Research Article

Robust Transmission of Speech LSFs Using Hidden Markov Model-Based Multiple Description Index Assignments

Paul Rondeau and Pradeepa Yahampath

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

Correspondence should be addressed to Pradeepa Yahampath, pradeepa@ee.umanitoba.ca

Received 26 July 2007; Revised 5 December 2007; Accepted 21 January 2008

Recommended by Thippur Sreenivas

Speech coding techniques capable of generating encoded representations which are robust against channel losses play an important role in enabling reliable voice communication over packet networks and mobile wireless systems. In this paper, we investigate the use of multiple description index assignments (MDIAs) for loss-tolerant transmission of line spectral frequency (LSF) coefficients, typically generated by state-of-the-art speech coders. We propose a simulated annealing-based approach for optimizing MDIAs for Markov-model-based decoders which exploit inter- and intraframe correlations in LSF coefficients to reconstruct the quantized LSFs from coded bit streams corrupted by channel losses. Experimental results are presented which compare the performance of a number of novel LSF transmission schemes. These results clearly demonstrate that Markov-model-based decoders, when used in conjunction with optimized MDIA, can yield average spectral distortion much lower than that produced by methods such as interleaving/interpolation, commonly used to combat the packet losses.

Copyright © 2008 P. Rondeau and P. Yahampath. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The coding and transmission of speech over lossy channels have received considerable attention due to the widespread deployment of packet networks and wireless mobile systems, for example, [1–4]. In packet networks, data losses occur due to congestion in the network. Furthermore, in systems with wireless links, severe channel fading can render received signals undecodable. The retransmission of lost data packets (commonly used with file transfers) is usually not an option in voice communication, as the resulting delay can be unacceptable in conversations. Therefore, the development of robust source coding techniques, which can ensure acceptable reconstruction of speech based on incomplete data arriving at the receiver, is important. The general framework for analysis and design of such source codes is multiple description (MD) coding [5]. The basic principle of MD coding is to generate $M$ ($\geq 2$) codewords (referred to as descriptions) for a given source input in such a manner that an acceptable reconstruction can be obtained with any subset of $m$ ($\leq M$) codewords. In general, MD coding encompasses many different techniques, ranging from index assignment (IA) for quantizers to MD transforms and filter banks [5]. Most state-of-the-art speech codecs are based on the linear predictive coding (LPC) principle [6] which involves dividing the stream of input speech samples into nonoverlapping blocks or frames, and quantizing and transmitting only a compact set of parameters for each frame. This parameter set, which depends on the specific codec, generally includes several different types of parameters. The focus of this paper is the robust transmission of line spectral frequency (LSF) portion of the parameter set, which represents the short-term power spectrum of the input speech frame [6]. The special characteristics of LSF parameters can be used to quantize them very efficiently, a problem that has been studied quite extensively in previous work, for example, [7, 8], and [6, Chapters 8 and 15]. Given that speech transmission often takes place over noisy channels, the reliable transmission of LSF vectors has also received considerable attention. In this context, techniques ranging from postprocessing-based error concealment to joint source-channel decoding have been investigated; see [4] for a concise review of some recent work in these directions. However, except [4], these works mainly focus on noisy channels with random bit errors and do not consider block errors such as packet losses. In the joint source-channel decoding approach considered
in [4], the residual correlation in the output of a suboptimal quantizer (which all practical quantizers are) is exploited at the receiver to optimally estimate the LSF parameters in the presence of channel noise. In order to do so, the quantizer output is modeled as a first-order Markov process, and the decoding problem is then cast as a minimum mean square error (MMSE) estimation problem in hidden Markov processes (HMPs) [9]. Such Markov-model-based joint source-channel decoders have been previously studied in general in a number of early works, see [10–13]. More recently, this approach has also been studied in the context of MD quantization [14]. While the work in [4] also considers decoding over packet-loss channels in addition to bit-error channels, interleaving-based packetization strategies are employed to introduce transmission diversity. However, interleaving is a restricted case of more general multiple description (MD) IA (MDIA) studied in this paper, and hence is not necessarily the best approach to achieving transmission diversity at the source coding level.

In this paper, we investigate the integration of channel-optimized MDIA techniques into speech LSF quantizers, in order to achieve transmission diversity on lossy channels. MDIA is a simple postprocessing operation applied to the output of a given quantizer and can easily be incorporated into a standard speech encoder without increasing its complexity. Unlike interleaving, MDIA does not necessarily require either a decoding delay or source correlation to achieve transmission diversity. More importantly, MDIA is more flexible and can be optimized to the channel conditions and the specific decoder in place. To our knowledge, MDIA has not been considered for LSF transmission before. While source correlation is not strictly needed, it is shown in this paper that the high correlation among the LSF coefficients can be effectively exploited in MDIA-based transmission in a number of ways. First, the decoders based on optimal estimation can be employed which exploits residual redundancy due to suboptimal quantization to reconstruct the LSFs from partial descriptions received at the decoder when packet losses occur [14]. To this end, we use Markov modeling of the quantizer output to formulate interframe and/or intraframe decoders for MDIA-based transmission, and then develop computationally tractable procedures for optimization of the IA matrices within that framework. The second approach we propose for exploiting the correlation among quantized LSFs in MDIA is to perform IA jointly on the outputs of a bank of quantizers often used with LSF vectors, for example, scalar quantization (SQ) [15] or split vector quantization (split VQ) [6, 16, 17]. This approach, which we refer to as vector IA, is shown to significantly enhance the robustness of transmitted LSFs against channel losses. In order to study the benefit of the proposed MDIA, several LSF transmission schemes have been experimentally evaluated. These experimental results clearly indicate the potential advantage of LSF transmission using optimized MDIA compared to, for example, interleaving-based packetization. In particular, it is found that substantial reductions in average spectral distortion may be achieved in the presence of channel losses by using vector IA optimized to a combined inter-/intraframe decoder. This performance gain can be obtained at a modest increase in computational complexity, mainly at the decoder side. In this paper, we mainly focus on MDIA design for SQ-based systems (e.g., FS-1016 [15]) in which the impact of Markov-model-based decoding is most significant, due to the high residual redundancy in the quantizer output. However, the proposed approach can also be applied to VQ-based transmission as well. We provide one such example.

The rest of the paper is organized as follows. Section 2 provides a description of the problem, while Section 3 presents a brief review of MD quantization and the IA problem, and introduces the notation used in this paper. Section 4 presents several decoding strategies which exploit the correlation in the quantizer output to obtain the best possible estimate of the transmitted LSF coefficients in the presence of channel losses. Subsequently, Section 5 addresses the key issue of designing MDIA matrices for the given decoders, including the vector IA design. Numerous experimental results comparing the performance of several MD transmission methods are presented in Section 6. Finally, some concluding remarks are made in Section 7.

