measure the relation between the expected perceptual distortions for the two quantization schemes, which shows that MDSQRA outperforms MDSTCQ for most packet-loss probabilities. Also, we measure the perceptual distortion obtained when applying MDSQRA on synthetic data and show that it matches the expected perceptual distortion. Furthermore, due to MDSTCQs suboptimal implementation, MDSQRA outperforms MDSTCQ in terms of measured perceptual distortion even at packet-loss probabilities where the expected perceptual distortion was lowest for MDSTCQ.

***Index Terms*—** Multiple description coding, parametric audio coding, spherical quantization, robust audio coding, sinusoidal coding

1. INTRODUCTION

The problem of reducing the bit-rate for a given audio quality has attracted a considerable amount of attention in the last couple of decades. In many applications, it is desired to reduce the bit-rate significantly without sacrificing the perceptible audio quality. Therefore, it is often not sufficient to just reduce the statistical redundancies of the audio signals, if a large amount of compression is required. The solution to this problem could be to use a perceptual audio coding scheme. In such a coding scheme, it is taken into account that certain parts audio signals are inaudible to the human ear. By introducing a perceptually weighted coding error, the coding gain can be improved compared to a lossless coding scheme. The perceptual weighting can be performed by applying a masking threshold below which signal components are inaudible. This masking threshold can be calculated from the time, frequency ad amplitude characteristics of the audio signal [1].

Parametric audio coding is a subclass of perceptual audio coding schemes, in which it is exploited that many audio signals can be represented by a few perceptually important parameters. A well-known parametric audio coding scheme is sinusoidal coding, where the audio signal is described as a sum of sinusoids. Sinusoidal coding thereby entails the task of estimating the most perceptually relevant parameters and quantization of the parameters to allow for transmission. There exist several methods for estimation of the sinusoidal parameters (see, e.g., [2]). Recently, there has also been developed several methods for quantization of the parameters. Common for these methods is that they minimize a perceptual distortion measure subject to a certain rate constraint. Some fundamental quantization schemes in this regard are polar quantization (PQ) in [3, 4] and spherical quantization (SQ) in [5]. Another scheme, which achieves a lower distortion compared to PQ and SQ, is trellis-coded quantization (TCQ) proposed in [6]. However, both SQ and TCQ are not suited for quantization of parameters to be transmitted over lossy networks. One error concealment method that can improve the transmission robustness is multiple description coding (MDC) [7]. A recent quantization scheme employing MDC is the multiple description spherical trellis-coded quantization (MDSTCQ) scheme proposed in [8].

One downside of the MDSTCQ scheme is that it can not be implemented directly. The reason for this is, that the phase and frequency quantizers depend on the quantized amplitudes. Since MDC is used on all parameters in MDSTCQ, the decoded amplitudes will be different if one or two side descriptions are received by the decoder. This will inevitably imply a encoder/decoder mismatch and a suboptimal implementation. Therefore, we propose a new quantization scheme, namely, multiple description spherical quantization with repetition coding of the amplitudes (MDSQRA). That is, in the proposed MDSQRA scheme, the amplitudes are coded repetitively, making MDSQRA directly realizable. In this paper, we consider the derivation of the quantizers and

evaluation of them using objective measurements. For details on how the quantizers can be implemented into a sinusoidal coder and for subjective listening tests see, e.g., [9].

The rest of the paper is organized as follows. In Section 2 we derive the proposed MDSQRA scheme for robust quantization of sinusoidal parameters. We measure the performance of the proposed quantizers and the experimental results are delivered in Section 3. Finally, Section 4 concludes on our work.

2. PROPOSED QUANTIZATION SCHEME

In parametric audio coding, the audio signals are modeled by a set of few relevant parameters. One popular parametric audio coding scheme is sinusoidal coding, where we for $n = 0, \ldots, N-1$ assume the following sinusoidal model

$$\hat{x}(n) = \sum_{l=1}^{L} a_l \cos(\nu_l n + \phi_l) \ , \tag{1}$$

with a_l, ν_l and ϕ_l denoting the amplitude, frequency and phase characterizing the lth sinusoid, respectively, and L is the model order. There are mainly two important tasks in a parametric audio coder: 1) estimation of the parameters best describing the audio signal (e.g., using the perceptual matching pursuit algorithm [2, 10]) and 2) quantization of the parameters to allow for transmission. In this paper, we consider the latter part, namely, quantization of the sinusoidal parameters.

