Research Article

Multiple-Description Multistage Vector Quantization

Pradeepa Yahampath

Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada R3T 5V6

Multistage vector quantization (MSVQ) is a technique for low complexity implementation of high-dimensional quantizers, which has found applications within speech, audio, and image coding. In this paper, a multiple-description MSVQ (MD-MSVQ) targeted for communication over packet-loss channels is proposed and investigated. An MD-MSVQ can be viewed as a generalization of a previously reported interleaving-based transmission scheme for multistage quantizers. An algorithm for optimizing the codebooks of an MD-MSVQ for a given packet-loss probability is suggested, and a practical example involving quantization of speech line spectral frequency (LSF) vectors is presented to demonstrate the potential advantage of MD-MSVQ over interleaving-based MSVQ as well as traditional MSVQ based on error concealment at the receiver.

1. INTRODUCTION

Multiple-description (MD) quantization [1, 2] has received considerable attention in recent research due to its potential applications in lossy communication systems such as packet networks. In order to achieve robustness against channel losses, an MD quantizer assigns two or more codewords (to be transmitted on separate packets) for each input sample (or more generally, a vector of parameters representing a frame of samples) in such a manner that the source input can be reconstructed with acceptable quality using any subset of codewords, with the best quality being obtained when the complete set is available. In this paper, we propose an MD *multistage vector quantizer* (MD-MSVQ) and an algorithm for optimizing such a quantizer jointly for a given source and a lossy channel whose loss probability is known. Multistage vector quantization (MSVQ) [3] (also known an *residual vector quantization*) is a computationally efficient technique for realizing high-dimensional vector quantizers (VQs) with good rate-distortion performance and has been considered for many applications, including speech [4, 5], audio [6], and image coding [7]. Given the importance of network-based multimedia applications, it is of considerable interest to study MSVQ in the context of packet-loss channels.

Since an MSVQ generates a set of codewords for each source vector, it naturally provides a means of transporting a given source vector in multiple-packets and thereby achieving some robustness against random packet losses. Motivated by this observation, a previous work [8] considered a particular transmission scheme in which the outputs of different stages in an MSVQ are interleaved in two different packets. It was shown that an MSVQ can be designed to produce lower distortion at a given packet-loss probability, by accounting for interleaving in the optimization of stage codebooks. Based on the experimental results obtained with both speech LSF coding and image coding, [8] concludes that interleaving-optimized MSVQ can yield lower distortion compared to the commonly used approach of repeating the information in last correctly received frame in the event of a packet loss. The goal of this paper is to formulate the problem in the setting of MD quantization, by recognizing the fact that stage interleaving given in [8] is a special case of a more general class of MD quantizers. In MD-MSVQ, each stage consists of a set of multiple description codebooks with an associated *index assignment* (IA) matrix [2]. The interleaving scheme considered in [8] essentially corresponds to an MD-MSVQ in which the IA matrix of the first stage is constrained to be a diagonal matrix, while those of the other stages are constrained to be either a row vector or a column vector. As will be seen, MD-MSVQ designs with more general IA matrices can exhibit a better rate-distortion tradeoff. We present an algorithm for optimizing an MD-MSVQ for a given source (training set) and a set of channel (packet)-loss probabilities. While MD-MSVQ can be applied to any source, the advantage of more general MD-MSVQ over the

interleaving-based scheme is demonstrated here using an example involving 10-dimensional MSVQ of speech LSF vectors based on an input weighted distortion measure. This paper focuses on 2-channel MD-MSVQ; however, the given formulation is applicable to an *n*-channel case as well.

2. MD-MSVQ: STRUCTURE AND OPERATION

A block diagram of a 2-channel, K-stage MD-MSVQ is shown in Figure 1, where the source input $\mathbf{X} \in \mathbb{R}^d$ is d-dimensional random vector. A 2-channel MD-MSVQ is essentially a set of 3 MSVQs, MSVQ_0, MSVQ_1, and MSVQ_2, operating in parallel. However, the three quantizers do not operate independently. Rather, the code vectors of the three quantizers of each stage are linked to form 3-tuples and the encoding is carried out simultaneously using a joint distortion measure. In MD coding terminology, MSVQ_0 is the *central quantizer* and MSVQ_1 and MSVQ_2 are the *side quantizers*. The kth stage of MSVQ_m is a d-dimensional VQ $\mathcal{Q}_m^{(k)}$ with $N_m^{(k)}$ code vectors and the rate $R_m^{(k)} = (1/d)\log_2 N_m^{(k)}$ bits/sample, where $m = 0, 1, 2$ and $k = 1, \ldots, K$. Let $\mathbf{U}_m^{(k)}$ denote the quantization error (residual) of $\mathcal{Q}_m^{(k)}$, $\hat{\mathbf{U}}_m^{(k)}$ the quantized version of $\mathbf{U}_m^{(k)}$, and $\hat{\mathbf{X}}_m^{(k)}$ the reconstructed version of the input $\mathbf{X}$ using the first k stages of MSVQ_m (for the sake of notational consistency, let $\mathbf{U}_m^{(0)} = \mathbf{X}$ and $\hat{\mathbf{U}}_m^{(0)} = \hat{\mathbf{X}}_m^{(1)}$). Then, it is easy to see that

$$\hat{\mathbf{X}}_m^{(k)} = \hat{\mathbf{X}}_m^{(1)} + \sum_{i=1}^{k-1} \hat{\mathbf{U}}_m^{(i)}, \quad \text{for } 1 < k \leq K, \ m = 0, 1, 2, \quad (1)$$

and it follows that the overall quantization error of MSVQ_m, $\mathbf{X} - \hat{\mathbf{X}}_m^{(K)}$, is the quantization error $\mathbf{U}_m^{(K)}$ of the last stage $\mathcal{Q}_m^{(K)}$.