2. Problem Description

In an LPC-based speech coder, each frame of input speech samples (typically of duration 20–30 milliseconds) is first analyzed into a set of parameters, before being quantized and encoded into a bit stream. At the receiver, the decoded parameters are used to synthesize the speech frame. An important part of the analysis parameters is the LSF coefficients representing the short-term power spectrum of the speech frame [7]. Each LSF coefficient is a frequency in the range of (0, Fs/2) (i.e., a positive real value), where Fs is the sampling frequency of input speech. For convenience, let the LSF parameters of the nth speech frame in a sequence be represented by an L-dimensional vector \(X_n = (X_{n,1}, X_{n,2}, \ldots, X_{n,L})\), where \(X_{n,l}\) for \(l = 1, \ldots, L\) is assumed to be a real positive random variable. Typically, most low bit-rate LPC-based speech codecs use \(L = 10\) [6]. An important property of LSF coefficient is the ordering property [7], that is, \(X_{n,1} < X_{n,2} < \ldots < X_{n,L}\). For simplicity, we assume that the quantizer for \(X_n\) is a bank of L scalar quantizers, each operating on one vector element (the approach can be easily generalized to other alternatives such as split-VQ). These quantizers are assumed to be fixed. Our goal is to postprocess the outputs of these quantizers before transmission and to define packetization strategies based on the principle of MDIA for transmitting each LSF vector using two descriptions (data packets). This results in a two-channel MD system (while MD systems with three or more channels can also be constructed, they are of higher complexity and of little use unless the channel loss probability is very high). Note that MDIA-based packetization does not change the reconstruction quality at the receiver if both packets are received. When packets are lost, the decoder performs optimal estimation of the corrupted LSF coefficients, based on the partial descriptions in the received packets. The specific form of the decoder thus depends on the estimation strategy used. In order to measure the distortion between the quantizer
input \( X_{i,n} \) and its reconstructed version \( \hat{X}_{i,n} \), the square error \( (X_{i,n} - \hat{X}_{i,n})^2 \) is used in this paper.

3. MD QUANTIZATION AND IA

Consider the quantization and transmission of a generic real variable \( X_n \in \mathbb{R} \) with the probability density function (pdf) \( f_X(x) \). This variable is assumed to be generated by a stationary and ergodic random process. For notational convenience, define the two index sets \( I_m \equiv \{1, \ldots, m\} \) and \( I_{m,\varnothing} \equiv \{1, \ldots, m, \varnothing\} \) (\( m \) is a positive integer). The additional symbol \( \varnothing \) denotes the channel output in the event of an erasure.

A block diagram of a two-channel MD quantizer is shown in Figure 1. The input sample \( X_n \) is to be quantized and communicated over two independent channels (or packets), supporting the rates \( R_1 \) and \( R_2 \) bits/sample, respectively. To this end, the encoder generates two descriptions \( (I_n^{(1)}, I_n^{(2)}) \) for each input sample \( X_n \), where \( I_n^{(m)} \in I_{N,m} \) and \( N_m = 2^R_m \), \( m = 1, 2 \). It is assumed that each channel has two random states. In one of the states, error-free transmission occurs while in the other state, a complete erasure occurs. The encoder has no knowledge of the state of the channels. On the other hand, the decoder can always identify when an erasure occurs on a channel, in which case it assigns the symbol \( \varnothing \) to the missing channel output. Hence, a two-channel system can be conveniently viewed as a single finite-state channel with 4 states, whose random state \( S \) is unknown to the encoder but known to the decoder. This model applies to communication over data networks with packet losses as well as to wireless communication where fading renders some packets undecodable. In either case, the channel has two states and MD quantization can be used to convert such a channel into a virtual channel with 4 states, in which the probability of a complete loss is much less. Denote the possible states by \( s_1, s_2 \) (no erasures on either channel), \( s_1 \) (only channel 1 produces a valid output), \( s_2 \) (only channel 2 produces a valid output), and \( s' \) (no channel produces a valid output). MD decoding is used to reconstruct the source sample in channel states \( s_1, s_2 \), and \( s_2 \), based on the received codeword(s). In the state \( s' \), the decoder has to use a suitable error concealment technique. Let the output of the two channels at time \( n \) be denoted by the index pair \( V_n = (I_n^{(1)}, I_n^{(2)}) \), where \( I_n^{(m)} \in I_{m,\varnothing} \) (recall that \( \varnothing \) denotes an erasure). The decoder \( D \) produces an approximation \( \hat{X}_n \) to the source sample \( X_n \), based on \( V_n \). The decoding problem is addressed in Section 4. Given the source pdf and the channel state probabilities, the fundamental problem in MD quantizer design is to choose the encoder and decoder to minimize the average distortion between the source sample \( X_n \) and its reconstruction \( \hat{X}_n \).

The MD encoder can be decomposed into two stages as shown in Figure 1. The input quantizer assigns an index \( U_n \in I_N \) to each input sample \( X_n \), based on a partition of the support region of \( X_n \) into \( N \) nonoverlapping intervals. In the subsequent index assignment (IA), the index \( U_n \) is mapped to an index pair by a many-to-one mapping \( y(U_n) = (I_n^{(1)}, I_n^{(2)}) \), where \( y : \{0, 1, \ldots, N-1\} \rightarrow \{0, 1, \ldots, N-1\} \) and \( N_1, N_2 \geq N \). Note that the total transmission rate is determined by \( N_1 \) and \( N_2 \) and not by the quantizer resolution \( N \). Given the rate constraints \( R_1 \) and \( R_2 \) and the channel state probabilities, an optimal MD quantization system minimizes the mean square error (MSE): \[
D = E\{(X_n - \hat{X}_n)^2\} \\
= D_{1,2} P(s_{1,2}) + D_1 P(s_1) + D_2 P(s_2) + D_{E} P(s'),
\]
where \( D_r = E\{(X_n - \hat{X}_n)^2 \mid S = s\} \) is the conditional MSE given the channel state \( s \) and \( P(s) \) is the probability that channel state is \( s \). The average distortion of reconstruction based on both descriptions, \( D_{1,2} \), is referred to as the central distortion whereas \( D_1 \) and \( D_2 \) are referred to as side-distortions. Note that central distortion is only a function of the input quantizer \( \epsilon \) and is independent of \( \gamma \). For a fixed \( \epsilon \), the side-distortions \( D_1 \) and \( D_2 \) are determined by \( \gamma \) and the decoder \( \delta \). Hence, we focus on the design of index assignment \( \gamma \) and the decoder \( \delta \) for a fixed quantizer, that is, we seek to find \( \gamma \) and \( \delta \) which minimize the average side distortion \( D_1 P(s_1) + D_2 P(s_2) \). Note that \( D_r \) is independent of the MD quantizer.

For design, it is convenient to view the IA as an assignment of each of the possible \( N \) values of the quantizer index \( U_n \) to a unique element of an \( N_1 \times N_2 \) matrix \( A \), where the row number \( l_n^{(1)} \) and the column number \( l_n^{(2)} \) represent the two codewords (descriptions) transmitted over the MD channels for the given quantizer index \( [18] \). More specifically, if \( (A_{1,ij} = k \), then \( \gamma(k) = (i, j) \). Alternatively, we can specify two many-to-one functions:

\[
y^{(1)}(k) = i, \quad y^{(2)}(k) = j,
\]
which give the row and column of the quantizer index \( k \) in \( A \). Note that it is not necessary to have all elements in the IA matrix assigned to an input quantizer index. Such elements represent output codeword pairs \( (i, j) \) which are never transmitted simultaneously. Given an input quantizer with the resolution \( N \), and a total transmission rate of \( R_1 + R_2 \) bits/sample (where \( N \leq 2^R \)), the underlying optimization problem is to choose the IA matrix \( A \) to minimize the average side-distortion. This essentially is a combinatorial optimization problem with \( (N_1 N_2)!/(N_1 N_2 - N)! \) possible solutions. Clearly, an exhaustive search for the optimal solution is impractical in all but trivial situations. Previously, reduced-complexity search algorithms such as the binary switching algorithm \([19] \), simulated annealing \([20] \), and deterministic annealing \([21] \) have been used to obtain good IA designs. In \([18] \), a less general approach is proposed in which asymptotic quantization theory is used to obtain structured IA for Gaussian sources.

