

# Soft Reconstruction of Speech in the Presence of Noise and Packet Loss

Farshad Lahouti, Amir K. Khandani

## Abstract

Exploiting the residual redundancy in a source coder output stream during the decoding process has been proven to be a bandwidth efficient way to combat the noisy channel degradations. In this paper, we consider soft reconstruction of speech spectrum, in GSM Adaptive Multi-Rate and IS-641 vocoders, transmitted over a channel disturbed with noise and/or packet loss. Several schemes are presented which exploit different levels of intraframe and interframe residual redundancy for improved source decoding at the receiver. A packetization strategy is proposed which is matched to the presented error concealment units. For decoders that exploit the residual redundancy, extensive complexity has been a serious concern, especially as the quantizer bitrate increases [14][15][26]. In this work, a novel method is presented to construct reduced complexity algorithms. The proposed methodology is based on the classification of the signal domain and efficient approximation of the residual redundancy or the a priori transition probabilities. The presented schemes provide high quality error concealment solutions for CELP coders.

## Keywords

Joint source channel coding, residual redundancies, source decoding, MMSE estimation, forward backward recursion, Markov models, speech coding, LPC, LSF, GSM-AMR, IS-641, packet loss concealment, erasure channel, multiple description coding, speech error concealment.

This work is funded in part by the Natural Sciences and Engineering Research Council of Canada and Communications and Information Technology Ontario. This work has been presented in part at the IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT, USA, 2001, and the 22nd Biennial Symposium on Communications, Kingston, ON, Canada, 2004. Farshad Lahouti is with the Department of Electrical & Computer Engineering, University of Tehran, 14395-515 Tehran, Iran. Amir K. Khandani is with the Coding & Signal Transmission Lab., Dept. of E&CE, University of Waterloo, Waterloo, ON, N2L 3G1, Canada, Email: lahouti@ieee.org, khandani@cst.uwaterloo.ca.

## I. INTRODUCTION

Voice communication over wireless or wireline packet networks is exposed to adverse levels of noise and packet loss. Speech communication with acceptable quality is facilitated through careful design of communication systems at different layers. Our focus in this work is to provide an efficient error concealment scheme and a matched packetization strategy, for effective reconstruction of speech in presence of noise and/or packet loss.

Recently joint source channel (JSC) coding techniques that exploit the residual redundancy [2] in the source coder output stream have found increasing attention [2]-[25]. One of the reasons for such interest is the fact that, these techniques provide improved signal protection against channel errors, with no additional bandwidth requirement, using only the redundancy left due to suboptimal source coding. Researchers have suggested to employ the residual redundancy for improving the performance of channel coders e.g., [3]-[7] or designing effective source decoders (error concealment units) e.g., [8]-[21]. Also, iterative schemes that exploit these redundancies during both channel and source decoding processes have been proposed [22]-[24]. In general, the problem is formulated in the form of a *Maximum A Posteriori* detection or a *Minimum Mean Squared Error* estimation.

For design of source decoders, JSC-based techniques that exploit the residual redundancy replace the conventionally heuristic approaches of error concealment with a formulation within the formal framework of estimation theory. There are two challenges, however, in using such techniques to combat the effect of noise and packet loss in speech communications applications. First, although these techniques provide effective solutions, but in general, they could be complex for practical applications as underlined e.g., in [14][15][26]. The issue of complexity becomes more severe for quantizers with higher rates, as the complexity grows exponentially with quantizer bitrate [26]. Secondly, with the exception of [28] which considers a packet erasure channel, the prior art in this area have so far concentrated on designing joint source channel coding schemes in presence of channel degradations at the bit level. However, in some important applications, the transmitted speech signal is exposed to both noise and packet loss. Examples of particular interest include the communications over wireless packet networks; and the scenario when the communication involves both a wireless and a (packet-based) wireline link. As described below, solutions to these two challenges are proposed in this article.

In the context of methods for combating packet loss in speech transmission, researchers have

taken several different approaches which can be categorized as follows:

(i) One category is comprised of methods that perform Packet Loss Concealment (PLC) at the receiver. Among them, waveform substitution techniques replace the missing segments of speech (in the time domain) with estimates constructed from previous or future available speech segments [29]-[33]. Recently, a waveform substitution algorithm [34] based on pitch detection has been proposed for G.711 PCM speech coding standard [35]. PLC methods for CELP-based coders often replace the missing parameters with the corresponding parameters of the previous frame [36] and use scaled down gains [37]. Similar error concealment techniques are used in the recently proposed standard for GSM-Adaptive Multi Rate (AMR) codec [38] and the North American standard IS-641 [40]. Methods that use both previous and future speech frames for packet loss concealment in CELP based coders are reported in [41][42].

(ii) Another category of PLC algorithms operate with certain coordination with the transmitter side. Methods based on Forward Error Correction, transmit redundant repair data to recover lost packets [44]-[48]. Diversity schemes based on Multiple Description Coding (MDC) also fall into the same category [49]-[51]. Both FEC-based and MDC-based techniques impose additional delay; however, as opposed to the former, the latter schemes may not necessarily require additional bandwidth.

(iii) A third category includes the approaches that also require certain support from the network. In [52] and [53], the speech coder output is transmitted using packets with different priorities and the scheme relies on the network to drop the packets with low priorities during congestion periods. Currently, such a support may only exist in a proprietary network.

In this work, we present error concealment solutions for efficient reconstruction of speech in presence of noise and/or packet loss. The suggested techniques exploit the redundancy in the source coder output stream in a joint source channel coding framework. A novel scheme is proposed which substantially reduces the complexity of the algorithms. This scheme is based on an approximation of the a priori transition probabilities. We also present a packetization strategy based on the Multiple Description Coding concepts which is matched to the suggested error concealment schemes.

Our particular attention is on the recent standard GSM-AMR speech codec and the IS-641 codec. We focus on the reconstruction of the Linear Predictive Coding (LPC) coefficients. These coefficients represent the short-term spectral information of speech within a frame and preserving

them plays a major role in the quality of the reconstructed speech. In the IS-641 codec, the LPC coefficients are quantized in the Line Spectral Frequency (LSF) representation [54] using a Split-VQ [55]. The same scheme is also used in all but one of the eight modes of the GSM-AMR codec. The presented techniques exploit different levels of intraframe and interframe residual redundancy and numerical results demonstrate that they provide effective solutions for reconstruction of speech spectrum parameters in the presence of noise and/or packet loss. Our comparisons show that substantial gains compared to the error concealment method of the standard is achieved and the level of gain is directly related to the amount of residual redundancy exploited.

The rest of this article is organized as follows. The notations and system model are described in section II. Section III presents the error concealment techniques. The method to approximate the a priori transition probabilities and discussion on how it leads to the complexity reduced algorithms are presented in section IV. Section V presents the packetization strategy. Numerical results and comparisons are included in section VI. Finally, concluding remarks are provided in the last section.

## II. PRELIMINARIES

### A. Notations

The notations used in this article are as follows. The capital letters, e.g.,  $I$ , represent random variables, while the small letters, e.g.,  $i$ , represent a realization. We replace the probability  $P(I = i)$  by  $P(I)$  in most instances when it does not lead to a confusion. The vectors are shown bold faced, e.g.,  $\mathbf{X}$ . The lower index indicates the time instant, e.g.,  $\mathbf{X}_n$  is the vector  $\mathbf{X}$  at time instant  $n$ . The upper index in parenthesis indicates components of a vector or bit positions representing an integer value, e.g.,  $\mathbf{X}_n = [X_n^{(1)}, \dots, X_n^{(N)}]$  where  $N$  is the dimension of the vector  $\mathbf{X}_n$ . A sequence of variables over time, e.g.,  $(I_{n_1}, \dots, I_{n_2})$ ,  $n_1 \leq n_2$  is denoted by  $\underline{I}_{n_1}^{n_2}$ . For simplicity we represent  $\underline{I}_n^1$  by  $\underline{I}_n$ . The  $N$  dimensional Cartesian product of a set  $\mathcal{J}$  is represented by  $\mathcal{J}^N$  that consists of  $N$  dimensional vectors whose components are taken from  $\mathcal{J}$ .

### B. System Overview

The block diagram of the system is shown in Figure 1. The source coder is a mapping from an  $N$ -dimensional Euclidean space,  $\mathcal{R}^N$ , into a finite index set  $\mathcal{J}$  of  $M$  elements. It is composed of two components: the quantizer  $Q$  and the index generator  $\mathcal{I}$ . The quantizer maps the input sample  $\mathbf{X} \in \mathcal{R}^N$  to one of the reconstruction points or *codewords* in the codebook  $\mathcal{C}$ ,  $\mathcal{C} \subset \mathcal{R}^N$ .



Fig. 1. Overview of the system

The index generator then maps this codeword to an *index (symbol)*  $I$  in the index set  $\mathcal{J}$ . The bitrate of the quantizer  $r$  is given by  $\lceil \log_2 M \rceil$  bits/symbol (or  $\lceil \log_2 M \rceil / N$  bits/dim).

At the receiver, for each transmitted  $r$ -bit index  $I = i$ , a vector  $J$  with  $r$  components is received which provides information about  $I$ . The reconstructor (source decoder) maps  $J$  to an output sample  $\hat{\mathbf{X}}$ . In this reconstruction, the source decoder may use the previously received signals or some of the future samples as well.

### C. Channel Model

The channel model described here includes both the effects of noise and packet loss. It is an approximate model for communications over wireless packet networks; or communications that involve both a wireless and a (packet-based) wireline link. Similar configurations are used in [57], [58], [59] and [60] to provide an end-to-end performance consideration.

