Bit-Erasure Channel Decoding for GMM-based Multiple Description Coding

Yannis Agiomyrgiannakis, Yannis Stylianou

Department of Computer Science, University of Crete, Greece
jagiom@csd.uoc.gr, yannis@csd.uoc.gr

1. Introduction

There is a growing need towards the transmission of multimedia over IP networks. Unfortunately, in IP networks there is no mechanism to assure the quality-of-service and the arrival of a packet may be delayed or the ordering of the packets may be altered. This is usually compensated by a buffer at the decoder, but the size of this buffer is restricted in real-time communications, resulting to increased packet losses. Multiple Description Coding (MDC) is a plausible framework to increase the robustness of communication upon packet losses. In MDC, the source is encoded in two side descriptions which are routed independently. If both descriptions arrive, we get a high-quality reconstruction of the source, referred to as “central reconstruction”. If only one description is received, an intermediate quality reconstruction is made, referred to as “side reconstruction”. The MDC framework can be seen as a mechanism to introduce redundancy into the bitstream by trading the distortion of the central reconstruction (referred to as central distortion) for the distortion of the side reconstructions (referred to as side distortions).

In the typical MDC framework, the network is modeled as a packet erasure channel (a description can be either received or lost); the possibility of introducing bit-errors is ignored. Introducing bit-error correction directly to MD bitstreams may be suboptimal [1]. However, this is not the case if the MD quantizer is optimized for both packet-erasure and bit-erasure channels [2]. The importance of error-resilient MDC becomes more apparent with the convergence between wireline and wireless networks due to the fact that wireless transmission suffers from bit-errors.

The first systematic approach on MDC was made by Vaishampayan in [3] with Multiple Description Scalar Quantization (MDSQ). This method is extended to transform coding by Vaishampayan and Barlho in [4]. A further extension to sources that can be modeled with a GMM is provided by Samuelsson and Plasberg in [5]. The latter method (referred to as GMM-MDSQ) in the rest of the paper) can find applications in speech coding and particularly in the quantization of spectral envelopes is sinusoidal speech codecs [6], [7], [8], [9]. This paper extends GMM-MDSQ to the case of bit-erasure channels. It is shown that the redundancy introduced by GMM-MDSQ can be used to reduce the penalty of bit-errors that occur in one description. The alternative would be to disregard the damaged description and conduct a side-reconstruction using the undamaged description. The proposed method can be used to make a reconstruction that is significantly better than this side reconstruction. It is demonstrated that in a practical setting of MDC for speech spectral envelopes, single and double bit-errors can almost be corrected in terms of the mean-squared error. However, correcting more bit-errors faces a significant computational barrier.

The proposed method is related to the bit-erasure channel decoding method initially introduced by Subramaniam in [10] (Chapter 6). When the number of bit-errors is known, the channel decoder provides a set of candidate indices that contains the transmitted indices. For example for a 30-bit encoding and 2 detected bit-errors there are 30×29×870 candidate indices and thus 870 candidate codewords. In [10], a reconstruction is made by averaging these vectors according to the conditional probability of the current vector given the information of past vectors (speech spectral envelopes). In this paper, the correlated but undamaged side description is used to significantly reduce the number of candidate codewords of the damaged side description while the averaging is made using the probability of the source. The extension of the proposed method to the case where both descriptions are damaged is also discussed.

The rest of the paper is organized as follows: Section 2 presents the necessary background for the paper, Section 3 describes the proposed bit-erasure channel decoding method for GMM-MDSQ and Section 4 presents an experimental evaluation. Finally, Section 5 concludes the paper.

2. Background

2.1. Multiple Description Scalar Quantization

In MDSQ, the correlations between the two descriptions are introduced via an index-assignment matrix that links the two side description codebooks with the central description codebook [3]. Let \( \lambda \sim p(\lambda) \) be the scalar random variable which describes the encoded data and \( r \) be the total rate of the MDSQ system. Assuming balanced descriptions with equal rates, each description will then have a rate of \( r = \frac{1}{2}r \) bits. Let \( C_0 = \{ \hat{\lambda}^{(0)}_i : i = 1, \ldots, 2^r \} \) be the codebook of the central description. \( C_1 = \{ \hat{\lambda}^{(1)}_j : j = 1, \ldots, 2^r \} \) be the codebook of description 1 and \( C_2 = \{ \hat{\lambda}^{(2)}_j : j = 1, \ldots, 2^r \} \) be the codebook of description 2.