The proposed quantizers are derived on the basis of minimization of a perceptual distortion measure subject to an entropy constraint. The utilized perceptual distortion measure was first introduced in [11] and is defined as

$$D = \frac{1}{2\pi} \int_0^{2\pi} \mu_{x(a,\phi,\nu)}(\omega) |\epsilon(\omega)|^2 d\omega \ , \tag{2}$$

where $\mu_{x(a,\phi,\nu)}(\nu)$ is a perceptual weighting function and $\epsilon(\omega)$ is the DTFT of the windowed error. The error can be expressed $\epsilon(\omega)$ as

$$\epsilon(\omega) = \sum_{n=n_0}^{n_0+N-1} w(n)(x(n) - \hat{x}(n)) e^{-j\omega n} \ , \tag{3}$$

with $x(n)$ being the observed signal and $w(n)$ is a window of length N. We assume sufficiently long windows such that the cross-term distortion of the sinusoids are neglectable. Therefore, we only consider minimization of the distortion contribution from quantization of one sinusoid. It follows from [5], that under high-rate assumptions, the perceptual distortion measure can be approximated by

$$D \approx \frac{\mu_{x(a,\phi,\nu}}{2\|w\|^{-2}} \left((a-\tilde{a})^2 + a\tilde{a}((\phi-\tilde{\phi})^2 + \sigma^2(\nu-\tilde{\nu})^2) \right) \ , \tag{4}$$

where $\sigma^2 = \frac{1}{\|w\|^2} \sum_{n=-N/2}^{N/2-1} w^2(n) n^2$, $\{\tilde{a}, \tilde{\phi}, \tilde{\nu}\}$ is the set of quantized parameters, and $\|w\|^2$ is the squared norm of the window. Using high-rate assumptions we can assume that $a\tilde{a} = \tilde{a}^2$ in (4). Furthermore, it can be seen from (4) that the amplitude, phase and frequency quantizers can be designed independently. As mentioned introductory we design the quantizers to be robust against transmission over lossy network channels. Therefore, the MDC principle is employed such that each quantizer generates two descriptions. If only one description is received in the decoder, a low quality reconstruction is obtained with a side distortion D_s where $s = \{1, 2\}$. If both descriptions are received the distortion reduces to the central distortion D_0. We design the quantizers for robust transmission over a packet erasure channel, where the packets are lost independently with probability p. The expected distortion for such a channel is

$$\mathrm{E}[D] = (1-p)^2 \mathrm{E}[D_0] + 2p(1-p)\mathrm{E}[D_s] + p^2 \sigma_x^2 \ , \tag{5}$$

with σ_x^2 being the signal variance. The side distortions are assumed balanced, i.e., $\mathrm{E}[D_1] = \mathrm{E}[D_2]$. It can be shown that the expected side distortion can be approximated by

$$\begin{aligned} \mathrm{E}[D_s] \approx & \frac{\|w\|^2}{24} \iiint f_{A,\Phi,\Upsilon}(a,\phi,\nu) \mu_{x(a,\phi,\nu)}(\tilde{\nu}) \\ & \times (g_a^{-2} + \tilde{a}^2 (g_\phi^{-2} + \sigma^2 g_\nu^{-2})) da d\phi d\nu \ , \end{aligned} \tag{6}$$

where $f_{A,\Phi,\Upsilon}(a,\phi,\nu)$ is the joint pdf of the amplitudes, phases and frequencies, and g_a, g_ϕ and g_ν are the quantization point densities. For the phase and frequency quantizers, the ratio between the central and side distortions is $\frac{1}{(2N_k)^2}$ since we use the modified multiple description scalar quantization (MMDSQ) principle in [12] on these parameters. Note that N_k is the number of refined reconstruction points for the kth parameter. We can then express the expected central distortion as

$$\begin{aligned} \mathrm{E}[D_0] \approx & \frac{\|w\|^2}{96} \iiint f_{A,\Phi,\Upsilon}(a,\phi,\nu) \mu_{x(a,\phi,\nu)}(\tilde{\nu}) \\ & \times \left(4g_a^{-2} + \tilde{a}^2 \left(\frac{g_\phi^{-2}}{N_\phi^2} + \sigma^2 \frac{g_\nu^{-2}}{N_\nu^2} \right) \right) da d\phi d\nu \ . \end{aligned} \tag{7}$$