A data packet transmitted through the channel may be randomly lost (e.g. due to network congestion), otherwise the packet arrives at the receiver with certain delay while its contents may possibly be corrupted due to bit-level noise. If the delay is beyond an acceptable limit for the application and/or there is an error in the packet header, the packet is considered lost. Therefore, the channel model used in this work assumes (i) a random independent loss at the packet level with probability *Packet Loss Rate(PLR)* and (ii) a random memoryless channel degradation at the bit level, e.g., AWGN with BPSK. The PLR indicates the ensemble probability of packet loss due to various reasons as mentioned above.

In the followings, for the development of the proposed source decoders, we assume that, at each instant of time, the probability distribution of  $P(J_n | I_n)$  (channel information) is available at the receiver. If the bit  $I_n^{(k)}$  is located in a packet that is lost during transmission, then  $P(J_n^{(k)} | I_n^{(k)}) = 0.5$ ; otherwise this probability reflects the effect of channel degradations at the bit level. Such an abstraction of the channel condition, when observed by the source decoder, provides several advantages. First, the proposed decoders can accommodate the change of channel probability of packet loss or the change in the statistics of bit-level noise. Second, they can take advantage of soft channel (decoder) outputs if available. And third, the decoders can be used to combat the

effect of noise, the effect of packet loss, or the effects of both noise and packet loss.

### III. RECONSTRUCTION OF LSF PARAMETERS

This section begins with a brief description of the LSF quantizer used in the IS-641 and the GSM-AMR speech codecs and proceeds with the presentation of several MMSE-based reconstruction schemes for this quantizer.

#### A. IS-641 and GSM-AMR LSF Quantizers

The IS-641 speech codec employs a predictive Split-VQ for the quantization of LSF parameters [40]. The input to the quantizer is a 10 dimensional LSF vector per each frame of speech. A first-order Moving Average scalar prediction is employed to exploit the redundancies between the adjacent frames. Next, the resulting 10 dimensional LSF prediction residue vector is quantized using a 3-split Split-VQ [55] with an overall rate of 26 bits per frame (bpf). In other words, the vector of prediction residues are split to 3 vectors of dimensions (3, 3, 4) and quantized using 3 full-search VQs with bit-rates (8, 9, 9) bits, respectively. This configuration is used in four out of the eight modes/rates available in the GSM-AMR standard [38]. Three other modes in this standard also use the same 3-split Split VQ, but with different bitrates. The last mode of the GSM-AMR codec uses a Split Matrix Quantizer [61] instead.

The GSM-AMR speech coder attempts to remove the interframe redundancy by using a first-order Moving Average (MA) predictor. Using such a predictor is motivated for keeping both the source coder complexity and the propagation of channel errors at a lower level. However, there is a mismatch between this simple first-order model and the interframe dependencies of the source. Therefore, there remains some level of interframe residual redundancy, which can be exploited for improved reconstruction at the receiver.

It is well-known that there is a high level of intraframe dependency among the LSF parameters within a frame (see e.g., [56] for an analysis). The source coder under consideration attempts to exploit this intraframe dependency by using a 3-split Split-VQ. However, due to its suboptimality, there remains some level of intraframe redundancy between the quantizer three output symbols, which can be utilized at the receiver for improved signal reconstruction over noisy channels.

### B. MMSE-based Reconstruction Schemes

The source decoder corresponding to the predictive 3-split VQ used for LSF parameters in the GSM-AMR and the IS-641 codecs is designed to produce the minimum mean squared error estimate of the LSF parameters  $\mathbf{X}_n$  given the received sequence  $\mathbf{J}_{n+\delta} = [J_{n+\delta}^{(1)}, J_{n+\delta}^{(2)}, J_{n+\delta}^{(3)}]$ , where  $\delta \geq 0$  is the delay allowed in the decoding process. Based on the fundamental theorem of estimation, this is given by

$$\hat{\mathbf{x}}_n = E[\mathbf{X}_n | \mathbf{J}_{n+\delta}]. \quad (1)$$

In [27], a Sequence MMSE decoder is presented for high quality reconstruction of predictively encoded signals over noisy channels. An Asymptotically Optimum MMSE decoder for reconstruction of memoryless and predictive coded signals is proposed in [26] and [27], respectively. Other related prior works are available e.g., in [8],[10],[14]-[17]. Here, our objective is to design an efficient reconstruction solution, where ultimately the decoders act as an error concealment unit at the receiver. Therefore, we focus on low complexity yet effective solutions (approximations to equation (1)), which may not necessarily be the best possible solutions (in terms of the decoded signal to reconstruction noise ratio).

The proposed MMSE decoders reconstruct each split of the LSF prediction residues separately. Subsequently, the reconstructed LSF prediction residue vector is fed to the ordinary predictive Split-VQ decoder of GSM-AMR. Unless distinguishing different splits is required, in the followings, we use the notation  $\mathbf{Y}$  to indicate any of the 3 splits of LSF prediction residues. Similarly,  $I$  and  $J$  denote the corresponding source coder output symbol and received signal over the channel, respectively. Naturally, for each of the splits its corresponding a priori information is used.

#### B.1 Decoders MS0 and MS1:

The MS1 decoder is the basic MMSE decoder given by

$$\hat{\mathbf{y}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{Y}_n | I_n] P(I_n | J_n), \quad (2)$$

and since

$$P(I_n | J_n) = C \cdot P(I_n) \cdot P(J_n | I_n),$$

$C = \frac{1}{P(J_n)}$ , the basic MMSE decoder exploits the residual redundancy in the form of the non-uniform symbol probabilities  $P(I_n)$ . If we assume that the encoder produces equally probable

symbols,  $P(I_n) = \frac{1}{M}$ , the decoder of equation (2) collapses to the following decoder:

$$\hat{\mathbf{y}}_n = C \cdot \sum_{I_n \in \mathcal{J}} E[\mathbf{Y}_n | I_n] P(J_n | I_n), \quad (3)$$

which in this article is referred to as the MS0 decoder. This decoder is an MMSE-based decoder which neglects all the residual redundancies. The notation  $C$  indicates a factor which normalizes the sum of the probabilities to one. Here,  $C = \frac{1}{M \cdot P(J_n)}$ .

### B.2 Decoder MS2:

The decoder MS2 exploits the interframe residual redundancies and is given by

$$\hat{\mathbf{y}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{Y}_n | I_n] P(I_n | \underline{\mathcal{J}}_{n+\delta}), \quad (4)$$

where  $\delta \geq 0$  is the delay allowed in the decoding process. Assuming that the source coder produced symbols form a first-order Markov model due to the residual redundancy and a memoryless channel as described in section II-C, the probabilities in equation (4) are given by

$$P(I_n | \underline{\mathcal{J}}_{n+\delta}) = P_{fwd}(I_n) \cdot P_{bwd}(I_n), \quad (5)$$

where

$$P_{fwd}(I_n) = P(I_n | \underline{\mathcal{J}}_n) = C \cdot P(J_n | I_n) \cdot \sum_{I_{n-1} \in \mathcal{J}} P(I_n | I_{n-1}) P_{fwd}(I_{n-1}) \quad (6)$$

is a forward recursive term exploiting the dependencies of  $I_n$  with previous symbols and

$$P_{bwd}(I_n) = P(\underline{\mathcal{J}}_{n+\delta}^{n+1} | I_n) = \sum_{I_{n+1} \in \mathcal{J}} P(J_{n+1} | I_{n+1}) P(I_{n+1} | I_n) P_{bwd}(I_{n+1}) \quad (7)$$

is a backward recursive term exploiting the dependencies of  $I_n$  with symbols in the future. The backward recursion starts from

$$P_{bwd}(I_{n+\delta-1}) = P(\underline{\mathcal{J}}_{n+\delta} | I_{n+\delta-1}) = \sum_{I_{n+\delta} \in \mathcal{J}} P(J_{n+\delta} | I_{n+\delta}) P(I_{n+\delta} | I_{n+\delta-1}), \quad (8)$$

and continues backward until  $P_{bwd}(I_n)$  is derived. In equation (6),  $C$  is a factor normalizing the sum of probabilities to one. The first-order Markov model is characterized by the transition probabilities  $P(I_n | I_{n-1})$  or  $P(I_{n+1} | I_n)$  in equations (6) and (7). We note that in each time step, the forward recursion of equation (6) proceeds one step forward in time, while the backward term is recomputed over the entire backward window as indicated in equations (7) and (8). The proof

for equations (2) to (8) is presented in the Appendix for a more general case. The MS2 decoder or its variants have been previously presented in [14][15][8]. This decoder is in fact, a simplified MMSE decoder as presented in [12] and [26] for the case of a first-order Markov model ( $\gamma = 1$ ).

### B.3 Decoder MS3:

The decoder MS3 exploits both intraframe and interframe residual redundancies for improved signal reconstruction at the receiver. Consider the scenario where the source coder receives the two inputs  $\mathbf{Y}_n$  and  $\check{\mathbf{Y}}_n$  and produces two output symbols  $I_n \in \mathcal{J}$  and  $\check{I}_n \in \check{\mathcal{J}}$  at a time. Let us assume that due to the complexity constraints, there is a dependency both in the sequence of data in time and between the two symbols at each time instant. The decoder MS3 exploits both types of residual redundancy and is given by [9]

$$\hat{\mathbf{y}}_n = \sum_{I_n \in \mathcal{J}} E[\mathbf{Y}_n | I_n] P(I_n | \underline{\mathcal{J}}_{n+\delta}, \check{\underline{\mathcal{J}}}_{n+\delta}), \quad (9)$$

where  $\delta$  is the delay allowed in the decoding process. Assuming two independent first-order Markov models to capture the intraframe and interframe dependencies and a memoryless channel, we have

$$\begin{aligned} P(I_n | \underline{\mathcal{J}}_{n+\delta}, \check{\underline{\mathcal{J}}}_{n+\delta}) &\approx P(I_n | \underline{\mathcal{J}}_{n+\delta}, \check{\mathcal{J}}_n) \\ &= C \cdot P(I_n | \mathcal{J}_{n+\delta}) P(\check{\mathcal{J}}_n | I_n, \underline{\mathcal{J}}_{n+\delta}) \\ &= C \cdot P(I_n | \mathcal{J}_{n+\delta}) P(\check{\mathcal{J}}_n | I_n) \\ &= C \cdot P_{fwd}(I_n) \cdot P_{bwd}(I_n) \cdot \sum_{\check{I}_n \in \check{\mathcal{J}}} P(\check{\mathcal{J}}_n | \check{I}_n) P(\check{I}_n | I_n). \end{aligned} \quad (10)$$

In equation (10),  $C = P(\underline{\mathcal{J}}_{n+\delta}) / P(\underline{\mathcal{J}}_{n+\delta}, \check{\mathcal{J}}_n)$  is a factor which normalizes the sum of probabilities to one. The forward and backward terms are given in equations (6) and (7), respectively. The last term takes into account the dependencies between the two symbols at each time instant, which is captured by the transition probabilities  $P(\check{I}_n | I_n)$ . The decoder presented in [10] is similar to the decoder in equation (10) up to the term  $P_{bwd}$ .