**Notation**

Any random variable associated with the \( l \)th LSF in the \( n \)th LSF vector (frame) in a time sequence of LSF vectors is denoted by \( Y_{l,n} \), where \( l \in \{1, \ldots, L\} \) and \( n \in \{0,1,\ldots\} \).
4. DECODING STRATEGIES

4.1. Memoryless decoder

First, consider the simplest decoding strategy of reconstructing $X_{l,n}$ on the basis of the channel output $V_{l,n}$. The MMSE source estimate is given by [22]

$$x_l^*(v) = E[X_{l,n} | V_{l,n} = v] = \sum_{u=1}^{N_l} g_{l,u} P(U_{l,n} = u | V_{l,n} = v),$$  

(3)

where $g_{l,u} = E[X_{l,n} | U_{l,n} = u]$ is the centroid [23, Chapter 6] of the $n$th quantization interval of the quantizer for the $l$th LSF, $u \in \mathbb{R}_{N_l}$. Though simple, this decoder does not use all the information available at the output of the channel for estimating the unknown $X_{l,n}$. As mentioned earlier, the suboptimal quantization of a correlated sequence of random variables typically results in a correlated sequence of quantizer indexes. Furthermore, the ordering property of LSF coefficients within a vector, coupled with fact that the consecutive LSF vectors represent a speech utterance implies that there is a high correlation between the elements of a given LSF vector, as well as between the consecutive LSF vectors [11]. When suboptimal quantization of LSFs is employed, much of this correlation remains in the quantized variables. Therefore, the decoder should examine an entire sequence of transmitted indexes in order to obtain the best possible estimate of the quantizer input. Following [14], we present below several sequence-based generalizations of (3).

4.2. Interframe decoder

In order to benefit from the temporal (interframe) correlation in the encoder output, we let the decoder output $\hat{X}_{l,n}$ for the $l$th LSF $X_{l,n}$ of the $n$th frame to be a function of the complete observation sequence $V_{l}^{(n+n_0)}$, up to a delay of $n_0$ frames. Accordingly, the optimal decoder is obtained by minimizing

$$E\{[X_{l,n} - \hat{X}_{l,n}(V_{l})_0^{(n+n_0)}]^2 | V_{l}^{(n+n_0)}\}. $$  

(4)

It follows that the optimal decoder is [22]

$$\hat{X}_{l,n}(V_{l})_0^{(n+n_0)} = E\{X_{l,n} | V_{l}^{(n+n_0)}\} = \sum_{u_{l_1}^{(n+n_0)}} E\{X_{l,n} | u_{l_1}^{(n+n_0)}\} P(u_{l_1}^{(n+n_0)} | V_{l}^{(n+n_0)}),$$  

(5)

where the sum is now over all possible ($N_l$)$^{n+n_0}$ quantizer index sequences $u_{l_1}^{(n+n_0)}$ and the channels are assumed to be memoryless. Note that the term $E\{X_{l,n} | u_{l_1}^{(n+n_0)}\}$ represents the centroids of a partition of the amplitude range of $X_{l,n}$ with respect to length $n+n_0$ sequences of quantizer output indexes. It is neither easy to compute these centroids nor possible to deal with ($N_l$)$^{n+n_0}$ such codewords. However, since $X_{l,n}$ depends much more strongly on $U_{l,n}$ than on $U_{l,n-1}^{(n+n_0)}$, it is reasonable to assume that $E\{X_{l,n} | u_{l_1}^{(n+n_0)}\} \approx E\{X_{l,n} | u_{l_1}^{(n+n_0)} = u\} = g_{l,u}$ (in practice, considering other quantization indices on the right-hand side of the expectation changes the value only slightly). Then, we obtain the following simple generalization of the memoryless decoder in (3):

$$\hat{X}_{l,n}(V_{l})_0^{(n+n_0)} = \sum_{u=1}^{N_l} g_{l,u} P(U_{l,n} = u | V_{l})_0^{(n+n_0)}. $$  

(6)

Note that if $\{U_{l,n}\}$ is an iid sequence, the above decoder reduces to the memoryless decoder. Thus, the MSE of the decoder in (6) is bounded above by that of the memoryless decoder in (3). As we will see, when $\{U_{l,n}\}$ is correlated, the MSE obtained with (6) can be significantly less than that obtained with (3). Next, we focus on computing the posterior probabilities $P(U_{l,n} = u | V_{l})_0^{(n+n_0)}$ for $u = 1, \ldots, N_l$. In order to exploit the statistical dependence between successive outputs from the quantizer and to obtain a computationally tractable solution, we assume $\{U_{l,n}\}$ to be a first-order Markov process with respect to $n$, for each $l = 1, \ldots, L$, that is, we assume $P(U_{l,n} | U_{l,n-1}, U_{l,n-2}, \ldots) = P(U_{l,n} | U_{l,n-1})$. This is a reasonable assumption, since two consecutive LSFs are much more strongly correlated than the others (which will also be reflected in the quantized LSFs). With this assumption, the decoder input $V_{l}^{(n+n_0)}$ is an observation sequence from the HMP $\{U_{l,n}, n = 1, 2, \ldots\}$ whose posterior state probabilities (given by (6)) can be conveniently computed using the well-known forward-backward
algorithm \([9]\). To this end, consider that 
\[
P(u_n | v^{l,n}_{1,1}) = \frac{P(u_n, v^{l,n}_{1,1})}{P(v^{l,n}_{1,1})} = \frac{P(v^{l,n}_{1,1})u_n P(u_{n+1}, v^{l,n}_{1,1})}{P(v^{l,n}_{1,1})},
\]
where \(P(v^{l,n}_{1,1}) = \sum_{u_n} P(v^{l,n}_{1,1} | u_n)P(u_n, v^{l,n}_{1,1})\) and we use the fact that given the current state, the future outputs of a Markov process are independent of the previous outputs. Now, defining 
\[
\begin{align*}
\alpha_{l,n}(u_n) & \triangleq P(u_n, v^{l,n}_{1,1}), \\
\beta_{l,n}(u_n) & \triangleq P(v^{l,n}_{1,1} | u_n),
\end{align*}
\]
equation (7) can be expressed as 
\[
P(u_n | v^{l,n}_{1,1}) = \frac{\alpha_{l,n}(u_n)\beta_{l,n}(u_n)}{\sum_{u_n} \alpha_{l,n}(u_n)\beta_{l,n}(u_n)}. 
\]
(9)
For each \(n\), the two quantities \(\alpha_{l,n}(u_n)\) and \(\beta_{l,n}(u_n)\), \(u_n \in \mathbb{L}_N\), can be computed using the two recursive relations \([9]\): 
\[
\alpha_{l,n}(u_n) = P(v^{l,n}_{1,1}) \sum_{u_{n-1}} P(v^{l,n}_{1,1} | u_{n-1}) \alpha_{l-1,n}(u_{n-1}) 
\]
(10)
for \(k = 1, \ldots, n + n_0\) and 
\[
\beta_{l,n}(u_n) = \sum_{u_{n+1}} P(v^{l,n}_{1,1} | u_{n+1}) \beta_{l-1,n}(u_{n+1}) 
\]
(11)
for \(k = n + n_0 - 1, \ldots, 1\), where \(P(v^{l,n}_{1,1} | u_n)\) is the mapping used in the encoder. For example, if \(u_n = k\) and \(v^{l,n}_{1,1} = (i, \emptyset)\), then 
\[
P(v^{l,n}_{1,1} | u_n) = I\{\gamma_j^{l}(k) = i\}.
\]