Let the quantization index of $\mathcal{Q}_m^{(k)}$ be $I_m^{(k)} \in \{1, \ldots, N_m^{(k)}\}$. Then, for a given input $\mathbf{X}$, the MD-MSVQ encoder transmits the outputs of MSVQ_1, $T_1 = (I_1^{(1)}, \ldots, I_1^{(K)})$ at the rate of $R_1 = \sum_{k=1}^{K} R_1^{(k)}$ (bits/sample) and those of MSVQ_2, $T_2 = (I_2^{(1)}, \ldots, I_2^{(K)})$ at the rate of $R_2 = \sum_{k=1}^{K} R_2^{(k)}$ over two independent channels (or, if you will, on two separate packets), which can breakdown (or be lost) randomly and independently. The outputs of the central quantizer MSVQ_0 are not transmitted. Instead, each code vector in $\mathcal{Q}_0^{(k)}$ is labeled by a unique pair of code vectors from $\mathcal{Q}_1^{(k)}$ and $\mathcal{Q}_2^{(k)}$ in such a manner that $(I_1^{(k)}, I_2^{(k)})$ uniquely determines $I_0^{(k)}$. Note however that a given code vector in either $\mathcal{Q}_1^{(k)}$ or $\mathcal{Q}_2^{(k)}$ can be associated with more than one code vector in $\mathcal{Q}_0^{(k)}$. The given relation can also be described by an *index assignment* (IA) matrix $\mathbf{A}^{(k)}$ of size $N_1^{(k)} \times N_2^{(k)}$, where $N_0^{(k)} \leq N_1^{(k)} N_2^{(k)}$ [2]. Suppose lth code vector of $\mathcal{Q}_0^{(k)}$ is associated with the ith code vector in $\mathcal{Q}_1^{(k)}$ and the jth code vector in $\mathcal{Q}_2^{(k)}$. Then, (i, j)th element of $\mathbf{A}^{(k)}$ is l. Note that it is possible to have some elements in $\mathbf{A}^{(k)}$ unassigned. These correspond to redundant pairs of codewords $(I_1^{(k)}, I_2^{(k)})$, which are never transmitted simultaneously. The key point here is that if both sets T_1 and T_2 are received by the decoder, then the corresponding set of central quantizer indexes $(I_0^{(1)}, \ldots, I_0^{(K)})$ can be determined

and the receiver can reconstruct the output of MSVQ_0 at the rate $R_1 + R_2$ bits/sample. On the other hand, if only T_1 or T_2 is received, the output of MSVQ_0 cannot be uniquely determined, in which case the receiver can reconstruct exactly the output of either MSVQ_1 (at rate R_1) or MSVQ_2 (at rate R_2). The reconstruction accuracy of the central quantizer and the two side quantizers cannot be chosen independently, and the goal of MD-MSVQ design is to optimize the stage codebooks so as to minimize an average distortion measure. Note that, if neither T_1 nor T_2 is received, then an appropriate loss concealment method has to be employed.

Distortion measure and encoding

Let the distortion caused by quantizing $\mathbf{X}$ into $\hat{\mathbf{X}}$ be measured by $D(\mathbf{X}, \hat{\mathbf{X}})$. Also, denote the *average* distortion of MSVQ_m by $\overline{D}_m \triangleq E\{D(\mathbf{X}, \hat{\mathbf{X}}_m^{(K)})\}$, where, $\overline{D}_0$ is the *central distortion* and $\overline{D}_1$ and $\overline{D}_2$ are the *side distortions* [2]. With the rates (R_1, R_2) fixed, two equivalent formulations are possible for the underlying optimization problem. First, we can minimize $\overline{D}_0$ subject to upper bounds on $\overline{D}_1$ and $\overline{D}_2$. This leads to the minimization of the Lagrangian [2]

$$\mathcal{L} = \overline{D}_0 + \lambda_1 \overline{D}_1 + \lambda_2 \overline{D}_2, \quad (2)$$

where the choice of $\lambda_1, \lambda_2 > 0$ determines the tradeoff between the central distortion and the side distortions. The second formulation is applicable if the probabilities p_1 and p_2 of not receiving T_1 and T_2 at the receiver, respectively, are known (e.g., packet-loss probabilities). In this case, the overall average distortion is given by