For the reconstruction of each of the three splits of LSF parameters in GSM-AMR, using decoder MS3, the required intraframe observation,  $\check{\mathcal{J}}_n$ , is selected to capture the highest level of intraframe dependency. To reconstruct the LSF prediction residues of the first split the information about the received symbol of the second split is used and vice versa. Also, to reconstruct that of the third split the information about the received symbol of the second split

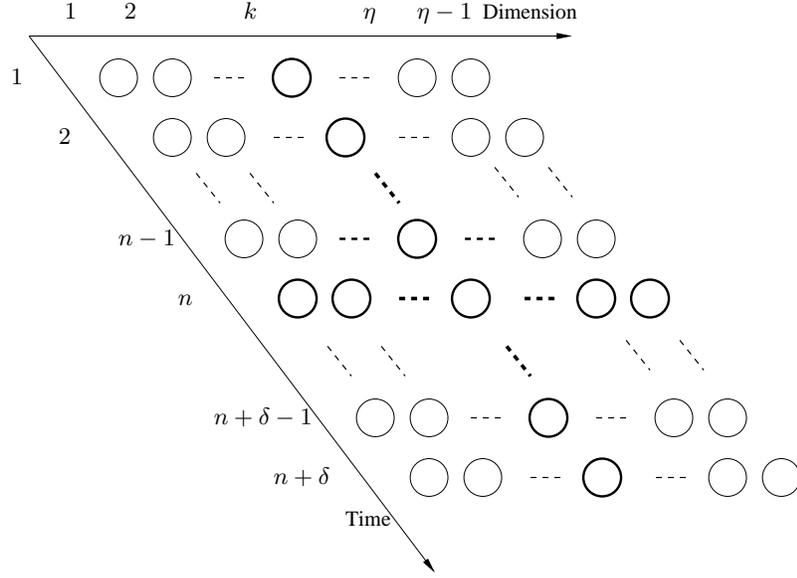


Fig. 2. An illustration of encoder output in MS4, each circle located at position  $k$  at time instant  $n$  represents the encoder output symbol  $I_n^{(k)}$ .

is used. This selection is due to the fact that the intraframe dependency of LSF parameters is higher between the neighboring parameters.

#### B.4 Decoder MS4:

The decoder MS3 presented above is a special case of a more general scheme that is described below. Consider the scenario where at each time instant the source coder, which is composed of  $\eta$  quantizers, encodes samples  $[\mathbf{Y}_n^{(1)}, \mathbf{Y}_n^{(2)}, \dots, \mathbf{Y}_n^{(\eta)}]$  to  $\eta$  symbols  $[I_1^{(1)}, I_2^{(2)}, \dots, I_n^{(\eta)}]$ , respectively, where  $I_n^{(k)} \in \mathcal{J}^{(k)}$  and  $\mathcal{J}^{(k)}$  is a set of  $M^{(k)}$  codewords,  $0 < k \leq \eta$ . Figure 2 illustrates such a scenario. Let us assume that due to design constraints, there is a residual redundancy both in the sequence of data in time and between the symbols at each time instant. The decoder MS4 exploits these dependencies and is given by

$$\hat{\mathbf{y}}_n^{(k)} = \sum_{I_n^{(k)} \in \mathcal{J}^{(k)}} E[\mathbf{Y}_n^{(k)} | I_n^{(k)}] P(I_n^{(k)} | \mathbf{J}_{n+\delta}), \quad 0 < k \leq \eta, \quad (11)$$

where  $\mathbf{J}_{n+\delta} = [\mathcal{J}_{n+\delta}^{(1)}, \mathcal{J}_{n+\delta}^{(2)}, \dots, \mathcal{J}_{n+\delta}^{(\eta)}]$ , and  $\delta$  is the delay allowed in the decoding process. Assuming two independent first-order Markov models to capture the intraframe and interframe dependencies and a memoryless channel, the symbol a posteriori probability  $P(I_n^{(k)} | \mathbf{J}_{n+\delta})$  in

equation (11) is given by

$$P(I_n^{(k)} | \underline{\mathbf{J}}_{n+\delta}) = C \cdot P_{fwd}(I_n^{(k)}) \cdot P_{bwd}(I_n^{(k)}) \cdot P_{rgt}(I_n^{(k)}) \cdot P_{lft}(I_n^{(k)}), \quad (12)$$

which is composed of four terms, namely the forward recursive term, the backward recursive term, and the left and right recursive terms (see Appendix for proof). The forward and backward terms are given in equations (6) and (7), respectively. The right recursive term exploits the dependencies of symbol  $I_n^{(k)}$  with the symbols  $[I_n^{(1)}, \dots, I_n^{(k-1)}]$  and is given by

$$\begin{aligned} P_{rgt}(I_n^{(k)}) &= P(J_n^{(1)}, \dots, J_n^{(k-1)} | I_n^{(k)}) \\ &= \sum_{I_n^{(k-1)} \in \mathcal{J}^{(k-1)}} P(J_n^{(k-1)} | I_n^{(k-1)}) P(I_n^{(k-1)} | I_n^{(k)}) P_{rgt}(I_n^{(k-1)}), \end{aligned} \quad (13)$$

where the recursion starts from

$$\begin{aligned} P_{rgt}(I_n^{(2)}) &= P(J_n^{(1)} | I_n^{(2)}) \\ &= \sum_{I_n^{(1)} \in \mathcal{J}^{(1)}} P(J_n^{(1)} | I_n^{(1)}) P(I_n^{(1)} | I_n^{(2)}), \end{aligned} \quad (14)$$

and continues towards right until  $P_{rgt}(I_n^{(k)})$  is derived. The left recursive term incorporates the dependencies of symbol  $I_n^{(k)}$  with the symbols  $[I_n^{(k+1)}, \dots, I_n^{(\eta)}]$  and is given by

$$\begin{aligned} P_{lft}(I_n^{(k)}) &= P(J_n^{(k+1)}, \dots, J_n^{(\eta)} | I_n^{(k)}) \\ &= \sum_{I_n^{(k+1)} \in \mathcal{J}^{(k+1)}} P(J_n^{(k+1)} | I_n^{(k+1)}) P(I_n^{(k+1)} | I_n^{(k)}) P_{lft}(I_n^{(k+1)}), \end{aligned} \quad (15)$$

where the recursion starts from

$$\begin{aligned} P_{lft}(I_n^{(\eta-1)}) &= P(J_n^{(\eta)} | I_n^{(\eta-1)}) \\ &= \sum_{I_n^{(\eta)} \in \mathcal{J}^{(\eta)}} P(J_n^{(\eta)} | I_n^{(\eta)}) P(I_n^{(\eta)} | I_n^{(\eta-1)}), \end{aligned} \quad (16)$$

and continues towards left until  $P_{lft}(I_n^{(k)})$  is calculated. At time instant  $n$ , first the left and right recursive terms are calculated once for symbols  $[I_n^{(1)}, \dots, I_n^{(\eta)}]$  and stored. Next, for each symbol the forward and backward terms are computed to be used in equation (12) and derive the desired a posteriori symbol probability. A similar decoder is developed in [25] which exploits both intraframe and interframe dependencies. It is noteworthy that the left and right recursions presented here lend themselves to more efficient implementations as the computations for different dimensions ( $k, 0 < k \leq \eta$ ) are identical; however, this is not the case in [25].

Exploiting both the intraframe and the interframe dependencies for efficient source decoding has received considerable attention in recent years. Adrat in a parallel, but independent, work developed a decoder similar to MS4 in [11]. The decoder MS4 exploits intraframe and interframe dependencies independently. These dependencies could be exploited jointly, e.g., using a second-order Markov model, for improved performance, but at the cost of extra complexity. This has been considered by several authors in the context of image coding and exploiting the dependencies in the horizontal and the vertical directions using different decoding rules, e.g., [18][19][20]. A similar idea is suggested in [21].