### 4.3. Intraframe decoder

The intraframe decoder does not make use of the correlation between the quantizer indexes of a single LSF vector. In order to utilize such “intraframe” correlation, we can model the dependence between two adjacent elements \(U_{l-1,n}\) and \(U_{l,n}\) in the \(n\)th frame by a first-order Markov process as well, that is, we assume 
\[P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n}) = P(U_{l,n} | U_{l-1,n}),\]
for \(l = 2, \ldots, L\), which suggests that the observations constitute a length \(n + n_0\) sequence of \(L\)-dimensional index vectors. However, given the \(l\)th LSF in consecutive speech frames and the other LSFs in the frame \(n\), \(U_{l,n}\) is weakly dependent on the residual LSFs in the other frames. By taking this observation into account, we can relate (17) to (7) and (13) as follows:

\[
P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n}) 
\approx P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n}) 
\approx P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n}) 
\approx P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n}) 
\approx P(U_{l,n} | U_{l-1,n}, \ldots, U_{l,n})
\]

in which the initial values can be chosen as \(\alpha_{l,n}(u) = P(\mathcal{V}_{l,n} \mid U_{l,n} = u)\) for all \(u \in \mathbb{L}_N\) and 
\[
\beta_{l,n}(u) = 1 \text{ for all } u \in \mathbb{L}_N.
\]

#### 4.4. Combined inter-/intraframe decoder

Clearly, the best possible source estimate is obtained by considering both interframe and intraframe correlations simultaneously. That is, the MMSE estimate for the \(l\)th LSF of frame \(n\) must be computed as 
\[
\hat{x}_{l,n} = \sum_{u_n} g_{l,n} P(U_{l,n} = u \mid v^{l,n}_{1,1}, \ldots, v^{l,n}_{1,L}) 
\]
(17)

Obviously, it is practically difficult to compute the conditional probabilities in this way, where the observations constitute a length \(n + n_0\) sequence of \(L\)-dimensional index vectors. However, given the \(l\)th LSF in consecutive speech frames and the other LSFs in the frame \(n\), \(U_{l,n}\) is weakly dependent on the rest of the LSFs in the other frames.
Importantly, this approximation results in decomposing a vector Markov process into two independent scalar Markov processes, as shown in Figure 2. Consequently, \( P(U_{l,n} = u \mid v_{l,1}^{(n+m)}) \) and \( P(U_{l,n} = u \mid v_{l,1}^n) \) can be computed independently, using (9)–(11) and (13)–(16), respectively. It may be noted that this decoder is a simplified version of the MS4 decoder proposed in [4]. The motivation for the particular derivation presented above is that, unlike the formulation in [4], the product form in (18) readily lends itself to IA optimization in a simple manner, as shown below.

### 5. IA DESIGN

#### 5.1. IA optimization by simulated annealing

In [20], a simulated annealing- (SA-) based algorithm has been proposed for obtaining good IA matrices for MD quantization. In essence, the basic approach is to search for the IA matrix which minimizes the average side-distortion. In essence, the basic approach is to search for the IA matrix which minimizes the average side-distortion in (19), we let

\[
\alpha_{l} | \beta_{l} = \frac{1}{U_{l,n}} \sum_{v_{l,1}^n} g(t) \alpha_{l}(t | k) \beta_{l}(t | k), \quad k = 1, \ldots, N_{l},
\]

where \( \alpha_{l}(t | k) \) and \( \beta_{l}(t | k) \) are the IA functions described in (2), and \( P(U = k) \) is the probability that the input quantizer index \( U = k \). The IA design problem is to minimize (19) with respect to the IA matrix, or equivalently \( \alpha_{l}(t | k) \) and \( \beta_{l}(t | k) \). This integer optimization problem can be solved by SA, using (19) as the cost function [20]. The algorithm is summarized in the appendix. More details can be found in [20].

When memory-based encoders, such as the interframe decoder, are involved, the reconstructed output \( \hat{X}_n \) and hence the side-distortion depends on a sequence of transmitted indexes. Specifically, the memory-based decoders presented in Sections 4.2 and 4.3 involve computing the recursive terms in (8) and (14), and in effect have time varying codebooks, implying that the optimal IA matrix will have to be adaptive. A more tractable alternative is to find a fixed IA matrix which is good on the average. One approach to determining such IA matrices would be to consider the average values (averaged over observation sequences) of the time-varying codebooks and then to use the resulting fixed codewords in (19) for IA optimization. However, computing average codebooks in this manner is computationally difficult within the SA-based IA optimization algorithm. Alternatively, we can view the decoder output as a function of time-dependent state variables \( \alpha(\cdot) \) and \( \beta(\cdot) \) given by (9) and (13). Then, we can compute fixed codebooks based on the average values of these state variables which are much easier to obtain by exploiting the recursive expressions used to compute them. We follow this basic idea.

#### 5.2. IA optimization for interframe decoder

Suppose that the quantizer output is \( U_{l,n} = k \) and only a single description is received at the decoder. If the channel state is \( S_n = s_j \), then the decoder receives \( y_{l,1}^{(j)}(k) \) as the channel output. For the purpose of determining the IA matrix for the \( l \)th LSF-based on minimizing side-distortion in (19), we let

\[
C_{l}(y_{l,1}^{(j)}(k)) = \frac{1}{U_{l,n}} \sum_{t=1}^{N_{l}} g(t) \alpha_{l}(t | k) \beta_{l}(t | k), \quad k = 1, \ldots, N_{l},
\]

where \( \alpha_{l}(t | k) \) and \( \beta_{l}(t | k) \) are the conditional averages of the state variables in (10) and (11), respectively, with respect to the channel output sequences \( V_{l,1}^{(n+m)} \). Clearly, these expressions are not easily evaluated. A good approximation is to average over \( V_{l,1}^{(n+1)} \) instead. While this is feasible, we found that it is sufficient to find the above averages for the case of an isolated loss, that...
is, while frame $n$ suffers a loss, the frames $n - 1$ and $n + 1$ are received correctly. For this case, we have

$$\tilde{\alpha}_{l,n}^{(j)}(r | k) = E[\alpha_{l,n}(r) | U_{l,n} = k, S_n = s_j, S_{n-1} = s_{s_j}],$$

$$\tilde{\beta}_{l,n}^{(j)}(r | k) = E[\beta_{l,n}(r) | U_{l,n} = k, S_n = s_j, S_{n+1} = s_{s_j}].$$  

(22)

First consider

$$\tilde{\alpha}_{l,n}^{(j)}(r | k) = \begin{cases} \sum \psi_i(r, t) E[\alpha_{l,n-1}(r) | U_{l,n} = k, S_{n-1} = s_{s_j}] & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}, \end{cases}$$

(23)

where the expectation is taken with respect to $P(U_{l,n-1} | U_{l,n} = k, S_{n-1} = s_{s_j})$. Now, since

$$E[\alpha_{l,n-1}(r) | U_{l,n} = k, S_{n-1} = s_{s_j}, S_n = s_j] = P(U_{l,n-1} = r | U_{l,n} = k),$$

(24)

it follows that

$$\tilde{\alpha}_{l,n}^{(j)}(r | k) = \begin{cases} \sum \psi_i(r, t) \psi_i(r, k) P(U_{l,n-1} = r) / P(U_{l,n} = k) & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}. \end{cases}$$

(25)

Next consider that

$$\tilde{\beta}_{l,n}^{(j)}(r | k) = \begin{cases} \sum \psi_i(t, r) E[\beta_{l,n+1}(r) | U_{l,n} = k, S_{n+1} = s_{s_j}] & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}, \end{cases}$$

(26)

where the expectation is taken with respect to $P(V_{l,n+1} | U_n = k, S_{n+1} = s_{s_j})$. Since

$$E[\beta_{l,n+1}(r) | U_{l,n} = k, S_{n+1} = s_{s_j}] = P(U_{l,n+1} = r | U_{l,n} = k),$$

(27)

it follows that

$$\tilde{\beta}_{l,n}^{(j)}(r | k) = \begin{cases} \sum \psi_i(t, r) \psi_i(k, r) & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}. \end{cases}$$

(28)

Note that with the above formulation, the IA matrix for each of the $L$ LSFs in a frame can be optimized separately.