$$
\begin{aligned}
&E\{D(\mathbf{X}, \hat{\mathbf{X}})\} \\
&= (1 - p_1)(1 - p_2)\overline{D}_0 + (1 - p_1)p_2\overline{D}_1 \\
&\quad + p_1(1 - p_2)\overline{D}_2 + p_1 p_2 \overline{D}_{\text{ec}} \\
&= (1-p_1)(1-p_2)\left[\overline{D}_0 + \frac{p_2}{(1-p_2)}\overline{D}_1 + \frac{p_1}{(1-p_1)}\overline{D}_2\right] + p_1 p_2 \overline{D}_{\text{ec}},
\end{aligned}
$$
$$(3)$$

where $\overline{D}_{\text{ec}}$ is the average distortion of the error concealment used when both T_1 and T_2 are lost. That is, if we let $\lambda_1 = p_2/(1 - p_2)$ and $\lambda_2 = p_1/(1 - p_1)$, minimizing $\mathcal{L}$ is equivalent to minimizing the overall average distortion $E\{D(\mathbf{X}, \hat{\mathbf{X}})\}$.

The optimal encoding in an MSVQ with K stages involves enumerating through all possible length K sequences of stage codewords to choose the one which yields the minimum distortion reconstruction of a given source vector. This can be achieved by considering the MSVQ encoder as a tree-encoder with a depth K [3], wherein each node of the kth depth level corresponds to a code vector from the kth stage codebook of the MSVQ. Since a full tree-search is impractical, reduced complexity search methods such as M-L algorithm [9] are used in practice to achieve near-optimal encoding. Similar search methods can be employed in MD-MSVQ as well. The only exception in this case is that each node of the kth depth level in the encoding tree now corresponds to a triplet of code
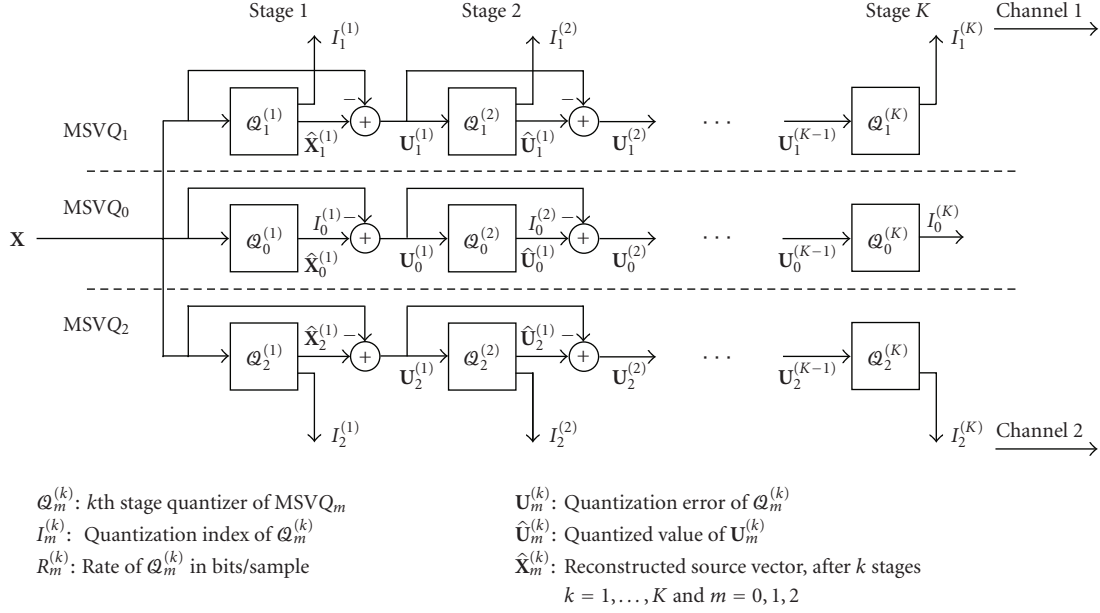
FIGURE 1: The structure of the proposed 2-channel MD-MSVQ encoder with K stages. The outputs of MSVQ$_1$ and MSVQ$_2$ are transmitted over two independent channels (packets). The output of MSVQ$_0$ is not transmitted.

vectors $(\mathbf{c}_0^{(k-1)}, \mathbf{c}_1^{(k-1)}, \mathbf{c}_2^{(k-1)})$ together with an associated path cost

$$
\begin{aligned}
&D^{(k)}\left(\mathcal{U}^{(k-1)}, \hat{\mathbf{c}}^{(k-1)}\right) \\
&= D\left(\mathbf{u}_c^{(k-1)}, \mathbf{c}_0^{(k-1)}\right) + \lambda_1 D\left(\mathbf{u}_1^{(k-1)}, \mathbf{c}_1^{(k-1)}\right) + \lambda_2 D\left(\mathbf{u}_2^{(k-1)}, \mathbf{c}_2^{(k-1)}\right),
\end{aligned}
\tag{4}
$$

where $\mathcal{U}^{(k-1)} \triangleq (\mathbf{u}_0^{(k-1)}, \mathbf{u}_1^{(k-1)}, \mathbf{u}_2^{(k-1)})$ denotes the quantization error triplet of the $(k-1)$th stage (due to the index assignment, it is sufficient to specify $\hat{\mathbf{c}}_0^{(k-1)}$ only which automatically determines the corresponding pair $(\hat{\mathbf{c}}_1^{(k-1)}, \hat{\mathbf{c}}_2^{(k-1)})$). Note that compared to an ordinary MSVQ (which corresponds to $\lambda_1 = \lambda_2 = 0$), the increase in encoding complexity of MD-MSVQ is only due to the use of this modified distortion measure, which is quite marginal.