The total expected distortion can then be approximated by inserting (6) and (7) into (5). We design the quantizers subject to an entropy constraint on one side description. By using high-rate assumptions we can approximate the entropy for one side description as

$$\begin{aligned} H_s \approx & h(a,\phi,\nu) \\ & + \iiint f_{A,\Phi,\Upsilon}(a,\phi,\nu) \log_2(g_a g_\phi g_\nu) da d\phi d\nu \\ & + \iiint f_{A,\Phi,\Upsilon}(a,\phi,\nu) \frac{\log_2 N_\phi N_\nu}{2} da d\phi d\nu \ , \end{aligned} \tag{8}$$

	Total expected distortion
MDSQRA	$\frac{3}{24}(1-p^2)\left(\frac{1-p}{8p}\right)^{\frac{1}{3}}\left(\frac{4p}{1+p}\right)^{\frac{2}{3}}\|w\|^2 2^{\frac{2}{3}\left(-\tilde{H}_s+\log_2(\sigma)+\frac{3}{2}\rho(a,\phi,\nu)+2b(a)\right)}+p^2\mathrm{E}[x^2]$
MDSTCQ	$\sqrt{\frac{\Gamma p}{8}}(1-p)^{\frac{3}{2}}\|w\|^2 2^{\frac{2}{3}\left(-\tilde{H}_s+\log_2(\sigma)+\frac{3}{2}\rho(a,\phi,\nu)+2b(a)\right)}+p^2\mathrm{E}[x^2]$

Table 1. Total expected distortions for MDSQRA and MDSTCQ.

where $h(a,\phi,\nu)$ is the joint differential entropy of the amplitudes, phases and frequencies. We introduce a new variable $\tilde{H}_s = H_s - h(a,\phi,\nu)$ and define a cost-function

$$J = (1-p)^2\mathrm{E}[D_0] + 2p(1-p)\mathrm{E}[D_s] + \lambda\tilde{H}_s\ , \quad (9)$$

with λ being the Lagrange multiplier. By minimizing the cost-function (9) using the Lagrange multiplier method, it can be shown that the optimal quantization point densities and numbers of refined reconstruction points are given by

$$g_a^2 = \left(\frac{1+p}{4p}\right)^{\frac{2}{3}}\frac{\mu_{x(a,\phi,\nu)}}{N^{\frac{2}{3}}} \times 2^{\frac{2}{3}\left(\tilde{H}_s-\log_2(\sigma)-\frac{2}{3}\rho(a,\phi,\nu)-2b(a)\right)} \quad (10)$$

$$g_\phi^2 = \tilde{a}^2\frac{4p}{1+p}g_a^2 \quad (11)$$

$$g_\nu^2 = \tilde{a}^2\sigma^2\frac{4p}{1+p}g_a^2 \quad (12)$$

$$N_\phi^2 = N_\nu^2 = \frac{1-p}{8p}\ , \quad (13)$$

where

$$\rho(a,\phi,\nu) = \iiint f_{A,\Phi,\Upsilon}(a,\phi,\nu) \times \log_2(\mu_{x(a,\phi,\nu)})dad\phi d\nu \quad (14)$$

$$b(a) = \int f_A(a)\log_2(a)da\ . \quad (15)$$

By inserting (10)-(13) into (6) and (7), which are then inserted into (5) we get an expression for the total expected distortion. The total expected distortion expression is given in Tab. 1 along with the total expected distortion of MDSTCQ [8]. Note that Γ in Tab. 1 is a correction factor used for approximating the expected distortion of MDSTCQ [6].