### 5.3. IA optimization for intraframe decoder

Similar to the interframe decoder, we now have to average the corresponding decoder state variables in (15) and (16) with respect to $V_{l,n}^{L,n}$. However, this is not computationally feasible for $L = 10$. Given the Markov assumption on the sequence $U_{l,n}^{L,n}$, it is reasonable to use the approximation $P(U_{l,n} = u | V_{l,n}^{L,n}) \approx P(U_{l,n} = u | V_{l,n}^{L,n})$, or equivalently, to approximate (15) and (16) for the $l$th LSF as

$$\alpha_{l,n}^{(j)}(u, n) \approx P(u, n, V_{l,n}^{L,n}),$$

$$\beta_{l,n}^{(j)}(u, n) \approx P(V_{l+1,n} | u, n).$$

(29)

Then, as in the case of the interframe decoder, we can determine the averages of the decoder state variables as

$$\tilde{\alpha}_{l,n}^{(j)}(t | k) = E[\alpha_{l,n}^{(j)}(t) | U_{l,n} = k, S_n = s_j],$$

$$\tilde{\beta}_{l,n}^{(j)}(t | k) = E[\beta_{l,n}^{(j)}(t) | U_{l,n} = k, S_n = s_j],$$

(30)

where the expectations are taken with respect to $P(V_{l+1,n}^{L,n} | U_{l,n} = k, S_n = s_j)$. First, consider

$$\tilde{\alpha}_{l,n}^{(j)}(t | k) = \begin{cases} \sum \psi_i(r, t) E[\alpha_{l,n-1}(r) | U_{l,n} = k, S_n = s_j] & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}, \end{cases}$$

(31)

where

$$E[\alpha_{l,n-1}(r) | U_{l,n} = k, S_n = s_j] = \sum_i P(U_{l-1,n} = r | V_{l+1,n} = y_i^{(j)}(i)) P(U_{l-1,n} = i | U_{l,n} = k).$$

(32)

Then, given the IA matrix for the $(l-1)$th LSF, $P(U_{l-1,n} = i | U_{l,n} = k) = \phi_i(k, k) P(U_{l-1,n} = i) / P(U_{l,n} = k)$ and $P(U_{l-1,n} = r | V_{l-1,n} = y_i^{(j)}(i))$ can be computed for different values of $i$ and $r$.

Next consider

$$\tilde{\beta}_{l,n}^{(j)}(t | k) = \begin{cases} \sum \psi_i(t, r) E[\beta_{l,n+1}(r) | U_{l,n} = k, S_{n+1} = s_j] & \text{if } y_i^{(j)}(k) = y_i^{(j)}(t), \\ 0 & \text{otherwise}, \end{cases}$$

(33)

where

$$P(V_{l+1,n} = y_i^{(j)}(r) | U_{l,n} = t) = \sum_i \phi_i(t, t) I[\gamma_i^{(j)}(r) = y_i^{(j)}(i)].$$

(34)

The latter probabilities can be obtained given the IA matrix of $(l+1)$th LSF. Since computing $\tilde{\alpha}_{l,n}^{(j)}(t | k)$ and $\tilde{\beta}_{l,n}^{(j)}(t | k)$ for the $l$th LSF requires the knowledge of the IA matrices of $(l-1)$th and $(l+1)$th LSF, respectively, we optimize the IA matrices of all $L$ LSFs simultaneously, that is, within the same SA algorithm (see the appendix).
5.4. IA optimization for combined decoder

Our formulation of the combined decoder allows us to compute the required codebooks in a simple manner. That is, given the average values of the state variables of the inter- and intraframe decoders obtained above, the expressions in (18) and (17) can be evaluated to find the codebooks \( C_i \) and \( C_{ii} \) to be used in (19). Then, the SA algorithm can be used as above.

5.5. Vector IA design

A practical speech encoder typically consists of a bank of quantizers, either scalar [15] or low-dimensional VQ (e.g., split VQ [8]). In such situations, there will still be a correlation between the outputs of individual quantizers which can be effectively exploited in MDIA to improve the reconstruction accuracy under packet losses. One approach to reaping this benefit is to perform IA jointly on a vector of quantizer indices, by assigning each possible vector of quantizer indices to a unique element in an IA matrix. Evidently, the complexity of the underlying IA design problem increases exponentially with the vector size. Hence, we restrict ourselves to joint IA design for pairs of quantizers. Without a loss of generality, we consider, in the following, the case of scalar quantizers (the formulation readily extends to VQ as well).

Let \( X_1 \) and \( X_2 \) denote the input samples (e.g., two LSFs) of a pair of scalar quantizers whose outputs are \( U_1 \in \mathbb{L}_N \) and \( U_2 \in \mathbb{L}_M \), respectively. In joint IA, one assigns the index pair \( (U_1, U_2) \) to another index pair \( (I^{(1)}, I^{(2)}) \), where \( I^{(m)} \in \mathbb{L}_{N_m} \) and \( N_m = 2^{R_m}, \) \( m = 1, 2 \). The indexes (or descriptions) \( I^{(1)} \) and \( I^{(2)} \) are then transmitted over two independent channels at rates \( R_1 \) and \( R_2 \) bits per sample, respectively. This mapping can be described by an assignment of each value of \((U_1, U_2)\) to a unique element of an \( N_1 \times N_2 \) matrix \( A \), where \( N_1 N_2 \geq NM \).