The two side descriptions, say \( i \) and \( j \), are routed independently to the receiver that has three decoders, a central decoder...
and two side decoders, one for each description. When both descriptions are received, the central decoder returns \( \hat{x}^{(0)} \in C_0 \). When only the first description is received, side decoder 1 returns \( \hat{x}^{(1)} \in C_1 \). Accordingly, when only the second description is received, side decoder 2 returns \( \hat{x}^{(2)} \in C_2 \). The encoder finds the pair of indices \((i, j)\) that minimizes the total distortion:

\[
D_{tot} = D_0 + \frac{\rho}{1 - \rho}(D_1 + D_2),
\]

where \( \rho \) is the packet loss probability, \( D_0 \) is the mean central distortion and \( D_1, D_2 \) are the mean distortions for the two side descriptions, respectively. The total distortion measure in (1) is based on the assumption that each description is routed through a symmetric channel and that both channels have the same packet loss probability \( \rho \). In practice, varying \( \rho \) provides a mechanism to exchange central distortion for side distortions. The index assignment mapping links a pair of side codebooks entries \((i, j)\) to one central codebook entry. Note that at higher loss probabilities \( \rho \), \( C_0 \) has less than \( 2^R \) codepoints. Therefore, not all possible pairs of indices \((i, j)\) correspond to a central description codepoint. The encoder can produce only the pairs of indices \((i, j)\) that are linked to a central description codepoint and those pairs will be referred to as valid indices. The corresponding central and side description codepoints will be referred to as valid codepoints.

Let \( Q_i^{(0)} \), \( Q_i^{(1)} \) and \( Q_i^{(2)} \) be the quantization cells associated with central codepoint \( \hat{x}^{(0)} \), \( \hat{x}^{(1)} \) and \( \hat{x}^{(2)} \), respectively. Let \( I_i(j) \equiv \{ j : \hat{x}^{(0)}_j \in C_0 \} \) be the set of valid indices \( j \) when \( i \) is known, and \( I_i(j) \equiv \{ j : \hat{x}^{(0)}_j \in C_0 \} \) be the set of valid indices \( i \) when \( j \) is known. Then, the side description quantization cells \( Q_i^{(0)}, Q_i^{(1)}, Q_i^{(2)} \) can be expressed as unions of the \( Q_i^{(j)} \) cells:

\[
Q_i^{(0)} = \bigcup_{j \in I_i(i)} Q_{i,j}^{(0)}, \quad Q_i^{(1)} = \bigcup_{i \in I_i(j)} Q_{i,j}^{(1)}, \quad Q_i^{(2)} = \bigcup_{i \in I_i(j)} Q_{i,j}^{(2)}.
\]

When only the quantization cells \( Q_{i,j}^{(0)} \) and the mappings \( I_i(j), I_j(i) \) are known, the optimal (in the mean-square-error sense) codepoints can be computed by taking expectations over the quantization cells using \( p(\lambda) \):

\[
\hat{x}^{(0)}_{i,j} = \int_{Q_{i,j}^{(0)}} \lambda p(\lambda|i,j) d\lambda,
\]

\[
\hat{x}^{(1)}_{i} = \sum_{j \in I_i(i)} \int_{Q_{i,j}^{(0)}} \lambda p(\lambda|i) d\lambda,
\]

\[
\hat{x}^{(2)}_{i} = \sum_{i \in I_i(j)} \int_{Q_{i,j}^{(0)}} \lambda p(\lambda|j) d\lambda,
\]

where \( p(\lambda|i,j), p(\lambda|i) \) and \( p(\lambda|j) \) is the pdf of \( \lambda \) inside the quantization cells \( Q_{i,j}^{(0)}, Q_{i,j}^{(1)} \) and \( Q_{i,j}^{(2)} \), respectively.

Construction techniques for the design of \( N(0, 1) \) MDSQ codebooks and the corresponding index assignment tables can be found in [3] and [11].

### 2.2. Transform Coding based on MDSQ

The scalar MDSQ codebooks can be used to construct a multiple description transform coding system (MDSQ-TC). A high-rate analysis for such a system is provided by Batllo and Vaishampayan in [4]. According to this analysis, the optimal bit allocation for resolution-constrained quantization of the multivariate Gaussian case (with diagonal covariance matrix) is:

\[
r_p = \frac{\frac{1}{2} R}{P} + \frac{1}{2} \log_2 \left( \frac{\sigma_p^2}{\prod_{i=1}^{P} \sigma_i^2} \right),
\]

where \( P \) is the number of encoded Gaussians, \( R \) is the total rate assigned to the Gaussians, \( r_p \) is the rate assigned to the \( p \)-th Gaussian random variable, and \( \sigma_p^2 \) is its variance.