Relation to stage interleaving

The interleaving scheme studied in [8] can easily be seen as a special case of the above described MD-MSVQ. In that scheme, the quantization indexes $(I_1, \ldots, I_K)$ of a K-stage (single description) MSVQ are divided into two sets, which are transmitted in two separate data packets. One packet carries $(I_1, I_3, I_5, \ldots)$ while the other carries $(I_1, I_2, I_4, \ldots)$. Note that the first-stage index is repeated in both packets, as the subsequent indexes are not meaningful without the first one. With the given packetization scheme, an approximation to the source vector can be obtained by using only the alternate stage indexes in either of the packets. This transmission scheme corresponds to a particular index assignment configuration in MD-MSVQ. Since the first-stage is a repetition code, we set $R_1^{(1)} = R_2^{(2)} = R_0^{(1)}$. In this case, the IA matrix has the size $N_0^{(1)} \times N_0^{(1)}$ and only the diagonal elements are

assigned. Now, in order to account for the transmission of alternate stage outputs on two channels (packets), we choose the stage index assignments to satisfy the following conditions. For even stages, $k = 2, 4, \ldots$, we set $R_1^{(k)} = R_0^{(k)}$ and $R_2^{(k)} = 0$. In this case, the IA matrices are column vectors of size $N_0^{(k)} \times 1$. For odd stages, $k = 3, 5, \ldots$, we set $R_2^{(k)} = R_0^{(k)}$ and $R_1^{(k)} = 0$, which implies that the IA matrices are row vectors of size $1 \times N_0^{(k)}$. The resulting MD-MSVQ is equivalent to stage-interleaving. Since the first stage is a repetition code, this scheme is inefficient when both packets are received (which is the most frequent event in practice). It will be seen that, by using more general IA matrices for all stages (e.g., by dividing the total bit rate of each stage equally between MSVQ$_1$ and MSVQ$_2$), we can achieve a better tradeoff between central and side-distortions, and hence a lower average distortion.

3. DESIGN AND OPTIMIZATION

The design of an MD-MSVQ entails the optimization of three MSVQs: MSVQ$_0$, MSVQ$_1$ and MSVQ$_2$ jointly to minimize (2), subject to constraints imposed by the IA matrices $A^{(k)}$, $k = 1, \ldots, K$. As the distortion measure, we consider the *input weighted square error* of the form [3, Chapter 10]

$$
D(\mathbf{x}, \hat{\mathbf{x}}) = (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{W_x} (\mathbf{x} - \hat{\mathbf{x}}),
\tag{5}
$$

where $\mathbf{W_x}$ is a $d \times d$ symmetric positive definite matrix whose elements are functions of the input vector $\mathbf{x}$ and $(\cdot)^T$ denotes the transpose. In this paper, we propose a codebook design algorithm based on [9], wherein stage codebooks are improved iteratively based on a training set of source vectors, much the same way as in the well-known Lloyd algorithm for ordinary VQ design [3]. In the context of ordinary

MSVQ, two basic approaches have been proposed for codebook optimization [9]: (i) sequential design, and (ii) joint design. In sequential codebook design [9], the kth stage is optimized to minimize the distortion of source reconstruction using up to k stages, assuming that the stages $1, \ldots, k-1$ are fixed, and the codebooks are optimized sequentially from the first stage to the last stage. In this paper, the sequential approach is adapted for MD-MSVQ. According to [9], while the joint method resulted in faster convergence, the final solutions reached by both methods were nearly identical in ordinary MSVQ design.

To start the algorithm, an initial set of stage codebooks and IA matrices are required. In this paper, we have used random initializations for both codebooks and IA matrices. A random IA matrix can be obtained by randomly populating the matrix $\mathbf{A}^{(k)}$ with possible values of $I_0^{(k)}$ such that each element is unique. The codebooks can be initialized by randomly picking vectors from the training set [3]. The initialization is performed sequentially, starting from the 1st stage, so that an input training set is available for every stage. Note that the encoding rule (4) defines simultaneously the quantization cells of all three quantizers of the given stage. In a design iteration, the quantization cells of a given quantizer $\mathcal{Q}_m^{(k)}$ are first estimated for the current codebook, and the codebook optimal for these quantization cells is then computed, as described below. In training set-based design, the quantization cells of a codebook are defined by the subsets of training vectors encoded into each code vector. Note that, once the IA matrices are defined, the codebooks are optimized for fixed IA matrices.