Similar to (2), this can also be described by a pair of many-to-one functions
\[
y^{(1)}(k, l) = i, \quad y^{(2)}(k, l) = j,
\]
which specify the quantizer index pair \((k, l)\) assigned to the matrix element \( A_{i,j} \). Now, if both \( I^{(1)} \) and \( I^{(2)} \) are received, \( U_1 \) and \( U_2 \) can be uniquely determined at the decoder. In this case the decoded outputs would be \( \hat{X}_1 = G_i(u_i) \) and \( \hat{X}_2 = G_j(u_j) \), where \( G_i(u_i) \) is the \( u_i \)th codeword in the codebook of \( i \)th quantizer. However, if only \( I^{(1)} \) or \( I^{(2)} \) is received, the pair \((U_1, U_2)\) cannot be uniquely determined in which case the codeword \( \hat{X}_i = E[X_i | I^{(m)} = j] \) is used as the output, \( i = 1, 2 \) (these codebooks can be easily estimated using a training set of source samples). For convenience, let \( C^{(m)}(j) = E[X_i | I^{(m)} = j], \) \( i = 1, 2, m = 1, 2, \) and \( j = 1, \ldots, N_m \). Then the total average side-distortion of the quantizer pair is given by
\[
D_{\text{side}}(y^{(1)}, y^{(2)}) = D_{\text{side}_1}(y^{(1)}, y^{(2)}) + D_{\text{side}_2}(y^{(1)}, y^{(2))},
\]
where
\[
D_{\text{side}_1}(y^{(1)}, y^{(2)}) = \sum_{k=1}^{N} \sum_{l=1}^{M} \left[ P(s_1)(C^{(1)}(y^{(1)}(k, l)) - 2G_i(k)C^{(1)}(y^{(1)}(k, l)))^2 + P(s_2)(C^{(2)}(y^{(2)}(k, l)) - 2G_j(k)C^{(2)}(y^{(2)}(k, l)))^2 \right]
\]
\times P(U_1 = k, U_2 = l)
\]
\[
D_{\text{side}_2}(y^{(1)}, y^{(2)}) = \sum_{k=1}^{N} \sum_{l=1}^{M} \left[ P(s_1)(C^{(1)}(y^{(1)}(k, l)) - 2G_i(k)C^{(1)}(y^{(1)}(k, l)))^2 + P(s_2)(C^{(2)}(y^{(2)}(k, l)) - 2G_j(k)C^{(2)}(y^{(2)}(k, l)))^2 \right]
\times P(U_1 = k, U_2 = l)
\]
\[
\text{is the side-distortion of the } i\text{th quantizer, } i = 1, 2, \text{ and } P(U_1 = k, U_2 = l) \text{ is the joint probabilities of the outputs.}
from the two quantizers. The joint IA matrix of the two quantizers can be optimized by minimizing (36) using the IA optimization algorithm given in the appendix.

It is worth noting here that MD coding using pairwise correlating transforms (PCT) [24] is also a vector IA method applicable to uncorrelated Gaussian sources and scalar quantizers. In PCT, vector IA is achieved by an integer-to-integer linear transform of a quantizer index pair. If the two sources being quantized are correlated, it is assumed that a decorrelating transform precedes the PCT. In contrast, the above described vector IA optimization approach makes no assumptions about the source distributions (i.e., works for non-Gaussian and correlated sources), and can be applied to VQ outputs.

6. NUMERICAL RESULTS AND DISCUSSION

This section presents experimental results to demonstrate the benefits of the proposed LSF transmission schemes based on MDIA. The experiments have been performed on a simulated packet-loss channel in which the probability of loss is the same for all packets. In order to design the IA matrices and to estimate the quantizer-output transition probabilities required by various decoders, a training set of LSF vectors has been used. The performance of the resulting systems has then been evaluated using a separate test set of LSF vectors. Both training and test sets were generated with the 4.8 kbps FS-1016 CELP [15] codec using the speech samples from the TIMIT database [25]. In the FS-1016 codec, a speech frame of 30 milliseconds is analyzed into a set of 17 parameters, which includes 10 LSF parameters (or a 10-dimensional LSF vector). The resulting training set consisted of 407 806 LSF vectors while the test set consisted of 172 061 vectors.

In the FS-1016 codec, an LSF vector is transmitted using 34 bits, by quantizing each of the 10 LSFs using a separate scalar quantizer. The bit allocation among the 10 LSFs is shown in the first row of Table 1. In our experiments with MD transmission schemes, we retain the same quantizers and the total transmission rate, but apply index as segments to split the output bit stream of each quantizer into two streams of 17 bits/vector each, which are packetized and transmitted separately (however, in the case vector IA, LSFs are packetized in pairs). Thus, all systems considered below produce the same reconstruction as the FS-1016 codec when no packet losses occur, that is, produces identical central distortions. The transmission systems compared in this section are based on the MDIA schemes described in Section 5 and the decoding strategies described in Section 4. Table 1 also shows the bit allocations (between two descriptions) used in various MDIA schemes. Note that interleaving of odd and even numbered LSF coefficients in a vector is a special case of MDIA where the IA matrix is either a column vector or a row vector.

In order to objectively measure the performance of the LSF transmission schemes, we use the log spectral distortion (SD) [8] between the input speech frame and its reconstructed version in the frequency range \((F_1, F_2)\), given by

\[
SD = \sqrt{\int_{F_1}^{F_2} \left(10 \log_{10} P(f) - 10 \log_{10} \hat{P}(f)\right)^2 df},
\]

where \(P(f)\) and \(\hat{P}(f)\) are the power spectra computed using the input LSF vector \(X\) and the reconstructed LSF vector \(X\), respectively [8, equations (6), (7)]. Following [8], we use \(F_1 = 0\) and \(F_2 = 3\) kHz as the perceptually significant range of frequencies. In lossless transmission, the average SD of the above-described 34 bits/frame SQ is 1.56 dB. We estimate the average side-distortion \(SD_{side}\) (i.e., the average distortion of LSFs reconstructed using only one of the descriptions) in two different ways. In the first method, we consider isolated losses, in which case the estimated value of side-distortion \(SD_{side}\) is independent of the channel loss probability. This enables us to compare the different transmission schemes, without being affected by simultaneous loss of both descriptions (the output for which cannot be handled

---

### Table 3: The advantage of IA optimized for inter-frame-decoding (estimates based on isolated loss events). A delay of one frame is used in all three schemes.

| IA Scheme                        | Decoding method      | \(SD_{ide}\) (dB) | Frames (%) |
|----------------------------------|----------------------|-------------------|------------|
| Interleaving                     | Interframe Markov decoder | 2.80              | 63.03      | 13.23 |
| Optimized for memoryless decoding| Interframe Markov decoder | 2.94              | 60.76      | 16.60 |
| Optimized for interframe decoding| Interframe Markov decoder | 3.00              | 53.95      | 20.70 |

### Table 4: The performance of intraframe decoding with different IA schemes (estimates based on isolated loss events).

| IA Scheme                              | \(SD_{ide}\) (dB) | Frames (%) |
|----------------------------------------|-------------------|------------|
| Interleaving                           | 3.17              | 61.72      | 22.86      |
| Scalar, optimized for memoryless decoding | 3.43              | 61.89      | 27.11      |
| Scalar, optimized for intraframe decoding | 3.08              | 64.57      | 19.46      |
| Vector, optimized for intraframe decoding | 2.80              | 65.87      | 12.67      |
by MD coding). Since the output of decoders with memory depends on the previous and subsequent packet losses, we also estimate the average side-distortion by considering random losses which can result in one or more consecutive losses as well as the loss of both descriptions. In this case, the estimated side-distortion depends on the channel loss probability (even for memoryless decoders). In all experiments, when both descriptions are simultaneously lost, the previously received LSF vector is repeated. Note that we do not report the overall average SD given by (1), as the central distortion is not a function of the channel-losses probability. This allows us to compare the given transmission schemes on the basis of their performance during packet losses.

First, we compare the performance of two simple systems in order to motivate the more complex IA schemes and Markov decoders. The first system transmits odd and even numbered LSF coefficients in each frame using two different packets (referred to as interleaving) and a decoding scheme in which the missing LSFs in a given frame are filled by the corresponding LSFs in the previously received frame (repetition). The second system uses the memoryless MD decoder in (3) with an optimized IA matrix. Table 2 shows the average side-distortion of these two systems. Also shown in the table are the percentages of reconstructed speech frames with spectral distortion exceeding 2 dB and 4 dB, respectively. These percentages of “outlier” speech-frames are also important, since the outlier frames usually result in audible distortion, even when the average SD is low [8]. It can be seen that the performance of the simpler system with interleaving- and repetition-based reconstruction is superior to that of the optimized but memoryless MD system (the former system essentially has a memory of one frame). The reason is that the latter scheme does not make use of the high correlation between consecutive LSF vectors. Rather, it only uses the correlation between the two descriptions of an LSF to reconstruct that LSF.