### 2.3. GMM-based MDSQ

Sources like the spectral envelopes have non-Gaussian distributions which are usually modeled via GMM. Efficient high-rate quantization for these sources can be made by GMM-based quantization, initially proposed by Subramaniam in [10]. In GMM-based quantization, the source is encoded with a set of transform coders and the “best” (in the mean-square-error sense) encoding is transmitted through the network together with the index of the corresponding encoder. Each of these transform coders is designed according to the statistics of a multivariate Gaussian component of a GMM. This idea is extended to multiple description coding by Samuelsson and Plasberg in [5]. We shall refer to this scheme as GMM-MDSQ-TC.

In GMM-MDSQ-TC, the data vector \( x \in \mathbb{R}^P \) is encoded with \( M \) MDSQ-TC coders and the “best” encoding is transmitted along with the index of the corresponding encoder. Let \( \alpha_{x,m} \) be the prior probability, \( \mu_{x,m} \) the mean and \( \Sigma_{x,m} \) the covariance matrix of the \( m \)-th Gaussian component of the GMM. An eigenvalue decomposition is made to each covariance matrix \( \Sigma_{x,m} \) :

\[
\Sigma_{x,m} = V_{x,m} \Lambda_{x,m} V_{x,m}^T
\]

where the columns of \( V_{x,m} \) are the eigenvectors of \( \Sigma_{x,m} \) and

\[
\Lambda_{x,m} = \text{diag}(\sigma_{m,1}^2, \sigma_{m,2}^2, \ldots, \sigma_{m,P}^2)
\]

is a diagonal matrix with the eigenvalues (variances) \( \sigma_{m,p}^2 \) on its diagonal.

The \( m \)-th MDSQ-TC encoder assumes that the statistics of \( x \) follow the statistics of the \( m \)-th Gaussian component of the GMM, namely \( N(\mu_{x,m}, \Sigma_{x,m}) \). Vector \( x \) is translated and rotated in order to obtain a zero mean vector \( x'_{m} \) with diagonal covariance matrix \( \Lambda_{x,m} \):

\[
x'_{m} = V_{x,m}^T(x - \mu_{x,m}).
\]

The uncorrelated vector \( x'_{m} \) is then quantized with a series of MDSQ encoders to obtain \( M \) sets of central and side reconstructions \( \{\tilde{x}_{0,m}, \tilde{x}'_{1,m}, \tilde{x}'_{2,m}\}, m = 1, \ldots, M \). The encoding that minimizes the distortion function

\[
D_m = ||x'_{m} - \tilde{x}_{0,m}'||^2 + \rho \left( ||x'_{m} - \tilde{x}'_{1,m}||^2 + ||x'_{m} - \tilde{x}'_{2,m}||^2 \right),
\]

where we recall that \( \rho \) is the packet loss probability, is selected for transmission. Note that \( D_m \) is based on the assumption that each description is routed through an independent symmetric channel with loss probability \( \rho \). Furthermore, the distortion is efficiently calculated using \( x'_{m} \) instead of \( x \) because the distortion measure is not affected by \( V_{x,m}^T \), which is a unitary transform. Let \( m^* \) be the encoding that minimizes \( D_m \). Then, the indices of the first and the second description are of the form:

\[
I_1 = \{ m', i_1, i_2, \ldots, i_P \}, \quad I_2 = \{ m', j_1, j_2, \ldots, j_P \},
\]
where \((i_p, j_p)\) refer to the index pairs of the MDSQ quantizer that encode the \(p\)-th dimension of \(x'_m\). The \(m'_\)th MDSQ quantizer decoder performs the inverse operation. If both descriptions are received, the decoder returns the central reconstruction \(\hat{x}'_{k,m'}\). If only one of these descriptions is received, then the corresponding reconstruction \(\hat{x}'_{1,m'}\) or \(\hat{x}'_{2,m'}\) is obtained. Finally, the resulting codevector is obtained by rotating and translating the reconstruction \(\hat{x}'_{k,m'}, k = 0, 1, 2\) according to the equation:

\[
x = \mu_{m'} + V_{m,m'} \hat{x}'_{k,m'}.
\] (12)

Details regarding the bit-allocation for GMM-MDSQ\(_{TC}\) can be found in [5].