From (1) and (4), it follows that minimizing the total average distortion of the kth stage, given the outputs of the stages $1, \ldots, k-1$, is equivalent to minimizing $E\{D^{(k)}(\mathcal{U}^{(k-1)}, \hat{\mathbf{U}}_0^{(k-1)})\}$. Let $\mathbf{c}_{m,j}^{(k)}$ be the code vector for the quantization cell $\Omega_{m,j}^{(k)}$ of $\mathcal{Q}_m^{(k)}$, where $j = 1, \ldots, N_m^{(k)}$ and $m = 0, 1, 2$. If the IA matrix $\mathbf{A}^{(k)}$ and the quantization cells are fixed, then the optimal value of $\mathbf{c}_{m,j}^{(k)}$ is given by the *generalized centroid* [3, equation 11.2.10]

$$\mathbf{c}_{m,j}^{(k)*} = \arg \min_{\mathbf{c}_{m,j}^{(k)}} E\left\{ D\left(\mathbf{U}_m^{(k-1)}, \mathbf{c}_{m,j}^{(k)}\right) \mid \mathbf{U}_m^{(k-1)} \in \Omega_{m,j}^{(k)} \right\}. \quad (6)$$

For the distortion measure in (5), the expectation in (6) becomes

$$\mathcal{J}\left(\mathbf{c}_{m,j}^{(k)}\right)$$
$$= E\left\{ \left(\mathbf{U}_m^{(k-1)} - \mathbf{c}_{m,j}^{(k)}\right)^T \mathbf{W_x} \left(\mathbf{U}_m^{(k-1)} - \mathbf{c}_{m,j}^{(k)}\right) \mid \mathbf{U}_m^{(k-1)} \in \Omega_{m,j}^{(k)} \right\}. \quad (7)$$

By letting $\nabla_{\mathbf{c}_{m,j}} \mathcal{J}(\mathbf{c}_{m,j}^{(k)}) = \mathbf{0}$, we obtain

$$E\left\{ \mathbf{W_x}\left(\mathbf{U}_m^{(k-1)} - \mathbf{c}_{m,j}^{(k)}\right) \mid \mathbf{U}_m^{(k-1)} \in \Omega_{m,j}^{(k)} \right\} = \mathbf{0}, \quad (8)$$

from which it follows that the optimal code vectors are given by

$$\mathbf{c}_{m,j}^{(k)*} = \left[E\left\{ \mathbf{W_x} \mid \mathbf{U}_m^{(k-1)} \in \Omega_{m,j}^{(k)} \right\} \right]^{-1}$$
$$\cdot E\left\{ \mathbf{W_x} \mathbf{U}_m^{(k-1)} \mid \mathbf{U}_m^{(k-1)} \in \Omega_{m,j}^{(k)} \right\}, \quad (9)$$

for $j = 1, \ldots, N_m^{(k)}$. The code vectors given by this expression can be conveniently estimated using a source training set as follows. In a given design iteration, the source training set is encoded using a tree-search (M-L algorithm) to minimize (4). This is equivalent to computing the quantization cells of each quantizer in the MD-MSVQ, which essentially generates a set of input vectors $\mathcal{T}_m^{(k)}$ for every stage $k = 1, \ldots, K$ of MSVQ$_m$ ($m = 0, 1, 2$), each partitioned into $N_m^{(k)}$ subsets $\mathcal{T}_{m,j}^{(k)}$, $j = 1, \ldots, N_m^{(k)}$, according to the codeword in $\mathcal{Q}_m^{(k)}$ into which those vectors were encoded. Then, the conditional expectations in (9) can be estimated using the weighted sample average computed from $\mathcal{T}_{m,j}^{(k)}$. Note that the weighting matrix $\mathbf{W_X}$ has to be computed from those source training vectors (i.e., inputs to the 1st stage) which produce the subset $\mathcal{T}_{m,j}^{(k)}$ at the kth stage. Once all the stage codebooks have been recomputed, the average distortion of the resulting system is estimated, and the codebook update iterations are repeated until the distortion converges.

4. NUMERICAL RESULTS AND DISCUSSION

In this section, the performance of several MD-MSVQs is evaluated and compared. For this purpose, we consider transmitting 10-dimensional speech LSF vectors over a channel with random packet loses, where the probability of losing any packet is the same. The LSF vectors required for training and testing the codebooks were generated with the Federal Standard MELP coder [10], using the speech samples from the TIMIT database [11] as the input. The designs were carried out using (5) as the distortion measure, with weighing matrix $\mathbf{W_x}$ chosen according to [12, equations (8), (9), (10), and (11)]. On the other hand, in order to objectively evaluate the performance of our LSF quantizer designs, the *frequency weighted spectral distortion* (FWSD) within the frequency band 0–4 kHz, given below, is used [10]:

$$\text{FWSD}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\frac{1}{B_0} \int_0^{4000} |B(f)|^2 \left| 10 \log_{20} \frac{|\hat{A}(f)|}{|A(f)|} \right|^2 df}, \quad (10)$$

where $A(f)$ and $\hat{A}(f)$ are the original and quantized LPC filter polynomials [12] (corresponding to LSF vectors $\mathbf{x}$ and $\hat{\mathbf{x}}$), respectively, $B(f)$ is the *Bark weighting factor* [10], and B_0 is a normalization constant (this distortion measure has been found to closely predict the perceptual quality of reconstructed speech [10]). It is generally accepted that spectral distortion less than 1 dB is inaudible in reconstructed speech [12].