### C. Complexity Considerations

The computational complexity of the decoders MS2, MS3 and MS4 mainly consists of the computations required for their corresponding recursive terms. In the set up of decoder MS4, these are given by

$$CC_{fwd} = \sum_{k=1}^{\eta} (2M^{(k)} + 3)M^{(k)} \quad (17)$$

$$CC_{bwd} = 3\delta \sum_{k=1}^{\eta} M^{(k)2} \quad (18)$$

$$CC_{rgt} = 3 \sum_{k=1}^{\eta-1} M^{(k+1)}M^{(k)} \quad (19)$$

$$CC_{lft} = 3 \sum_{k=1}^{\eta-1} M^{(k+1)}M^{(k)}, \quad (20)$$

in terms of required number of floating point operations<sup>1</sup>. The memory requirement of the presented decoders is mainly comprised of the memory required to store the a priori transition probabilities. In general, these are matrices of size e.g.,  $M^{(k)} \times M^{(k)}$ ,  $0 < k \leq \eta$  for the case of interframe probabilities. Both the computational complexity as indicated in equations (17)-(20) and memory requirements grow exponentially with quantizer bit-rate ( $M^{(k)} = 2^{r^{(k)}}$ ). For quantizers with large bitrates, therefore, the decoders complexity will increase beyond the limits of practical interest. Consequently, methods to reduce the complexity of the algorithms are of great interest. This issue is also underlined in [14][15][26] for related scenarios. It is noteworthy that calculating the transition probabilities could also become problematic with the increase of quantizer bitrate as it requires huge sets of source samples. In the next section, we propose a

<sup>1</sup>In this work, each addition, multiplication or comparison is considered as one floating point operation (flop).

method for approximating the transition probabilities to address the issues of calculating the transition probabilities and managing the computational complexity and memory requirement of the decoders.

#### IV. APPROXIMATING THE A PRIORI TRANSITION PROBABILITIES

In this section, we present a technique to efficiently approximate the a priori transition probabilities, based on which low complexity approximate solutions to the presented decoders that employ the residual redundancies can be constructed.

Consider a vector quantizer with a codebook  $\mathcal{C}$  and the corresponding index set  $\mathcal{J}$  of  $M$  elements. Assume that due to the residual redundancy, the sequence of quantizer output in time forms a first-order Markov model. To derive an approximation to the a priori transition probabilities  $P(I_n|I_{n-1})$ , we classify the source (and hence the codebook and the index set) to  $M'$ ,  $M' \leq M$  classes, i.e.,

$$\mathcal{J} = \{\mathcal{J}_1, \mathcal{J}_2, \dots, \mathcal{J}_{M'}\},$$

in which each class  $\mathcal{J}_k$  has  $m_k$  members ( $\sum_{k=1}^{M'} m_k = M$ ). We assume that the classification is performed in a way that the probability of transition from a codeword to another codeword only depends on the class they are located in, i.e.,

$$\begin{aligned} P(I_n = i_n | I_{n-1} = i_{n-1}) &= P(I_n = i'_n | I_{n-1} = i'_{n-1}) = P_{kl}, \\ \forall i_n, i'_n \in \mathcal{J}_k, \forall i_{n-1}, i'_{n-1} \in \mathcal{J}_l. \end{aligned} \tag{21}$$

Should we have the codebook classified in this manner, we would only need  $M' \times M'$  transition probabilities to characterize the Markov model of the encoder output sequence. Subsequently, the relationship between the class and symbol transition probabilities is given by the following simple derivations:

$$\begin{aligned} P(I_n \in \mathcal{J}_k | I_{n-1} \in \mathcal{J}_l) &= \sum_{I_n \in \mathcal{J}_k} P(I_n | I_{n-1} \in \mathcal{J}_l) \\ &= \sum_{I_n \in \mathcal{J}_k} \sum_{I_{n-1} \in \mathcal{J}_l} P(I_n | I_{n-1}) P(I_{n-1} | I_{n-1} \in \mathcal{J}_l) \\ &= m_k \cdot P(I_n | I_{n-1}), \end{aligned} \tag{22}$$

in which the last result is derived using the assumption of equation (21). Equation (22) shows that the codeword (symbol) transition probabilities can be approximated as a scaled version of

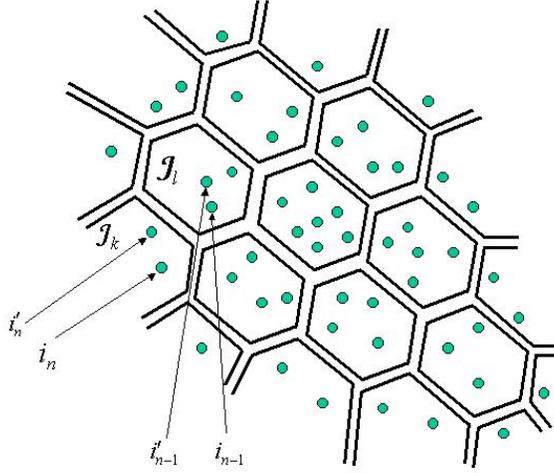


Fig. 3. Classification of the source codewords. The probability of transition between codewords is approximated by the corresponding class transition probability. See equations (21) and (22).

the class transition probabilities. Therefore, the problem now collapses to that of determining a transition probability matrix of size  $M' \times M'$ , which can be found with much less data. Subsequently, this matrix can be stored much more efficiently, reducing the memory requirement by a factor of  $(\frac{M}{M'})^2$ . The intraframe transition probabilities can be also approximated in a similar fashion. Figure 3 is a representative diagram of such a scenario.

In order to classify the codebook in a way such that the equation (21) holds, we suggest LBG [63] quantization of the source with  $M'$  levels and defining the classes as the quantization Voronoi regions. Subsequently, we can classify the codewords of the size  $M$  codebook. We note that this does not guarantee the validity of the assumption of equation (21) and it is only an approximate technique. However, our numerical results demonstrate its fruitfulness.

It is straight forward to see that, using this technique, the presented decoders which exploit the residual redundancies (MS2, MS3 and MS4) can be implemented in a much more computationally efficient manner. This is due to the fact that the transition probabilities now depend only on the classes, and the summations in the forward, backward, right and left recursive equations of

(6), (7), (13) and (15) can be broken over the classes accordingly. We have

$$\begin{aligned} P_{fwd}(I_n) &= P(I_n|\underline{J}_n) \\ &= C \cdot P(J_n|I_n) \cdot \sum_{k=1}^{M'} P(I_n|I_{n-1} \in \mathcal{J}_k) \sum_{I_{n-1} \in \mathcal{J}_k} P_{fwd}(I_{n-1}), \end{aligned} \quad (23)$$

and

$$\begin{aligned} P_{bwd}(I_n) &= P(\underline{J}_{n+\delta}^{n+1}|I_n) \\ &= \sum_{k=1}^{M'} P(I_{n+1}|I_n) P_{bwd}(I_{n+1}) \sum_{I_{n+1} \in \mathcal{J}_k} P(J_{n+1}|I_{n+1}). \end{aligned} \quad (24)$$

The right and left recursive terms of equations (13) and (15) will be given by similar approximate expressions. This leads to a noticeably smaller computational complexity. In the set up of decoder MS4 (section III-B.4), assuming that the codebook  $\mathcal{J}^{(k)}, 0 < k \leq \eta$  is partitioned to  $M'^{(k)}$  classes, the computational complexity is now given by

$$CC'_{fwd} = \sum_{k=1}^{\eta} 3M^{(k)} + 2M'^{(k)2} \quad (25)$$

$$CC'_{bwd} = 3\delta \sum_{k=1}^{\eta} M'^{(k)2} \quad (26)$$

$$CC'_{lft} = 3 \sum_{k=1}^{\eta-1} M'^{(k+1)} M'^{(k)} \quad (27)$$

$$CC'_{rgt} = 3 \sum_{k=1}^{\eta-1} M'^{(k)} M'^{(k+1)}, \quad (28)$$

in terms of the required number of floating point operations. If any of the recursive terms other than the forward term is used, the computational complexity also includes an overhead of  $\sum_{k=1}^{\eta} M^{(k)}$  to calculate the channel related probabilities of classes from their corresponding symbol probabilities at each time instant (see e.g., equation (24) for the case of backward recursive term). In equations (25)-(28), it is assumed that the a priori information is stored in the form of scaled class transition probabilities, e.g.,  $m_l^{-1} P(I_n \in \mathcal{J}_l | I_{n-1} \in \mathcal{J}_k)$  for the case of dependencies over time.

Figure 4 presents the performance of decoder MS2 for the reconstruction of a source encoded with a scalar quantizer at rates  $r = 3$  and  $r = 8$  bits/sample ( $M = 8$  and  $M = 256$ ). The source is a 10'th-order Gauss-Markov source with coefficients given in Table I. These coefficients correspond to the LPC coefficients of one 20ms frame of speech. The original transition probabilities

coef. 1 – 5	-1.3822	0.3399	0.1772	0.6760	-0.6396
coef. 6 – 10	0.0719	-0.1386	0.6024	-0.4561	0.1560

TABLE I

## COEFFICIENTS OF THE SYNTHESIZED GAUSS-MARKOV SOURCE

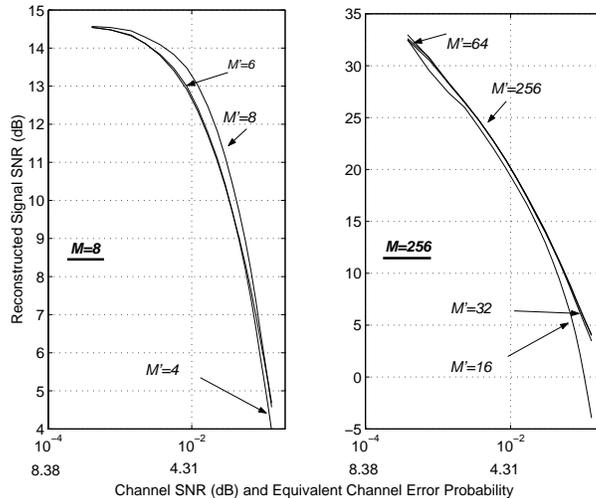


Fig. 4. Performance of the MS2 decoder ( $\delta = 0$ ) for transmission of a Gauss-Markov source over a Soft Output Channel. The source is encoded with a scalar quantizer,  $M = 8$  and  $M = 256$ , and different number of classes  $M'$  is used.

are given by  $M \times M = 8 \times 8$  and  $256 \times 256$  sized matrices. In Figure 4, the performance of decoder MS2 with different number of classes  $M'$  is depicted. It is clear that negligible performance degradation is caused by selecting the number of classes,  $M'$ , as small as 4 and 32 for cases of  $M = 8$  and  $M = 256$ , respectively. In return, this indicates a respective 4 and 64 times reduction of memory requirement and 2.7 and 46 times reduction of computational complexity.