3. EXPERIMENTAL RESULTS

We will now present the experimental results obtained from the performance measurements of the proposed quantization scheme. First, we compare MDSQRA and MDSTCQ using their respective expected perceptual distortion expressions. The expressions used for this simulation are found in Tab. 1. Note that the only difference between the quantizers are two distortion factors which depend on the packet-loss probability p and Γ. These factors, which are proportional to the expected perceptual distortion approximations, can be written as

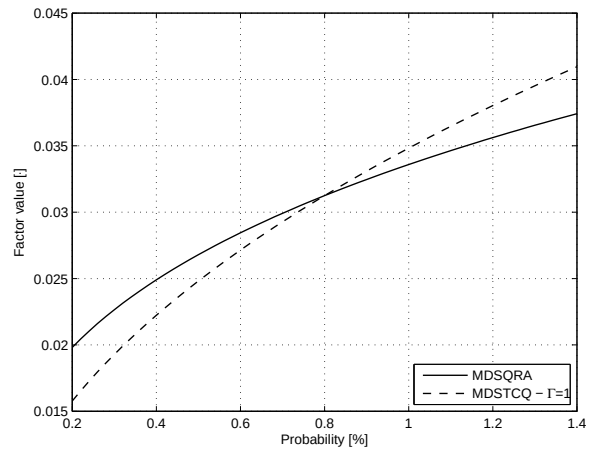

Fig. 1. Comparison of the quantizer distortion factors $k_{\text{MDSQRA}}(p)$ and $k_{\text{MDSTCQ}}(p,\Gamma)$ for a selected packet-loss probability range with $\Gamma = 1$.

$$k_{\text{MDSQRA}}(p) = \frac{3}{24}(1-p^2)\left(\frac{1-p}{8p}\right)^{\frac{1}{3}}\left(\frac{4}{1+p}\right)^{\frac{2}{3}} \quad (16)$$

$$k_{\text{MDSTCQ}}(p,\Gamma) = \left(\frac{\Gamma p}{8}\right)^{\frac{1}{2}}(1-p)^{\frac{2}{3}}\ . \quad (17)$$

In Fig. 1, we have plotted the factors in (16) and (17) for a limited range of packet-loss probabilities and for $\Gamma = 1$. Setting $\Gamma = 1$ leads to the lowest expected perceptual distortion for MDSTCQ and it corresponds to an infinite number of states and an infinite dimension in the MDSTCQ scheme [6]. From these results, it can be seen that the expected perceptual distortion of MDSTCQ is lower than the distortion of MDSQRA for $0.2\% < p < 0.8\%$ for this experimental setup. For $0.8\% < p < 1.4\%$, however, the expected perceptual distortion of MDSQRA is lower than for MDSTCQ. That is, we can not unequivocally conclude, that the expected perceptual distortion of MDSQRA will be lower than for MDSTCQ for all packet-loss probabilities and vice versa.

However, it is not sufficient to compare the expected perceptual distortions directly, since MDSTCQ requires a suboptimal implementation. Therefore, we have measured the

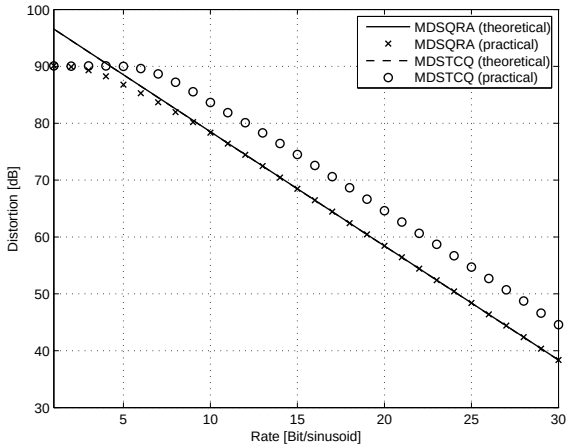

Fig. 2. Expected and measured perceptual distortions for MDSQRA and MDSTCQ. The quantizers were designed for $p = 0.78\%$ and $\Gamma = 1$.