Next, we investigate the benefit of using Markov-model-based interframe decoding (Section 4.2) which exploits the correlation between the values of a given LSF in consecutive frames. In this case, we compare in Table 3 the performance of interframe decoders with that of a system which uses interleaving at the transmitter and optimal linear interpolation of missing LSF values at the receiver (i.e., to estimate the missing LSF values in the frame $n$, the corresponding LSFs in frames $n-1$ and $n+1$ are used). Since this requires a decoding delay of one frame, we allow the same delay in the Markov decoders as well. Note that the system with interframe decoding and IA optimized to that decoder yields a noticeable reduction in side-distortion, which appears to be mainly due to a reduction in the outlier frames with average SD more than 4 dB during packet losses. This example also shows the advantage of IA optimization for the interframe decoder. The interframe decoder, when used with an IA optimized for memoryless decoding, does not perform much better than the system based on interleaving/interpolation. As mentioned earlier, the average side-distortion of interframe decoders can be influenced by packet-loss patterns (e.g., occurrence of several consecutive losses). Therefore, in Figure 3, we show the average side-distortion as a function of channel loss probability, estimated by transmission of long sequences of LSF vectors. Several observations are noteworthy. First, note that the system with optimized IA and the interframe Markov decoder maintains a clear advantage over other methods at all loss probabilities shown, both with and without a decoding delay. However, there is a significant improvement to be gained by allowing a decoding delay of one frame. An exception is the interleaving scheme with linear interpolation (curve (e)), whose performance degrades rapidly at high loss probabilities. This shows the advantage of Markov-model-based decoding over simple interpolation in the presence of consecutive losses.

The intraframe Markov decoder presented in Section 4.3 exploits the correlation between LSF values within a given frame. The performance of several systems based on intraframe Markov decoding is presented in Table 4. Note that the system with optimized vector IA yields the best performance in this case. The results in Tables 3 and 4 motivate us to consider combined inter-/intraframe Markov decoding, which can be expected to yield a substantial improvement in performance during packet losses. Table 5 presents the performance of combined decoders based on several MDIA schemes and no decoding delay, while Table 6 presents the performance of the same systems, when a decoding delay of one frame is allowed. Also, Figure 4 shows the average side-distortion of these schemes as a function of the channel loss probability. The benefit of optimizing the IA matrix for the decoder in use is obvious. In particular, the systems with vector IA optimized to a combined inter-/intraframe decoder achieve about 1.0 dB reduction in SD compared to the commonly used transmission schemes such as interleaving/interpolation (e.g., Table 3). In this case, we fully exploit the correlations among the quantizer outputs (both inter- and intraframe), both at the encoder and decoder. It is of

![Figure 3: The comparison of average side-distortion of different LSF transmission schemes based on interframe decoding, under random packet losses: (a) optimized IA+memoryless MD decoding (no delay), (b) interleaving+repetition (no delay), (c) interleaving+interframe Markov decoding (no delay), (d) optimized IA+interframe Markov decoding (no delay), (e) interleaving+interpolation (delay = 1 frame), (f) interleaving+interframe Markov decoding (delay = 1 frame), and (g) optimized IA + interframe Markov decoding (delay = 1 frame).]
Table 5: The performance of combined inter-/intraframe decoding with different IA schemes and zero decoding delay (estimates based on isolated loss events).

| IA Scheme                                      | $SD_{side}$ (dB) | 2–4 dB | > 4 dB |
|-----------------------------------------------|------------------|--------|--------|
| Interleaving                                   | 2.74             | 66.73  | 10.10  |
| Scalar, optimized for interframe decoding      | 2.89             | 65.10  | 14.22  |
| Scalar, optimized for combined decoding        | 2.66             | 61.92  | 9.81   |
| Vector, optimized for combined decoding        | 2.47             | 53.89  | 7.12   |

Table 6: The performance of combined inter-/intraframe decoding with different IA schemes and a decoding delay of one frame (estimates based on isolated loss events).

| IA Scheme                                      | $SD_{side}$ (dB) | 2–4 dB | > 4 dB |
|-----------------------------------------------|------------------|--------|--------|
| Interleaving                                   | 2.46             | 62.33  | 5.13   |
| Scalar, optimized for interframe decoding      | 2.54             | 60.53  | 7.84   |
| Scalar, optimized for combined decoding        | 2.37             | 56.84  | 5.25   |
| Vector, optimized for combined decoding        | 2.19             | 47.10  | 3.70   |

Figure 4: The comparison of average side-distortion of several LSF transmission schemes based on combiner inter-/intraframe decoding, under random packet losses: (a) interleaving, (b) optimized scalar IA (no delay), (c) optimized vector IA (no delay), (d) optimized scalar IA (delay = 1 frame), (e) optimized vector IA (delay = 1 frame).

course obvious that interleaving is a suboptimal approach to IA. Note also that the use of optimized IA adds little to the complexity of the encoder (added complexity is of course in the design stage only). While the Markov decoders presented in this paper have a higher complexity than a simple interpolating decoder, this complexity is quite manageable and can be justified by the resulting performance improvements.

While we have considered above the application of MDIA to scalar quantized LSFs, the approach can also be applied to encoders which use VQ. In order to do so, we considered split VQ [23] of 10-dimensional LSF vectors using a bank of 5 two-dimensional VQs, each having a rate of 5 bits/vector (a total rate of 25 bits/frame). In loss-free transmission, the average SD of this quantizer was 1.47 dB, which is better than the 34 bits/frame scalar quantizer (SQ) considered above. For MDIA-based transmission of the split VQ outputs, we considered two scenarios. In the first one, 25 bits/frame were used for MDIA, so that the transmission bit rate is the same as the output bit rate of the split VQ. In the second scenario, the output of each VQ is allocated 6 bits, equally divided among the two MD channels, yielding a total transmission rate of 30 bits/frame. In both cases, we use combined inter-/intraframe decoding (with one-frame delay) and optimized scalar IA. Table 7 compares the average side-distortion of these two split VQ systems and that of the best-performing 34 bits/frame, vector IA-based SQ scheme from Table 6. All three systems use combined inter-/intraframe decoding. It can be seen that split VQ at 25 bits/frame performs poorly compared to SQ at 34 bits/frame, even though in loss-free transmission the split-VQ yields a lower SD. This is because the residual redundancy in the split VQ output is comparatively less and consequently the effectiveness of Markov decoding during a packet-loss is reduced. However, with MDIA at 30 bits/frame, the split VQ performs comparably with the vector IA-based SQ. Note that in this case, MDIA adds an excess rate (or a redundancy) of 5 bits/frame to reduce the side-distortion.

7. CONCLUDING REMARKS

The robust transmission of speech LSF coefficients over lossy channels using MDIA optimized to Markov-model-based decoders, which exploit the residual correlation in quantized LSF frames for channel loss concealment, has been investigated. An analytical framework for obtaining good IA matrices for intra- and interframe decoders has been proposed and a simulated annealing-based IA optimization algorithm has been presented. Experimental results have shown that Markov-model-based decoders with optimized IA as proposed in this paper can achieve substantial reductions in spectral distortion of LSF vectors reconstructed in the presence of packet-losses, compared to commonly used
interleaving-based schemes. In particular, a combined inter-/intraframe decoder based on scalar quantization of LSF vectors and optimized vector IA has been found to be very effective in our comparisons.