## V. PACKETIZATION STRATEGY

Interleaving during packetization, as a Multiple Description Coding strategy, creates robustness to packet loss through proper diversification (distribution) of the data across different packets (descriptions). Subsequently, a (source) decoding technique, matched to the packetization scheme, is employed which exploits the diversity for improved reconstruction. For packetization of CELP-based coded speech, a popular method is to interleave the encoded speech bitstream across packets on a frame by frame basis [45][48]. This is matched to the employed error conceal-

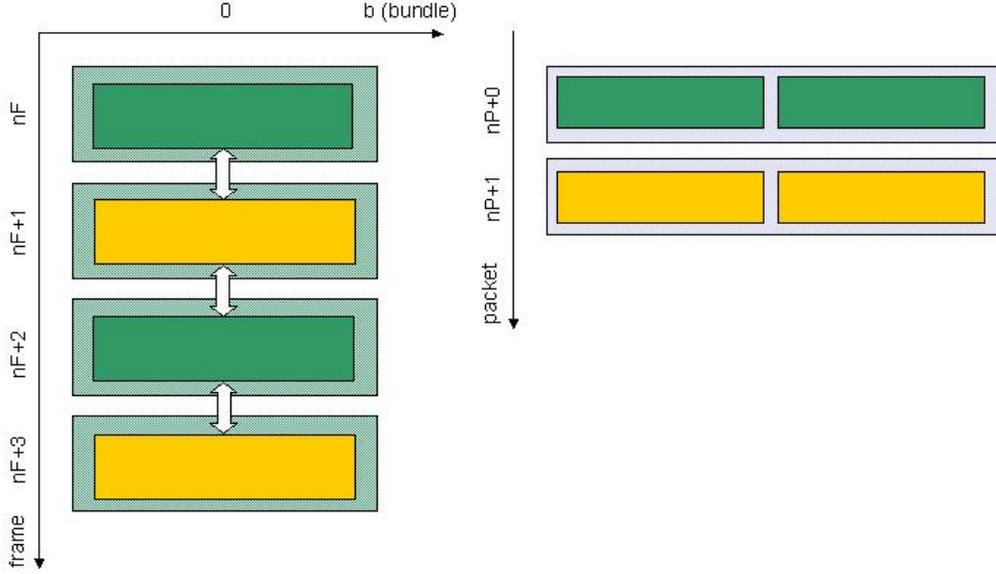


Fig. 5. Packetization for  $F = 4$  and  $P = 2$  ( $B = 1$ ). The arrows represent the dependencies due to residual redundancies.  $n \geq 0$  is a notion of time indicating the set of frames to be interleaved.

ment scheme, which often relies on the interframe dependencies and estimates the parameters of the missing speech frames from the corresponding data in the previous frames. Examples of such error concealment schemes are suggested for CELP-based speech codecs of GSM-AMR [38] and IS-641 [40]. Figure 5 shows a representative diagram for interleaving a set of  $F$  frames of data ( $F = 4$ ), across  $P$  packets ( $P = 2$ ).

In this work, we propose a packetization strategy that is matched to the presented source decoders exploiting both interframe and intraframe dependencies. Consider the scenario where the source coder outputs certain number of bits at each time instant (frame). Let us assume that these bits may be partitioned into  $B$  bundles of data where in general, there are dependencies between the (neighboring) data bundles at each time instant, i.e., within a frame, and over time, i.e., between frames. The proposed packetization rule takes into account these dependencies and is given by

$$b + f + nF \equiv p + nP \pmod{P}, \quad (29)$$

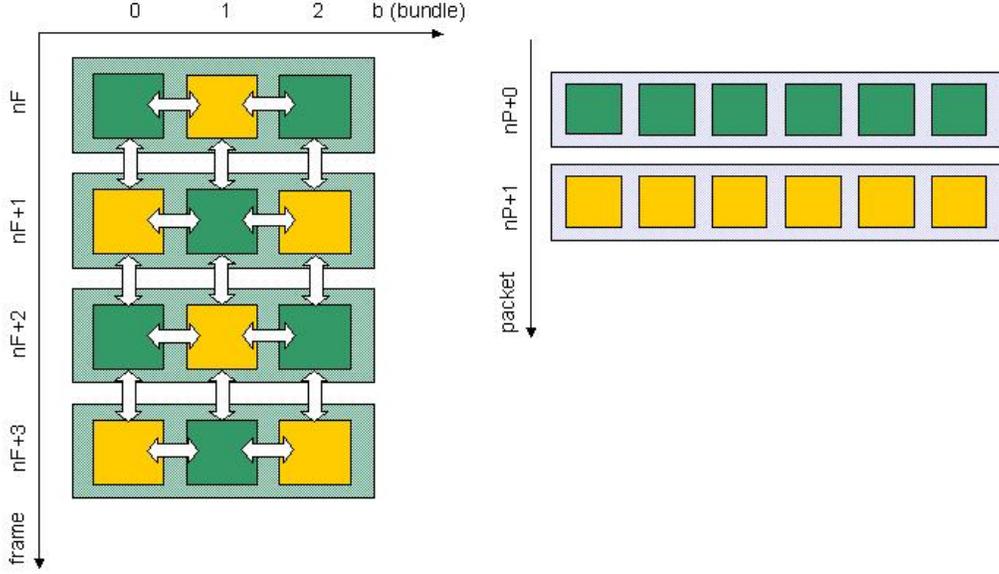


Fig. 6. Packetization for  $F = 4$ ,  $B = 3$  and  $P = 2$ . The arrows represent the dependencies due to residual redundancies.

which distributes a set of  $F$  frames of data, each composed of  $B$  bundles, across  $P$  packets ( $P$  indicates the diversity depth). In equation (29), we have  $n, b, B, f, F, p, P \in \mathbb{Z}$ , where  $0 \leq b < B$  is the bundle number,  $f + nF$ ,  $0 \leq f < F$ , indicates the frame number,  $p + nP$ ,  $0 \leq p < P$ , indicates the packet number and  $(\text{mod } P)$  is the modulo operation in base  $P$ . Also,  $n \geq 0$  is a notion of time indicating the set of frames to be interleaved. For a given  $n, B, F$  and  $P$ , equation (29) determines in which packet  $p$ , the bundle  $b$  of frame  $f$  is to be located. Figure 6 illustrates the case for  $B = 3$ ,  $F = 4$  and  $P = 2$ . As seen, if one of the packets is lost, the other packet contains all the neighboring bundles of the missing data which are utilized by the source decoder to reconstruct the missing bundles. For the case of  $B = 1$ , equation (29) is simplified to the interleaving scheme of [7] as exemplified in Figure 5.

It is noteworthy that the proposed packetization strategy imposes a maximum delay of  $F - 1$  frames. Equation (29) presents a general formulation with three design parameters  $B, F$  and  $P$ . The design is performed based on a trade-off of diversity, delay, and bandwidth (due to packet

header), which ultimately affect the performance. Real time voice communications have a tight delay budget, while non-interactive applications such as streaming tolerate larger delays and therefore, can benefit from packetization schemes with higher diversity.

## VI. PERFORMANCE EVALUATION

In this section, the performance of the proposed source decoding and packetization schemes for reconstruction of LSF parameters of speech, encoded with GSM-AMR codec and, transmitted over a channel disturbed by noise and/or packet loss is presented.

### A. Experiment Setup

We use a training database of 175726 LSF vectors, each representing the spectrum of a 20ms frame of speech. This database contains a combination of clean speech and speech with background noise from a number of male and female speakers. Another outside test database of 30000 LSF vectors derived from a recorded clean speech is used to test the performance of the proposed decoders<sup>2</sup>. The average spectral distortion measure (measured in the frequency range of  $f_1 = 60$  Hz to  $f_2 = 3500$  Hz) is employed to measure the objective quality of the reconstructed LPC coefficients. This criterion is a function of the distortion introduced in the power spectral density of speech in each particular frame. The spectral distortion in the  $n$ th frame is given by

$$SD^{(n)} = \sqrt{\frac{1}{f_2 - f_1} \int_{f_1}^{f_2} [10 \log_{10}(P^{(n)}(f)) - 10 \log_{10}(\hat{P}^{(n)}(f))]^2 df}, \quad (30)$$

in which

$$P^{(n)}(f) = \frac{1}{|A^{(n)}(\exp(j2\pi f/F_s))|^2}, \quad (31)$$

and

$$\hat{P}^{(n)}(f) = \frac{1}{|\hat{A}^{(n)}(\exp(j2\pi f/F_s))|^2} \quad (32)$$

are the original and reconstructed power spectral density of the  $n$ th frame, respectively. The terms  $A^{(n)}(z)$  and  $\hat{A}^{(n)}(z)$  are the corresponding original and reconstructed LPC filters. We also use signal to reconstruction noise ratio (dB) to evaluate the effect of error on individual LSF parameters.