3. Bit-Erasur Channel Decoding for GMM-based MDSQ

Assume that description 1 (index \(I_1\)) is received without bit-errors, while description 2 contains bit-errors. The channel decoder provides a set of candidate indices \(\hat{U}_2 = \{I_1, I_2, \ldots, I_{P}\}\) for the second description. The indices \(I_1, I_2, \ldots, I_{P}\) are decoded to the following set of component and scalar quantization indices:

\[
I_1 = \{m'_i, i_1, i_2, \ldots, i_P\},
\] (13)

\[
I_{2,\ldots,P} = \{m'_i, i_1, i_2, \ldots, i_P\}, \quad l = 1, \ldots, P
\] (14)

where \(P\) is the number of dimensions. Only a subset of the candidate indices \(U_2\) valid, in the sense that the pair \((I_1, I_2)\) can be produced from the encoder. Let \(U_2^{'\prime} \subseteq U_2\) be the set of valid candidate indices:

\[
U_2' = \{I_2, \ldots, I_{P}\} \quad m = m' \& (i_p, j_p) \text{ is valid } \forall p \in \{1 \ldots, P\},
\] (15)

where we recall that a pair of indices \((i_p, j_p)\) is considered to be valid if the MDSQ index assignment matrix maps the indices to a central codebook entry. Note that the component index \(m'\) is transmitted twice, in each description. Therefore every valid candidate index must have the same component index. For convenience, let us renumber the elements in \(U_2'\) so that \(U_2' = \{I_2', \ldots, I_{P}'\}\) where \(P' = \) the size of the set \(U_2'\).

Following the nomenclature of Section 2.3, let \(x'_{m'} = [x_{m'_1}, x_{m'_2}, \ldots, x_{m'_P}]^T\) be the \(P\) uncorrelated variables that are encoded with the \(m\)-th MDSQ quantizer, and \(\sigma_{m'_1}, \sigma_{m'_2}, \ldots, \sigma_{m'_P}\) be the corresponding variances. Let \(r_{m'}\) be the number of bits allocated to each of the \(x_{m'_p}\), \(p = 1, \ldots, P\). A set of pre-trained MDSQ codebooks for the \((0, 1)\) Gaussian is obtained using the methods in [3],[11]: one pack of MDSQ codebooks (central and side codebooks) for each loss probability \(p\) and each rate \(r_p\). The GMM-MDSQ\(_{TC}\) encoder/decoder operates with the subset of codebooks that corresponds to the loss probability of the network. Let \(\hat{x}'_{1,m'}(r)\) be the \((i, j)\)-th entry of the central description codebook trained for \(m'(0)\) variables at a side description rate of \(r\) bits, and \(S_{r}(r)\) the length of the corresponding quantization cell. Let \(y_{(l)}\), \(l = 1, \ldots, L'\) be the valid candidate codevectors:

\[
y_{(l)} = \begin{bmatrix}
\sigma_{m'_1} \hat{x}'_{1,m'_1}(r_{m'_1}) \\
\sigma_{m'_2} \hat{x}'_{2,m'_2}(r_{m'_2}) \\
\vdots \\
\sigma_{m'_P} \hat{x}'_{P,m'_P}(r_{m'_P})
\end{bmatrix}.
\] (16)

Let \(Q^{(l)}\) be the \(P\)-dimensional quantization cell associated with codevector \(y_{(l)}\):

\[
Q^{(l)} = Q_{i_1,j_1,1} \times Q_{i_2,j_2,2} \times \cdots \times Q_{i_P,j_P,P},
\] (17)

where \(Q_{i_p,j_p,p}\) is the (scalar) central description quantization cell associated with the codepoint \(\sigma_{m'_1} \hat{x}'_{1,m'_1}(r_{m'_1})\). The cell \(Q_{i_p,j_p,p}\) is an interval of length \(S_{r}(r_{m'_p})\). The optimal MSE (Mean Square Error) reconstruction is given by the following formula:

\[
\hat{x}'_{1,m'}^{(l)} = \sum_{l'=1}^{L'} \left( \frac{\int x'_{m'}(r) p(x'_{m'})dx'}{\int \sum \{ p(x'_{m'})dx\} } \right),
\] (18)

where \(p(x'_{m'})\) is the pdf of the \(m'_p\)-th Gaussian component, namely \(N(0, \sigma_{m'_p})\). A computationally attractive high-rate approximation can be made if we assume that \(p(x; m)\) is approximately constant inside each quantization cell \(Q^{(l)}\), therefore:

\[
\hat{x}'_{1,m'} \approx \sum_{l=1}^{L'} \frac{w_l y_{(l)}}{w},
\] (19)

where \(w_l = \prod_{p=1}^{P} p(\hat{x}'_{i_p,j_p,p}(r_{m'_p})) S_{r}(r_{m'_p})\sigma_{m'_p}\).