MD-MSVQ systems compared in this paper are summarized in Table 1. In this table, the kth stage of an MD-MSVQ is specified by the triplet $(N_0^{(k)}, N_1^{(k)}, N_2^{(k)})$ where $N_m^{(k)}$, $m = 0, 1, 2$, is the number of code vectors in central and side codebooks. Accordingly, the transmission rates on two MD channels are $R_1^{(k)} = \log_2 N_1^{(k)}$ and $R_2^{(k)} = \log_2 N_2^{(k)}$ bits/vector, respectively. Note that if $N_0^{(k)} = N_1^{(k)} = N_2^{(k)}$, then only the diagonal elements of the IA matrix are used, and consequently, the two transmitted descriptions $\hat{\mathbf{X}}_1^{(k)}$ and $\hat{\mathbf{X}}_2^{(k)}$ will

TABLE 1: MD-MSVQ systems used for comparison. The triplet $(N_0^{(k)}, N_1^{(k)}, N_2^{(k)})$, $m = 0, 1, 2$, for stage k is the number of code vectors in central and side codebooks. R_1 and R_2 are the total rates in bits/vector of $MSVQ_1$ and $MSVQ_2$. R is the total transmission rate per LSF vector.

	Stage 1	Stage 2	Stage 3	Stage 4	R_1	R_2	R
System A	(128,16,8)	(64,8,8)	(64,8,8)	(64,8,8)	13	12	25
System B	(128,128,128)	(64,64,1)	(32,1,32)	—	13	12	25
System C	(32,32,32)	(32,32,1)	(32,1,32)	(32,32,1)	15	10	25
System D	(128,-,-)	(64,-,-)	(64,-,-)	(64,-,-)	—	—	25

TABLE 2: The average frequency-weighted spectral distortion of the systems in Table 1, optimized for different packet-loss probabilities P_L. $\overline{SD}_{central}$ is the central distortion, $\overline{SD}_{side}$ is the side distortion, and $\overline{SD}_{average}$ is the total average distortion.

P_L	$\overline{SD}_{central}$ (dB)				$\overline{SD}_{side}$ (dB)				$\overline{SD}_{average}$ (dB)			
	A	B	C	D	A	B	C	D	A	B	C	D
.001	0.86	1.35	1.23	0.86	7.57	2.49	2.87	8.52	0.87	1.35	1.23	0.86
.005	0.92	1.35	1.23	0.86	5.85	2.47	2.91	8.52	0.97	1.36	1.25	0.90
.01	0.97	1.35	1.23	0.86	4.98	2.44	2.90	8.52	1.05	1.37	1.27	0.95
.05	1.16	1.37	1.29	0.86	3.26	2.36	2.69	8.60	1.37	1.47	1.43	1.30
.1	1.30	1.41	1.38	0.86	2.75	2.29	2.51	8.63	1.60	1.60	1.62	1.64
.15	1.39	1.49	1.43	0.86	2.73	2.22	2.37	8.72	1.76	1.76	1.76	2.04
.20	1.46	1.54	1.48	0.86	2.72	2.14	2.27	8.74	1.92	1.88	1.87	2.44

be identical (i.e., a repetition code). This is the case in the first stage of *System B* and *System C*. Also note that the rest of the stages in these two systems have rate 0 (codebook size of 1) for one of the descriptions. Thus, these two systems are equivalent to stage interleaving MSVQ described in [8]. On the other hand, *System A* uses a general index assignment scheme in which the total rate allocated to each stage is split more evenly between the two MD channels. All three systems have the same total bit-rate as the standard MELP coder [10] (in the 2.4 kbps MELP coder, 54 bits are used for each frame, out of which 25 bits are allocated for the LSF vector). Furthermore, the rate allocation for each stage in *System A* is also the same as in the standard MELP coder. Hence, when optimized for very low packet-loss probabilities, it yields the same distortion as the standard coder. This is not the case with the other two systems. Note also that *System C* has a smaller central codebook for the first-stage compared to *System A*, while having the same number of stages. On the other hand, *System B* has the same central codebook size for the first stage as in *System A*, but at the expense of having only 3 stages. As will be seen below, this results in different central-side distortion tradeoffs. *System D* in Table 1 is a traditional, single description MSVQ with a total rate of 25 bits/vector, used here as a reference for comparison. To deal with the packet losses in this case, we adopt the error-concealment strategy recommended for standard speech codecs such as the 3GPP adaptive multirate (AMR) speech codec [13]. That is, in the event of the loss of nth packet, the current LSF is reconstructed according to $\hat{\mathbf{X}}(n) = \alpha\hat{\mathbf{X}}(n-1) + (1-\alpha)\overline{\mathbf{X}}$, where $\overline{\mathbf{X}}$ is the mean value of the LSF vectors and $\alpha = 0.95$.