As a reference for comparison, we use the error concealment scheme suggested for the GSM-AMR [38][39] and IS-641 [40] speech codecs. In this scheme, a CRC is used for the first two splits

<sup>2</sup>The speech databases used in this work are provided by Nortel Networks.

of the Split-VQ. If the CRC fails due to bit error, a Bad Frame Indicator flag (BFI) is set, and the error concealment process is activated. In this case, the LSF parameters are replaced by

$$\hat{x}_n^{(j)} = \alpha \hat{x}_{n-1}^{(j)} + (1 - \alpha) \bar{x}^{(j)}, \quad \alpha = 0.95, j = 1, \dots, 10, \quad (33)$$

where  $\hat{x}_n^{(j)}$  indicates the  $j$ 'th reconstructed LSF parameter at time instant  $n$ , and  $\bar{x}^{(j)}$  is the average value of  $j$ 'th LSF parameter. In our simulations of IS-641, we simulate the same procedure. The quantization indices of all the three splits of a frame are located in the same packet. If the BFI is set due to bit error or the LSF parameters are missing due to packet loss, the error concealment process mentioned above is activated.

### B. Complexity Considerations for Proposed MMSE-based Decoders

The required a priori transition probabilities, modeling the interframe dependencies of GSM-AMR encoded LSF parameters, are described with one  $2^8 \times 2^8$  matrix for the first split and two  $2^9 \times 2^9$  probability transition matrices for the other two splits. One can see that finding these values requires a very large speech database which can make the task impractical. As well, the memory required to store these matrices is more than 2 Megabytes. Therefore, approximation of these values is inevitable. Using the method described in section IV with a class size of  $M' = 32$  for each of the splits, the problem is reduced to calculating three  $2^5 \times 2^5$  transition probability matrices and only 12 kilobytes of memory requirement. As well, we use the same technique (with  $M' = 32$ ) to approximate the transition probabilities between the splits, which model the intraframe residual redundancies of GSM-AMR encoded LSF parameters.

In terms of computational complexity of the decoders, this approximation leads to a reduction by an impressive factor of 118 for the forward recursion and a factor of 192 for the backward, left and right recursive terms (see equations (17)-(20) and (25)-(28)). In the following section, we use these configurations and study the performance of the system. The choice of  $M' = 32$  with  $M = 512$  and  $M = 256$  is motivated by the results presented in Figure 4 and is further supported by the simulation results of the next section.

### C. Numerical Results

Six decoding schemes are considered here. First the error concealment scheme from the standard (GSM-AMR, IS-641) is considered as a reference. The methods MS0 and MS1 are the basic MMSE reconstruction algorithms neglecting all the residual redundancies with MS1 exploiting

the symbol a priori probability  $P(I_n)$  (equations (3) and (2), respectively). The method MS2 exploits only interframe redundancies for reconstruction (equations (4) and (5)). The method MS3 exploits both intraframe and interframe residual redundancies at the receiver using an approximate formulation (equations (9) and (10)). Finally, the method MS4 is the most comprehensive scheme considered here, which also exploits intraframe and interframe redundancies (equations (11) and (12)). We use the packetization scheme of Figure 6 with decoders MS3 and MS4 which exploit the intraframe and interframe redundancies. The three bundles of each frame, as demonstrated in Figure 6, now correspond to the three splits of the Split-VQ, containing 8,9 and 9 bits, respectively. For all other decoders that base their estimation of missing/erroneous LSF parameters on interframe dependencies, we use the packetization scheme of Figure 5. In [28], based on a Gaussian Mixture Modelling of the probability distribution of the LSF parameters in GSM-AMR, a loss concealment algorithm is proposed which estimates the missing splits based on the present ones.

Table II presents the performance of the above mentioned schemes for reconstruction of the Split-VQ encoded LSF parameters transmitted using a BPSK modulation over an AWGN channel with soft outputs. Also, the performance of these schemes in presence of packet loss is provided in Table III (packet erasure channel). It is clear from the tables that employing the symbol a priori probability, interframe and intraframe residual redundancies constantly improves the reconstruction quality. The method, MS3 is always better than MS2, MS2 is always better than MS1 and MS1 is always better than MS0 even in good channel conditions. In poor channel conditions, the MMSE decoders rely more on the source a priori information. Therefore, the performance advantage of MS1, MS2 and MS3 is higher in these channel conditions. As expected, the method MS4 provides certain level of performance improvement over the MS3 decoder due to its more accurate formulation. However, the amount of such improvement suggests that the decoder MS3 still provides a reasonable approximation and given its complexity advantage, it could be an attractive choice for the current application. The advantage of exploiting the intraframe residual redundancies are higher compared to that of interframe redundancies and this is due to the use of the first order MA prediction at the source coder, which partially removes the interframe dependencies.

Table IV presents the performance gain of the proposed decoders over the standard method in presence of both noise and packet loss. The gain is given in terms of improvement of signal

Channel SNR (dB)	BER	IS-641	MS0	MS1	MS2	MS3	MS4
1.00	0.0560	4.99	4.26	3.63	3.51	3.16	3.07
2.00	0.0370	4.08	3.55	3.04	2.94	2.65	2.58
3.00	0.0220	3.15	2.81	2.44	2.36	2.14	2.10
4.00	0.0120	2.28	2.13	1.89	1.84	1.69	1.67
5.00	0.0059	1.65	1.60	1.47	1.44	1.37	1.35
6.00	0.0023	1.27	1.26	1.20	1.19	1.14	1.15
7.00	0.0008	1.09	1.09	1.08	1.07	1.06	1.06

TABLE II

SPECTRAL DISTORTION (dB) OF THE TEST LSF DATABASE RECONSTRUCTED USING SIX DECODING SCHEMES FOR TRANSMISSION OVER AN AWGN CHANNEL WITH SOFT OUTPUTS AND BPSK MODULATION, NO PACKETIZATION IS USED AND  $\delta = 0$ .

PLR(%)	IS-641	MS1	MS2	MS3	MS4
25	3.28	2.76	2.62	2.50	2.43
20	2.84	2.42	2.31	2.21	2.15
15	2.45	2.09	2.00	1.92	1.88
10	1.98	1.73	1.65	1.61	1.58
5	1.55	1.40	1.36	1.34	1.32
1	1.10	1.08	1.07	1.07	1.06

TABLE III

SPECTRAL DISTORTION (dB) OF THE TEST LSF DATABASE RECONSTRUCTED USING FIVE DECODING SCHEMES FOR TRANSMISSION OVER A PACKET ERASURE CHANNEL

to reconstruction noise ratio for each of the reconstructed LSF parameters. From Table IV, it is observed that the proposed schemes achieve substantial gains over the standard. The average gain over the standard reaches 4.14 dB for the MS4 decoder. The level of performance improvement provided by different considered methods follow a pattern similar to what described above; by exploiting more of the available residual redundancy during the source decoding process, a higher quality reconstruction of the source is achieved. The provided results confirm the effectiveness

Method	LSF1	LSF2	LSF3	LSF4	LSF5	LSF6	LSF7	LSF8	LSF9	LSF10	Mean
MS1	2.66	3.23	2.66	2.49	2.60	2.55	2.49	2.53	2.34	2.36	2.59
MS2	2.99	3.71	3.39	2.91	3.20	3.11	3.12	3.52	2.80	2.81	3.16
MS3	3.44	4.39	4.68	4.38	3.76	3.91	3.84	3.80	3.18	3.00	3.84
MS4	3.58	4.52	4.95	4.63	4.30	4.80	4.03	4.01	3.33	3.30	4.14

TABLE IV

GAIN (dB) IN SIGNAL TO RECONSTRUCTION NOISE RATIO OF THE PROPOSED ERROR CONCEALMENT SCHEMES AS COMPARED TO THE SCHEME USED IN GSM-AMR CODEC, IN PRESENCE OF NOISE (CHANNEL SNR=2dB, BER=0.037) AND PACKET LOSS (PLR=5%).

of the proposed decoders.

As discussed before, the suggested packetization process imposes a delay of up to  $F - 1$  frames. This delay contributes to the diversity of the packetization scheme. As well, the decoders MS2, MS3 and MS4 utilize this delay for more effective reconstruction exploiting the dependency of a frame with the subsequent frames. As mentioned, this is a design parameter for constructing effective packetization schemes.

## VII. CONCLUSIONS

Several techniques for efficient error concealment of LSF parameters in GSM-AMR codec, transmitted over a channel disturbed with noise and/or packet loss are presented. These schemes are MMSE-based decoders that exploit the intraframe and the interframe residual redundancies. The decoders maintain a manageable level of complexity, thanks to a proposed methodology for efficient approximation of the a priori transition probabilities. Also, a packetization strategy, that is matched to the considered decoders, based on the concepts of Multiple Description Coding is presented. Numerical results are provided, which demonstrate the effectiveness of the suggested reconstruction schemes.

One possible direction of future research is the extension of the presented decoding methods to also exploit the memory of the channel for improved reconstruction at the receiver. This is expected to be achieved at the cost of increased complexity; however, the amount of gain and the increase in complexity is subject to further investigations. A related prior work, which can serve as a good starting point for such studies is presented in [17].

Another useful direction is to explore other, possibly more accurate, approximations of the a priori transition probabilities. Also of interest, is to explore the possible gains due to using higher order Markov models to exploit the available residual redundancies. A good starting point is [26].