While we have mainly focused on the application of MDIA to scalar quantization, in principle the approach can also be applied to VQ. In this paper, we have considered an example involving 2-dimensional split VQ of 10-dimensional LSF vectors. Possible future work includes the application of MDIA to other VQ schemes, such as predictive VQ used in GSM-AMR codec [16] and IS-641 codec [17]. Interleaving-based transmission for this type of quantizers has been previously considered in [4].

APPENDIX

BASIC IA OPTIMIZATION ALGORITHM

The SA-based MDIA optimization algorithm in [20] is an extension of the noisy channel IA optimization algorithm described in [26]. While [20] considered scalar IA design, the algorithm given below can be used for both scalar and vector IA optimization, by using the appropriate cost function (distortion measure).

(1) Set initial temperature \( T = T_0 \) (a high value); select an initial IA matrix \( A = A_0 \).

(2) Randomly perturb \( A \) to \( A' \) and evaluate the change \( \Delta J \) in the cost function in either (19) or (36).

(3) If \( \Delta J < 0 \), replace \( A \) by \( A' \); if \( \Delta J \geq 0 \), replace \( A \) by \( A' \) with probability \( \exp(-\Delta J/T) \).

(4) If the number of iterations, which resulted in a cost decrease, exceeds a prescribed limit \( k_i \) or if too many iterations \( (k_m) \) have been performed without any cost decrease go to Step (5) (otherwise go to Step (2)).

(5) Lower the temperature as \( aT \rightarrow T \); if \( T \) is below some prescribed freezing temperature \( T_f \) or if the current solution appears stable, terminate the algorithm (otherwise go to Step (2)).

The space of candidate solutions in our problem consists of all possible instances of the IA matrix \( A \). Thus, the perturbation of a given solution \( A \) to \( A' \) can be achieved by interchanging two randomly chosen elements of \( A \). The rate of convergence of the algorithm depends on the choice of values chosen for parameters \( T_0, T_f, a, k_i, \) and \( k_m \).

Note

In the case of IA optimization for the intraframe decoder, all \( L \) IA matrices must be optimized jointly, due to the dependence of cost function of a given LSF on the IA matrices of the adjacent LSFs (see Section 5.3). In this case, random perturbations are applied to all \( L \) IA matrices simultaneously and SA iterations are carried out until all \( L \) IA matrices converge.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their insightful comments and suggestion which help improve the quality of this paper.

REFERENCES

[1] J. Suzuki and M. Taka, “Missing packet recovery techniques for low-bit-rate coded speech,” IEEE Journal on Selected Areas in Communications, vol. 7, no. 5, pp. 707–717, 1989.

[2] B. W. Wah and D. Lin, “LSP-based multiple-description coding for real-time low bit-rate voice over IP,” IEEE Transactions on Multimedia, vol. 7, no. 1, pp. 167–178, 2005.

[3] J. Samuelsson and J. H. Plasberg, “Multiple description coding based on Gaussian mixture models,” IEEE Signal Processing Letters, vol. 12, no. 6, pp. 449–452, 2005.

[4] F. Lahouti and A. K. Kandani, “Soft reconstruction of speech in the presence of noise and packet loss,” IEEE Transactions on Speech and Audio Processing, vol. 15, no. 1, pp. 44–56, 2007.

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

[6] W. C. Chu, Speech Coding Algorithms, Wiley-Interscience, New York, NY, USA, 2003.

[7] F. Soong and B. Juang, “Line spectrum pair (LSP) and speech data compression,” in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’84), vol. 9, pp. 37–40, San Diego, Calif, USA, March 1984.

[8] 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.

[9] L. R. Rabiner and B. Juang, “An introduction to hidden Markov models,” IEEE Acoustics, Speech & Signal Processing Magazine, vol. 3, no. 1, part 1, pp. 4–16, 1986.

[10] N. C. Phamdo and N. Farvardin, “Optimal detection of discrete Markov sources over discrete memoryless channels-applications to combined source-channel coding,” IEEE Transactions on Information Theory, vol. 40, no. 1, pp. 186–193, 1994.

[11] F. I. Alajaji, N. C. Phamdo, and T. J. Fuja, “Channel codes that exploit the residual redundancy in CELP-encoded speech,”

---

Table 7: A comparison of split VQ and SQ systems with MDIA and combined inter-/intraframe decoding (all systems use a decoding delay of one frame).

| Transmission rate | IA Scheme | \( SD_{\text{trc}} \) (dB) | Frames (%) |
|-------------------|-----------|-----------------|-----------|
| SQ                | Vector    | 2.19            | 47.10     |
| split VQ 25 bits/frame | Scalar    | 2.58            | 58.70     | 9.10     |
| split VQ 30 bits/frame | Scalar    | 2.19            | 51.80     | 2.60     |

Note: In the case of IA optimization for the intraframe decoder, all \( L \) IA matrices must be optimized jointly, due to the dependence of cost function of a given LSF on the IA matrices of the adjacent LSFs (see Section 5.3). In this case, random perturbations are applied to all \( L \) IA matrices simultaneously and SA iterations are carried out until all \( L \) IA matrices converge.
IEEE Transactions on Speech and Audio Processing, vol. 4, no. 5, pp. 335–336, 1966.

[12] N. C. Phamdo, A. Alajaji, and N. Farvardin, “Quantization of memoryless and Gauss-Markov sources over binary Markov channels,” IEEE Transactions on Communications, vol. 45, no. 6, pp. 668–675, 1997.

[13] D. Miller and M. Park, “A sequence-based approximate MMSE decoder for source coding over noisy channels using discrete hidden Markov models,” IEEE Transactions on Communications, vol. 46, no. 2, pp. 222–231, 1998.

[14] P. Yahampath, “A recursive side-decoder for multiple description quantization,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’05), vol. 3, pp. 493–496, Philadelphia, Pa, USA, March 2005.

[15] J. P. Campbell Jr., T. E. Tremain, and V. C. Welch, “The federal standard 1016 4800 bps CELP voice coder,” Digital Signal Processing, vol. 1, no. 3, pp. 145–155, 1991.

[16] 3rd Generation Partnership Project, 3GPP TS 26.090 Adaptive multi-rate (AMR) speech codec: Transcoding functions, ver 5.0.0 (2002-06).

[17] “TDMA Cellular/PCS- radio interface, enhanced full-rate voice codec (ACELP),” ANSI/ITA/EIA-136-Rev.c part 410, June 2001.

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

[19] N. Görtz and P. Leelapornchai, “Optimization of the index assignment for multiple description vector quantizers,” IEEE Transactions on Communications, vol. 51, no. 3, pp. 336–340, 2003.

[20] P. Yahampath, “On index assignment and the design of multiple description quantizers,” in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’04), vol. 4, pp. 597–600, Montreal, Quebec, Canada, May 2004.

[21] P. Kouligi, S. L. Regunathan, and K. Rose, “Multiple description quantization by deterministic annealing,” IEEE Transactions on Information Theory, vol. 49, no. 8, pp. 2067–2075, 2003.

[22] H. V. Poor, An Introduction to Signal Detection and Estimation, Springer, Berlin, Germany, 2nd edition, 1994.

[23] A. Gersho and R. M. Gray, Vector Quantization and Signal Compression, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992.

[24] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, “Multiple description coding using pairwise correlating transforms,” IEEE Transactions on Image Processing, vol. 10, no. 3, pp. 351–366, 2001.

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

[26] N. Farvardin, “A study of vector quantization for noisy channels,” IEEE Transactions on Information Theory, vol. 36, no. 4, pp. 799–809, 1990.