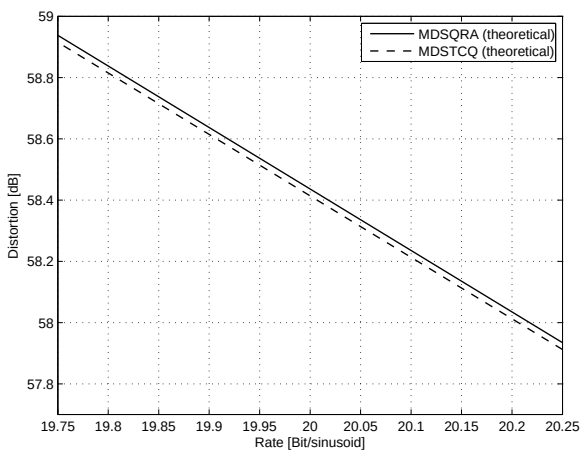

Fig. 3. Zoomed plot of the expected and measured perceptual distortions for MDSQRA and MDSTCQ. The quantizers were designed for $p = 0.78\%$ and $\Gamma = 1$.

perceptual distortion of MDSQRA and MDSTCQ as a result of quantization of synthetic sinusoidal parameters. The perceptual distortion was measured using the expression in (4).

Both MDSQRA and MDSTCQ were designed to have $N = 4$ number of refined reconstruction points for the quantizers employing MDC. For MDSTCQ, the factor Γ was set to one, resulting in both MDSQRA and MDSTCQ being designed for $p = 0.78\%$. Note that MDSTCQ is implemented using the suboptimal implementation proposed in [8]. The synthetic sinusoids were generated by assuming Rayleigh dis-

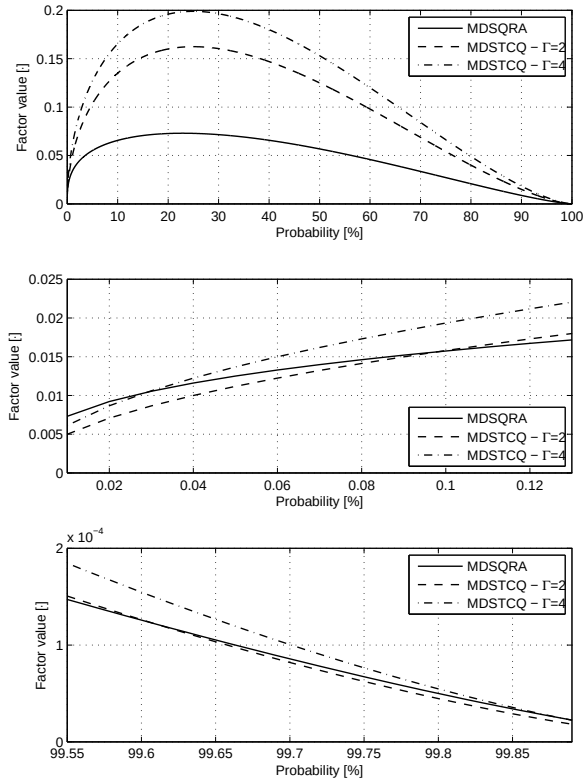

Fig. 4. Comparison of the quantizer distortion factors $k_{\text{MDSQRA}}(p)$ and $k_{\text{MDSTCQ}}(p, \Gamma)$ with $\Gamma = \{2, 4\}$, respectively (top). Also, zoomed plots for the lower (middle) and upper (bottom) packet-loss probability ranges are shown.

tributed amplitudes and frequencies with the pdf

$$f_X(x) = \frac{x}{\beta} e^{-\frac{x^2}{2\beta}} \ . \tag{18}$$

We chose $\beta = \{1000, 0.25\}$, respectively, for the amplitudes and frequencies. The phase was assumed to be uniformly distributed in the interval $0 \leq \phi < 2\pi$.

The quantization point densities of the quantizers were found from a training set consisting of 100,000 sinusoidal parameter sets. Following, we used the designed quantizers to quantize 10,000 sinusoidal parameter sets for different target entropies. In Fig. 2 we have plotted the expected and measured perceptual distortions as functions of the target entropies for both MDSQRA and MDSTCQ. First, notice that the measured perceptual distortion of MDSQRA converges against the expected perceptual distortion for an increasing bit-rate. Due to the use of high-rate assumptions it is expected that the expected and measured perceptual distortion do not match at low bit-rates. Second, it can be seen that the suboptimal implementation of MDSTCQ leads to a significant

increase in perceptual distortion. That is, even though the expected perceptual distortion of MDSTCQ is lower than for MDSQRA in this experiment, MDSQRA outperforms MDSTCQ in terms of measured perceptual distortion. To verify that the expected perceptual distortion of MDSTCQ is lower than for MDSQRA at $p = 0.78\%$, see the zoomed plot in Fig. 3.