## APPENDIX

The symbol a posteriori probabilities, described in section III-B.4, are calculated by the following equation as detailed below. Figure 2 illustrates such a scenario.

$$\begin{aligned}
P(I_n^{(k)} | \mathbf{J}_{n+\delta}) &\approx P(I_n^{(k)} | \underline{J}_{n+\delta}^{(k)}, J_n^{(1)}, \dots, J_n^{(k-1)}, J_n^{(k)}, \dots, J_n^{(\eta)}) \\
&= C_1 \cdot P(I_n^{(k)}, \underline{J}_n^{(k)}, \underline{J}_{n+\delta}^{(k),n+1}, J_n^{(1)}, \dots, J_n^{(k-1)}, J_n^{(k)}, \dots, J_n^{(\eta)}) \\
&= C \cdot P(I_n^{(k)} | \underline{J}_n^{(k)}) \cdot P(\underline{J}_{n+\delta}^{(k),n+1} | I_n^{(k)}) \cdot P(J_n^{(1)}, \dots, J_n^{(k-1)} | I_n^{(k)}) \cdot P(J_n^{(k)}, \dots, J_n^{(\eta)} | I_n^{(k)}) \\
&= C \cdot P_{fwd}(I_n^{(k)}) \cdot P_{bwd}(I_n^{(k)}) \cdot P_{rgt}(I_n^{(k)}) \cdot P_{lft}(I_n^{(k)}), \tag{34}
\end{aligned}$$

where  $C_1 = 1/P(\underline{J}_{n+\delta}^{(k)}, J_n^{(1)}, \dots, J_n^{(k-1)}, J_n^{(k)}, \dots, J_n^{(\eta)})$  and  $C = C_1 \cdot P(\underline{J}_n^{(k)})$ . In equation (34), it is assumed that each symbol is dependent to the symbols to its left, right, previous and after (in time) through a first order Markov model, but it is independent of the other symbols. The memoryless assumption of the channel is also utilized. Equation (34), shows that the desired symbol a posteriori probability is composed of four multiplying terms; namely the forward recursive, backward recursive, right recursive and left recursive terms. The forward and backward terms exploit the residual redundancies in time, while the right and left recursive terms exploit the intraframe redundancies present between different source coder produced symbols at each time instant.

The forward term is given by

$$\begin{aligned}
P(I_n | \underline{J}_n) &= C_1 \cdot P(I_n, \underline{J}_n) \\
&= C_1 \cdot P(\underline{J}_{n-1}) \cdot P(I_n | \underline{J}_{n-1}) \cdot P(J_n | I_n, \underline{J}_{n-1}) \\
&= C \cdot P(J_n | I_n) \cdot P(I_n | \underline{J}_{n-1}) \\
&= C \cdot P(J_n | I_n) \cdot \sum_{I_{n-1}} P(I_n | I_{n-1}, \underline{J}_{n-1}) \cdot P(I_{n-1} | \underline{J}_{n-1}) \\
&= C \cdot P(J_n | I_n) \cdot \sum_{I_{n-1}} P(I_n | I_{n-1}) \cdot P(I_{n-1} | \underline{J}_{n-1}), \tag{35}
\end{aligned}$$

in which  $C_1 = 1/P(\underline{J}_n)$  and  $C = P(\underline{J}_{n-1})/P(\underline{J}_n)$ . The backward term is calculated as follows:

$$\begin{aligned}
P(\underline{J}_{n+\delta}^{n+1}|I_n) &= \sum_{I_{n+1}} P(\underline{J}_{n+\delta}^{n+1}|I_{n+1}, I_n) \cdot P(I_{n+1}|I_n) \\
&= \sum_{I_{n+1}} P(I_{n+1}|I_n) \sum_{I_{n+2}} P(\underline{J}_{n+\delta}^{n+1}|I_n, I_{n+1}, I_{n+2}) \cdot P(I_{n+2}|I_{n+1}) \\
&= \sum_{I_{n+1}} P(I_{n+1}|I_n) \sum_{I_{n+2}} P(I_{n+2}|I_{n+1}) \cdots \sum_{I_{n+\delta}} P(I_{n+\delta}|I_{n+\delta-1}) \cdot P(\underline{J}_{n+\delta}^{n+1}|\underline{I}_{n+\delta}^n).
\end{aligned}$$

Considering the memoryless assumption of the channel, we have  $P(\underline{J}_{n+\delta}^{n+1}|I_n, \underline{I}_{n+\delta}^n) = \prod_{k=1}^{\delta} P(J_{n+k}|I_{n+k})$ , which simplifies the backward equation to

$$\begin{aligned}
P(\underline{J}_{n+\delta}^{n+1}|I_n) &= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1})P(I_{n+1}|I_n) \sum_{I_{n+2}} P(J_{n+2}|I_{n+2})P(I_{n+2}|I_{n+1}) \cdots \\
&\quad \sum_{I_{n+\delta}} P(J_{n+\delta}|I_{n+\delta})P(I_{n+\delta}|I_{n+\delta-1}) \\
&= \sum_{I_{n+1}} P(J_{n+1}|I_{n+1})P(I_{n+1}|I_n)P(\underline{J}_{n+\delta}^{n+2}|I_{n+1}), \tag{36}
\end{aligned}$$

which is calculated recursively at each time instant, starting from

$$P(J_{n+\delta}|I_{n+\delta-1}) = \sum_{I_{n+\delta}} P(J_{n+\delta}|I_{n+\delta})P(I_{n+\delta}|I_{n+\delta-1}),$$

and continuing backward in  $\delta$  steps until  $P(\underline{J}_{n+\delta}^{n+1}|I_n)$  is found. Equation (36) for the backward recursive term can be rewritten as

$$P_{bwd}(I_n) = \sum_{I_{n+1}} P(J_{n+1}|I_{n+1})P(I_{n+1}|I_n)P_{bwd}(I_{n+1}). \tag{37}$$

Similarly, the right recursive term is given by

$$\begin{aligned}
P(J_n^{(1)}, \dots, J_n^{(k-1)}|I_n^{(k)}) &= \sum_{I_n^{(k-1)}} P(J_n^{(k-1)}|I_n^{(k-1)})P(I_n^{(k-1)}|I_n^{(k)}) \\
&\quad \cdot \sum_{I_n^{(k-2)}} P(J_n^{(k-2)}|I_n^{(k-2)})P(I_n^{(k-2)}|I_n^{(k-1)}) \cdots \\
&\quad \sum_{I_n^{(1)}} P(J_n^{(1)}|I_n^{(1)})P(I_n^{(1)}|I_n^{(2)}) \\
&= \sum_{I_n^{(k-1)}} P(J_n^{(k-1)}|I_n^{(k-1)})P(I_n^{(k-1)}|I_n^{(k)})P(J_n^{(1)}, \dots, J_n^{(k-2)}|I_n^{(k-1)}), \tag{38}
\end{aligned}$$

which is calculated recursively at each time instant, starting from

$$P(J_n^{(1)}|I_n^{(2)}) = \sum_{I_n^{(1)}} P(J_n^{(1)}|I_n^{(1)})P(I_n^{(1)}|I_n^{(2)}),$$

and continuing towards right in  $k - 1$  steps.

In the same fashion, the left recursive term is given by

$$\begin{aligned}
P(J_n^{(k+1)}, \dots, J_n^{(\eta)} | I_n^{(k)}) &= \sum_{I_n^{(k+1)}} P(J_n^{(k+1)} | I_n^{(k+1)}) P(I_n^{(k+1)} | I_n^{(k)}) \\
&\cdot \sum_{I_n^{(k+2)}} P(J_n^{(k+2)} | I_n^{(k+2)}) P(I_n^{(k+2)} | I_n^{(k+1)}) \dots \\
&\sum_{I_n^{(\eta)}} P(J_n^{(\eta)} | I_n^{(\eta)}) P(I_n^{(\eta)} | I_n^{(\eta-1)}) \\
&= \sum_{I_n^{(k+1)}} P(J_n^{(k+1)} | I_n^{(k+1)}) P(I_n^{(k+1)} | I_n^{(k)}) P(J_n^{(k+2)}, \dots, J_n^{(\eta)} | I_n^{(k+1)}), \quad (39)
\end{aligned}$$

which is calculated recursively at each time instant, starting from

$$P(J_n^{(\eta)} | I_n^{(\eta-1)}) = \sum_{I_n^{(\eta)}} P(J_n^{(\eta)} | I_n^{(\eta)}) P(I_n^{(\eta)} | I_n^{(\eta-1)}),$$

and continuing towards left in  $\eta - k$  steps.

#### REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423 and 623-656, 1948.
- [2] K. Sayood, J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders," *IEEE Trans. Commun.*, vol.39, No.6, pp. 838-845, 1991.
- [3] J. Hagenauer, "Source-controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, No. 9, Sept. 1995.
- [4] K. Sayood, L. Fuling and J. D. Gibson, "A constrained joint source/channel coder design," *IEEE Trans. Sel. Areas Commun.*, vol. 12, No. 9, pp. 1584-1593, Dec. 1994.
- [5] F. I. Alajaji, N. Phamdo and T. E. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech and Audio Proc.*, vol. 4, No. 5, Sept. 1996.
- [6] A. Ruscitto and E. M. Biglieri, "Joint source and channel coding using turbo codes over rings," *IEEE Trans. Commun.*, vol. 46, No. 8, pp. 981-984, Aug. 1998.
- [7] T. Fazel and T. E. Fuja, "Joint source-channel decoding of block-encoded compressed speech," in *Proc. Conf. Information Sciences and Systems*, pp. FA5-1-FA5-6, Mar. 2000.
- [8] T. Fingscheidt and P. Vary, "Softbit speech decoding: A new approach to error concealment," *IEEE Trans. on Speech and Audio Proc.*, vol. 9, No. 3, Mar. 2001.
- [9] F. Lahouti and A. K. Khandani, "Approximating and exploiting the residual redundancies- Applications to efficient reconstruction of speech over noisy channels," *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proc.*, vol.2, pp.721-724, Salt Lake City, UT, May 2001.
- [10] M. Adrat, J. Spittka, S. Heinen and P. Vary, "Error concealment by near optimum MMSE-estimation of source codec parameters," *IEEE Workshop on Speech Coding*, pp. 84-86, 2000.