The average FWSD of MD-MSVQs optimized for different packet-loss probabilities are shown in Table 2. Several observations are noteworthy. First, the advantage of more general index assignments compared to stage interleaving index

assignments is clear. In particular, *System A* has much lower central distortion at low-loss probabilities, compared to *System B* and *System C*. This is primarily due to the use of repetition codes for the first stage in the latter two systems. Furthermore, in *System A*, the rate of the central quantizer in each stage is determined by the channel-loss probability. That is, at low-loss probabilities all the elements in the IA matrices are assigned to a code vector in the central codebook, that is, the size of the k-stage central codebook is $N_1^{(k)} \times N_2^{(k)}$. Thus, the quantizer is biased towards lowering the central distortion which dominates the average distortion at low-loss probabilities. As the loss probability increases, some of the elements in the IA matrices will be left unassigned and hence the number of code vectors in the central codebook is reduced, that is, the central codebook size becomes less than $N_1^{(k)} \times N_2^{(k)}$. This allows for central distortion to be traded-off for side-distortion to achieve the minimum average distortion for the given loss probability (i.e., $N_0^{(k)} = N_1^{(k)} \times N_2^{(k)}$ shown in Table 1 for *System A* are actually the size of the initial codebook, and the size of the final codebook produced by the design algorithm depends on the channel-loss probability). On the other hand, restricted IA schemes in *System B* and *System C* do not allow the size of the central codebook to vary as a function of the channel-loss probability. Rather, it is only possible to vary the values of the fixed number of code vectors during the optimization. It can be seen that, in comparison to MD systems, the average FWSD of the traditional *System D* is quite poor at higher loss probabilities. The fact that the central distortion of *System D* is independent of the channel probability is obvious, since in this case the quantizer is not adapted to the loss probability. However, in comparison to MD-MSVQ systems, the side distortion of *System D* is quite high. The side distortion in *System D* is due to the error in predicting the current LSF, based on the

TABLE 3: The percentage of decoded frames with FWSD in 2–4 dB range, >4 dB range, and the percentage of frames with FWSD in 2–4 dB range at the output of central decoder ($MSVQ_0$) only.

P_L	2–4 dB				> 4 dB				2–4 dB (central)			
	A	B	C	D	A	B	C	D	A	B	C	D
.001	0.45	7.40	4.41	0.45	0.18	0.03	0.05	0.10	0.43	6.85	4.21	0.43
.005	0.73	7.82	4.84	0.46	0.77	0.08	0.18	0.50	0.50	7.12	4.25	0.43
.01	1.37	8.24	5.43	0.48	1.30	0.14	0.03	1.0	0.72	7.24	4.28	0.43
.05	8.46	12.27	10.48	0.53	2.30	0.62	1.18	4.8	2.29	7.29	4.62	0.43
.1	17.4	17.5	17.90	0.60	2.25	1.23	1.78	9.84	4.84	7.34	6.62	0.43
.15	23.9	23.60	22.72	0.64	2.65	1.88	2.48	14.7	6.54	8.93	7.19	0.43
.20	30.9	27.68	26.93	0.72	3.29	2.64	3.17	19.6	7.88	9.60	7.52	0.43

previously reconstructed LSF (which depends on the correlation between consecutive LSF vectors). As the loss probability increases, the probability of losing two consecutive LSFs increases and so does the prediction error. Hence *System D* exhibits the undesirable property that the side distortion increases with the channel-loss probability.

In addition to the average spectral distortion, another widely used predictor of quality of speech reconstructed from quantized LSFs is the percentage of speech frames having spectral distortion above a certain threshold. Experimental results have shown that such *outlier statistics* of quantized LSF frames have a direct relationship to the perceptual quality of speech [12]. In particular, it has been observed that the distortion in reconstructed speech is inaudible if the average spectral distortion of LSFs is not more than 1 dB, while having less than 2% of speech frames with more than 2 dB spectral distortion and no speech frames with spectral distortion greater than 4 dB [12]. These criteria are used as the basis for comparison in Table 3. It can be observed here that, while the percentage of outlier frames in *System A* is comparatively higher at low-loss probabilities, it becomes comparable to those in *System B* and *System C* as the loss probability increases. This is consistent with the results in Table 2, where *System A* shows a much more pronounced tradeoff between central and side distortions. In order to more clearly demonstrate the advantage of *System A* over the interleaving-based systems, we also list in Table 3 (last four columns) the percentage of frames with FWSD between 2–4 dB at the output of the central decoder (the percentage of frames at the central decoder output with FWSD >4 dB was less than 0.1% in all four systems). It can be noted that, while in all systems most of the outlier frames occur during packet losses, *System A* produces much lower percentage of outlier frames in central decoding compared to *System B* and *System C*. This advantage was evident in the speech output produced by *System A*. This is due to the fact that, even though the intermittent packet losses degrades the output quality of some speech frames, the listening experience appeared to be significantly affected by the output of the central decoder (i.e., transparent quality may be obtained most of the time, accompanied by occasional artifacts during losses). Although the central decoder performance of *System D* is unaffected by the channel quality, the percentage of outlier frames with FWSD greater than 4 dB is substantially higher than in the MD-MSVQ sys-
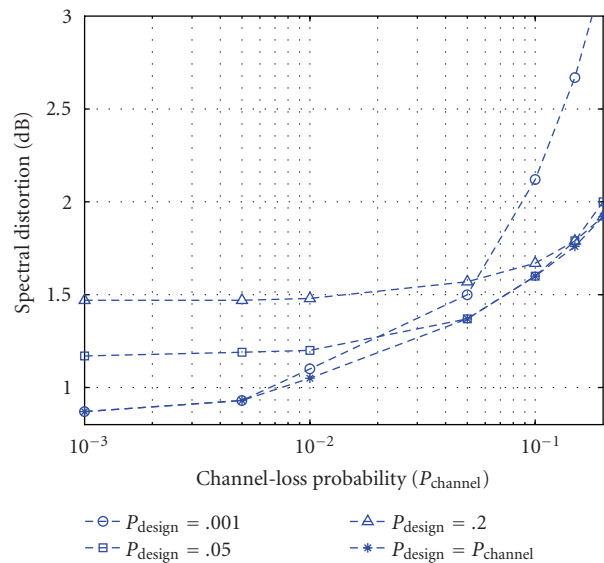