The decoded GMM-MDSQ\(_{TC}\) reconstruction can be obtained by rotating and translating \(\hat{x}'_{1,m'}\) according to the statistics of the \(m'_p\)-th Gaussian component of the GMM (see equation (12)):

\[
\hat{x}_1 = \mu_{m'} + V_{m,m'} \hat{x}'_{1,m'}.
\] (20)

The reconstruction formula (19) has a complexity that increases linearly with \(L'\), but the number \(L'\) of valid candidate codevectors increases rapidly with the number of detected bit-errors. Assume that each description is quantized with 30 bits, and that the channel code only detects bit-errors. The number of candidate codevectors when there are \(K\) bit-errors is \(L = \binom{30}{K}\). For \(K = 1 \Rightarrow L = 30\), for \(K = 2 \Rightarrow L = 30 \times 29 = 870\), while for \(K = 3 \Rightarrow L = 30 \times 29 \times 28 = 24360\). The number of valid candidate codevectors is smaller than \(L\): \(L' \leq L\), but the complexity remains overwhelming for more than 2 bit-errors because all candidate codevectors should be checked for validity. A way to tackle the complexity is to split the bits of each description into two parts and use a different error detection code for each part.

The case where both descriptions are damaged with bit-errors can be handled in a similar manner, using equation (19). The only difference in this case is that the candidate indices and the valid indices (those that could originate from the GMM-MDSQ\(_{TC}\) encoder) should now constructed from both descriptions. Such a scheme provides a better reconstruction that the average of the source (the alternative solution when both descriptions are damaged), but even such a reconstruction cannot be considered as good for a speech coding application. However, if additional correlated information is available, the integration in equation (19) can be made using the conditional pdf of the source instead of the source pdf in order to provide an improved reconstruction.
4. Experiments and Results
An evaluation is made using 20-dimensional RCC (Real Cepstrum Coefficient) spectral envelopes derived from 20 ms speech frames. Each frame was analyzed with a harmonic sinusoidal model that estimates the amplitudes and the phases of the harmonics by solving a least squares problem. Harmonic amplitudes were fitted to a 20th order regularized cepstral envelope in Bark scale [9],[12]. These features are suitable for sinusoidal speech coding [9], voice conversion [12] and they are similar to cepstral coefficients used in [13],[14] for speech coding. We trained a GMM with 16 components using 400.000 samples from the training set of the TIMIT database, while the test-set consisted of 10.000 samples from the corresponding test-set of TIMIT. The RCC source was encoded with 60 bits using two balanced 30 bit descriptions. Let $D_0$ be the distortion of the central reconstruction and $D_1, D_2$ the distortions of the corresponding side reconstructions. A number of 1 and 2 bit-errors were introduced to description 2 and the proposed decoder was evaluated using the measured distortions $D_{ber,1}, D_{ber,2}$ respectively. The candidate indices for each vector were computed by introducing 1-2 random bit-errors to the 30 bit description and changing every possible set of two bits; a total of

\[
\begin{align*}
30! & = 30 \text{ candidates for the 1 bit error case and a total of} \\
1 & \text{total of}
\end{align*}
\]

$D_{ber,1}$ and $D_{ber,2}$ correspond to the average distortions of the central and the two side descriptions, respectively. $D_{ber,1}$ and $D_{ber,2}$ correspond to the average distortion when description 2 contains 1 and 2 bit-errors. Distortions are depicted in dB scale.

**MDC quantizer.**

6. References


Figure 1: GMM-MDSQ$_{RC}$ for an 20-dimensional RCC source of speech spectral envelopes. Each description is encoded using 30 bits. $D_0,D_1$ and $D_2$ correspond to the average distortions of the central and the two side descriptions, respectively. $D_{ber,1}$ and $D_{ber,2}$ correspond to the average distortion when description 2 contains 1 and 2 bit-errors. Distortions are depicted in dB scale.