Finally, we have compared the factors in (16) and (17) for the whole range of packet-loss probabilities. However, in this simulation we chose two different Γs for MDSTCQ, i.e., $\Gamma = \{2, 4\}$, respectively. The factor Γ can be found through simulation and from [6] we conclude that Γ being somewhere between 2 and 4 is typical. The results from this simulation are depicted in Fig. 4. From the figure it is clearly seen that MDSQRA outperforms MDSTCQ for most packet-loss probabilities. Also, MDSQRA might also perform MDSTCQ at the other packet-loss probabilities in terms of measured perceptual distortion due to the suboptimal implementation of MDSTCQ.

4. DISCUSSION

In this paper, we considered a novel multiple description coding scheme for quantization of sinusoidal parameters, namely, multiple description spherical quantization with repetition coding of the amplitudes (MDSQRA). The proposed quantizers are simple, closed-form and computationally efficient since the multiple description scalar quantization principle is employed. One of the major advantages of the proposed quantization scheme is, that it is directly implementable as opposed to multiple description spherical trellis-coded quantization (MDSTCQ) proposed in [8]. Another advantage is a reduced computational complexity. We found that only a single factor, dependent on the packet-loss probability and a scaling factor which relates TCQ to SQ, differs between the proposed quantizers and MDSTCQ. Through simulations we plotted the distortion factors of MDSQRA and MDSTCQ and found that MDSQRA outperforms MDSTCQ for most packet-loss probabilities. Also, we applied the proposed quantizers and MDSTCQ on synthetic sinusoidal parameters and measured the perceptual distortion. These simulations showed, that even though the expected perceptual distortion of MDSTCQ is lower than for MDSQRA at some packet-loss probabilities, MDSQRA might still outperform MDSTCQ for the exact same packet-loss probabilities in terms of measured perceptual distortion. In other work [9], we described how the proposed quantization scheme can be implemented into a real audio coder, and verified the simulation results in this paper using listening tests. In future work, the proposed quantizers should be generalized to support more than two descriptions, which would be preferable for high packet-loss probabilities.

5. REFERENCES

[1] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Emerald Group Publishing Limited, 5. edition, 2008.

[2] M. G. Christensen and S. H. Jensen, "On perceptual distortion minimization and nonlinear least-squares frequency estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 14, no. 1, pp. 99–109, Jan. 2006.

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

[4] R. Vafin, D. Prakash, and W. B. Kleijn, "On frequency quantization in sinusoidal audio coding," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 210–213, 2005.

[5] P. Korten, J. Jensen, and R. Heusdens, "High-resolution spherical quantization of sinusoidal parameters," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 966–981, 2007.

[6] M. H. Larsen, M. G. Christensen, and S. H. Jensen, "Variable dimension trellis-coded quantization of sinusoidal parameters," *IEEE Signal Process. Lett.*, vol. 15, pp. 17–20, 2008.

[7] V. K. Goyal, "Multiple description coding: compression meets the network," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[8] M. H. Larsen, M. G. Christensen, and S. H. Jensen, "Multiple description trellis-coded quantization of sinusoidal parameters," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5287–5291, Oct. 2008.

[9] J. R. Jensen, M. G. Christensen, M. H. Jensen, S. H. Jensen, and T. Larsen, "Robust parametric audio coding using multiple description coding," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1083–1086, Dec. 2009.

[10] R. Heusdens, R. Vafin, and W. B. Kleijn, "Sinusoidal modeling using psychoacoustic-adaptive matching pursuits," *IEEE Signal Process. Lett.*, vol. 9, no. 8, pp. 262–265, 2002.

[11] S. Par, A. Kohlrausch, R. Heusdens, J. Jensen, and S. H. Jensen, "A perceptual model for sinusoidal audio coding based on spectral integration," *EURASIP J. on Applied Signal Processing*, vol. 2005, pp. pp. 1292–1304, 2005.

[12] C. Tian and S. S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 329–332, Apr 2005.