FIGURE 2: The sensitivity of MD-MSVQ (*System A*) to variations in packet-loss probability. P_{design} refers to the channel-loss probability for which the given system was optimized. Note that the system with $P_{design} = P_{channel}$ is optimal for the given channel.

tems. This was also evident in the speech output produced by *System D*, which sounded markedly poor at loss probabilities above 5%. Thus, the advantage of MD-MSVQ over traditional MSVQ with error concealment is clear. It is also worth emphasizing the fact that MD-MSVQ is a generic technique in the sense that it does not rely on correlation between consecutive vectors to deal with channel losses. Indeed, the performance of an MD-MSVQ system can be further enhanced by exploiting the intervector correlation at the receiver (e.g., by appropriately combining MD decoding with prediction-based error concealment).

Since an MD-MSVQ is optimized for a specific channel-loss probability, it is also of importance to investigate the robustness of MD-MSVQ against variations in the loss probability, that is, when the actual-loss probability $P_{channel}$ is different from the design value P_{design}. In Figure 2, we present the average FWSD of 4 different MD-MSVQs with $P_{design} = .001, .05, .2$, and $P_{design} = P_{channel}$, evaluated at loss probabilities ranging from $P_{channel} = .001$ to .2. It can be

concluded that MD-MSVQs are robust against the variations in the channel-loss probability around the design value. Also note that MD-MSVQs optimized for higher-loss probabilities show a relatively small variation in the FWSD over the given range of loss probabilities, compared to the one optimized for a low loss probability (P_{design} = .001). It is thus possible to adapt MD-MSVQ to varying channel conditions and maintain near-optimal performance, by having a number of codebooks optimized to a set of different loss probabilities.

5. CONCLUDING REMARKS

An algorithm for designing MD-MSVQ based on an input-weighted square error to match the channel-loss probability, together with experimental results obtained by transmitting 10-dimensional speech LSF vectors over a random packet-loss channel, has been presented. It has been shown that previously studied stage interleaving-based MSVQ [8] is included in MD-MSVQ as a special case of stage index assignment, and that by choosing more general index assignments, one can achieve a better rate-distortion trade-off. Thus, MD-MSVQ is a potential approach to realizing robust high-dimensional VQ for network-based communication of speech, audio, and image sources. It is also worth pointing out that the given approach may be extended to realize more general tree-structured VQ (TSVQ) [3] in MD form, as MSVQ is a special case of TSVQ.

REFERENCES

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

[2] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, Boston, Mass, USA, 1992.

[4] B.-H. Juang and A. H. Gray Jr., "Multiple stage vector quantization for speech coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '82)*, vol. 7, pp. 597–600, Paris, France, May 1982.

[5] V. Krishnan, D. V. Anderson, and K. K. Truong, "Optimal multistage vector quantization of LPC parameters over noisy channels," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 1, pp. 1–8, 2004.

[6] W.-Y. Chan and Gersho, "High fidelity audio transform coding with vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '90)*, vol. 2, pp. 1109–1112, Albuquerque, NM, USA, 1990.

[7] S. Kossentini, M. T. J. Smith, and C. F. Barnes, "Image coding using entropy-constrained residual vector quantization," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1349–1357, 1995.

[8] H. Khalil and K. Rose, "Multistage vector quantizer optimization for packet networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1870–1879, 2003.

[9] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, pp. 373–385, 1993.

[10] L. M. Supplee, R. P. Cohn, J. S. Collura, and A. V. McCree, "MELP: the new federal standard at 2400 bps," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, vol. 2, pp. 1591–1594, 1997.

[11] National Institute of Standards & Technology (NIST), "The DARPA TIMIT Acoustic Continuous Speech Corpus (CD-ROM)," NIST, 1990.

[12] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3–14, 1993.

[13] 3rd Generation Partnership Project (3GPP), "Adaptive multi-rate (AMR) speech codec; error concealment of lost frames," Technical Specification 3G TS 26.091, 3GPP, Valbonne, France, 1999, www.3gpp.org.