- [11] M. Adrat, "Iterative source-channel decoding for digital mobile communications," Ph.D. Thesis, Aachen University, ABDN Vol. 16 (ed. P. Vary), ISBN 3-86073-835-6, 2003.
- [12] F. Lahouti and A. K. Khandani, "An efficient MMSE source decoder for noisy channels," *Proc. Int. Symp. Telecommun.*, pp. 784-787, Tehran, Iran, Sept. 2001.
- [13] F. Lahouti and A. K. Khandani, "Sequence MMSE source decoding over noisy channels using the residual redundancies", *Annual Allerton Conference on Communication, Control and Computing*, Illinois, Urbana-Champaign, October 2001.
- [14] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels- Applications to combined-source channel coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 186-103, 1994.
- [15] D. J. Miller and M. Park, "A sequence-based approximate MMSE decoder for source coding over noisy channels using discrete hidden Markov models," *IEEE Trans. Commun.*, vol.46, No.2, pp. 222-231, 1998.
- [16] N. Phamdo, F. Alajaji and N. Farvardin, "Quantization of memoryless and Gauss-Markov sources over binary Markov channels," *IEEE Trans. Commun.*, vol. 45, No. 6, pp. 668-675, June 1997.
- [17] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory*, vol. 45, No. 4, pp. 1293-1307, 1999.
- [18] S. Emami and S. L. Miller, "DPCM picture transmission over noisy channels with the aid of a Markov model," *IEEE Trans. Image Proc.*, vol. 4, No. 11, pp. 1473-1481, 1995.
- [19] M. Park and D. J. Miller, "Improved Image decoding over noisy channels using MMSE estimation and a Markov mesh," *IEEE Trans. Image Proc.*, vol. 8, No. 6, pp. 863-867, 1999.
- [20] R. Link and S. Kallel, "Optimal use of Markov models for DPCM picture transmission over noisy channels." *IEEE Trans. Image Proc.*, vol. 48, No. 10, pp. 1702-1711, 2000.
- [21] S. Heinen, "An optimal MMSE-estimator for source codec parameters using intra-frame and inter-frame correlation," *Proc. Int. Symp. Inform. Theory*, pp. 237, June 2001.
- [22] T. Hindelang, T. Fingscheidt, N. Seshadri and R. V. Cox, "Combined source/channel (de-)coding: can a priori information be used twice?," *Proc. IEEE Int. Symp. Inform. Theory*, pp. 266, 2000.
- [23] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE J. Select. Areas Commun.*, vol. 19, No. 9, pp. 1662-1670, Sept. 2001.
- [24] M. Adrat, P.Vary and J. Spittka, "Iterative source-channel decoder using extrinsic information from softbit-source decoding," *Proc. IEEE Int. Confs. Acoust., Speech and Sig. Proc.*, vol. 4, pp. 2653-2656, Salt Lake City, UT, May 2001.
- [25] T. Fingscheidt, T. Hindelang, R. V. Cox, and N. Seshadri, "Joint source-channel (de-)coding for mobile communications" *IEEE Trans. Commun.*, vol. 50, No.2, 2002.
- [26] F. Lahouti and A. K. Khandani, "Efficient MMSE source decoding over noisy channels", *IEEE Trans. Inform. Theory*, vol. 50, No. 9, Sep. 2004.
- [27] F. Lahouti and A. K. Khandani, "Reconstruction of predictively encoded signals over noisy channels using a sequence MMSE decoder," *IEEE Trans. Commun.*, vol. 52, No. 8, Aug. 2004.
- [28] R. Martin, C. Hoelper, and I. Wittke, "Estimation of missing LSF parameters using Gaussian Mixture

- Models,” *IEEE Int. Confs. Acoustics, Speech and Signal Processing*, Salt Lake City, May 2001.
- [29] J. Tang, “Evaluation of double sided periodic substitution (DSPS) method for recovering missing speech in packet voice communications,” *Proc. Tenth Annual Int. Phoenix on Computers and Communications*, pp. 454-458, 1991.
- [30] O.J. Wasem, D. J. Goodman and C. A. Dvorak, “The effect of waveform substitution on the quality of PCM packet communications,” *IEEE Trans. Acoust. Speech and Signal Processing*, vo. 36, No. 3, pp. 342-348, 1988.
- [31] K. Cluver and P. Noll, “Reconstruction of missing speech frames using sub-band excitation,” *Proc. IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analysis*, pp. 277-280, 1996.
- [32] J. Lindblom and P. Hedelin, “Packet loss concealment based on sinusoidal extrapolation,” *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 173-176, 2002.
- [33] E. Gunduzhan and K. Momtahan, “Linear prediction based packet loss concealment algorithm for PCM coded speech,” *IEEE Trans. Speech and Audio Processing*, vol. 9, No.8, pp. 778-785, Nov. 2001.
- [34] ITU-T Recommendation G.711 Appendix I, *A high quality low-complexity algorithm for packet loss concealment with G.711*, 1999.
- [35] ITU-T Recommendation G.711, *Pulse Code Modulation (PCM) of voice frequencies*, Nov. 1988.
- [36] R. V. Cox, and W. B. Kleijn, and P. Kroon, “Robust CELP coders for noisy backgrounds and noisy channels,” *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, pp. 739-742, 1989.
- [37] S. A. Atungsiri, A. M. Kondozi and B. G. Evans, “Error control for low-bit-rate speech communication systems,” *IEE Proc. I: Communications, Speech and Vision*, vol. 140, No. 2, pp. 97-103, April 1993.
- [38] 3GPP TS 26.090, “Adaptive Multi-Rate (AMR) speech codec; Transcoding functions,” version 5.0.0 (2002-06), 3rd Generation Partnership Project (3GPP).
- [39] 3GPP TS 26.091, “Adaptive Multi-Rate (AMR) speech codec; Error concealment of lost frames” version 5.0.0 (2002-06), 3rd Generation Partnership Project (3GPP).
- [40] ANSI/TIA/EIA-136-Rev.C part 410, “TDMA Cellular/PCS - Radio Interface, Enhanced Full-Rate Voice Codec (ACELP),” Formerly IS-641, TIA published standard, June 2001.
- [41] J. Wang and J. D. Gibson, “Parameter interpolation to enhance frame erasure robustness of CELP coders in packet networks,” *Proc. IEEE Int. Confs. Acoust., Speech and Sig. Proc.*, vol.2, pp.721-724, Salt Lake City, UT, May 2001.
- [42] F. Mertz, H. Taddei, I. Varga, and P. Vary, “Voicing controlled frame loss concealment for adaptive multi-rate (AMR) speech frames in voice-over-IP,” *Proc. 8th European Conf. Speech Commun. and Technology (Eurospeech)*, Geneva, Switzerland, Sept. 2003.
- [43] H. Sanneck, A. Stenger, K. Ben Younes, and B. Girod, “A new technique for audio packet loss concealment,” *IEEE Global Telecommunications Conference*, pp. 48-52, 1996.
- [44] Shacham, N.; McKenney, P., “Packet recovery in high-speed networks using coding and buffer management,” *Proc. IEEE INFOCOM*, vol.1, pp. 124-131, 1990.
- [45] D. Sridhara and T. Fuja, “Performance of the Federal Standard 2.4 kbps MELP vocoder over ATM networks,” *Proc. Conf. on Information Sciences and Systems*, Princeton, NJ, 2000.
- [46] J. C. Bolot, S. Fosse-Parisis and D. Towsley, “Adaptive FEC-based error control for Internet telephony,”

- Proc. IEEE INFOCOM*, vol. 3 , pp. 1453 -1460, 1999.
- [47] C. Perkins et. al., "RTP payload for redundant audio data," IETF RFC 2198, Sept. 1997.
- [48] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," IETF RFC2733, Dec. 1999.
- [49] N. S. Jayant and S. W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure," *IEEE Trans. Commun.*, vol. 29, No. 2, pp. 101-109, 1981.
- [50] Y.-L. Chen and B.-S. Chen, "Model-based multirate representation of speech signals and its application to recovery of missing speech packets," *IEEE Trans. Speech and Audio Proces.*, vol. 5, No.3 , pp. 220-231, 1997.
- [51] G. Kubin and W.B. Kleijn, "Multiple-description coding (MDC) of speech with an invertible auditory model," *Proc. IEEE Workshop on Speech Coding*, pp. 81 -83, 1999.
- [52] 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.
- [53] M. Yong, "Study of voice packet reconstruction methods applied to CELP speech coding," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2 , pp. 125-128, 1992.
- [54] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal Acoustical Society America*, vol.57, p. 535, Apr.1975
- [55] K. K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech and Audio Process.*, vol.1, no.1, pp. 3-14, 1993.
- [56] F. Lahouti and A. K. Khandani, "Quantization of LSF parameters using a trellis modeling," *IEEE Trans. Speech and Audio Proc.*, pp. 400-412, vol. 11, No. 5, 2003.
- [57] J. Chakareski, P. A. Chou, and B. Girod, "Rate-distortion optimized streaming from the edge of the network," *IEEE Workshop on Multimedia Signal Processing*, St. Thomas, US Virgin Islands, 2002.
- [58] J. Chakareski, P. A. Chou, B. Aazhang, "Computing rate-distortion optimized policies for streaming media to wireless clients," *Proc. Data Compression Conference*, pp. 53-62, 2002.
- [59] H. Sanneck, N. Le, M. Haardt and W. Mohr, "Selective packet prioritization for wireless VoIP," *Fourth Int. Symp. on Wireless Personal Multimedia Communication*, Aalborg, Denmark, September 2001.
- [60] P. C. Cosman, J. K. Rogers, P. G. Sherwood, and K. Zeger, "Combined forward error control and packetized zerotree Wavelet encoding for transmission of images over varying channels," *IEEE Trans. Image Proc.*, vol. 9, no. 6, June 2000.
- [61] C. S. Xydeas and C. Papanastasiou, "Split matrix quantization of LPC parameters," *IEEE Trans. Speech and Audio Process.*, vol. 7, no. 2, pp. 113-125, 1999.
- [62] L. E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, pp. 164-171, 1970.
- [63] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84-95, 1980.