# Multiple Descriptions and Missing Data Estimation for Voice over Packet-Switched Networks

*Rainer Martin and Frank Mertz*
*Institute of Communication Systems and Data Processing*
*Aachen University of Technology, 52056 Aachen*
*Phone: +49 241 806984, Fax: +49 241 8888 186*
*Email: {martin | mertz}@ind.rwth-aachen.de*

**Abstract**

In this contribution we consider the transmission of voice over packet-switched networks using Multiple Description encoding and optimal estimation of missing information. We investigate Multiple Description encoding techniques for the transmission of spectral parameters of a speech coder and present an approach for estimating lost parameters. The reconstruction of missing information is based on intra-frame Minimum Mean Square Error (MMSE) estimation with *a priori* knowledge modelled by Gaussian Mixture Models (GMM). We show that the GMM approach leads to implementations which are efficient with respect to computational complexity and memory requirements. Experimental results are given for the estimation of missing Line Spectral Frequency (LSF) parameters.

## 1. Introduction

Future telecommunications networks will most likely be based on a packet-switched *"all-IP"* architecture with the benefits of utilizing a single network for all telecommunication services and integrating traditionally separated services into powerful *unified* applications. This also means that the plain old telephone service will migrate from circuit-switched to packet-switched technology. As packet-switched networks such as the Internet suffer from congestion, packet delays, and packet loss, network protocols and speech transmission technology must be adapted to achieve at least the same quality as with the familiar circuit-switched networks.

In this contribution we investigate methods for the transmission of coded speech parameters over packet-switched networks. Since packet losses on the Internet are bursty and speech transmission cannot tolerate large transmission delays traditional error mitigation techniques are not always successful. The application of forward error correction (FEC) methods in an environment with packet loss will require the interleaving of several speech frames and hence will increase the transmission delay. For the same reason, automatic repeat request (ARQ) methods are not feasable. We therefore focus on Multiple Descriptions (MD) encoding methods. Multiple Descriptions systems send two (or more) individually good descriptions of the speech source via independent channels, hoping that at least one description is received. However, when all descriptions are received the combined information of these descriptions should be as large as possible. Thereby, the method maximizes the quality in the error-free case while at the same time it adds redundancy for improving the performance on channels with packet losses. We explore the benefits of such a scheme for the Line Spectral Frequency (LSF) parameters of a speech coder in conjunction with Minimum Mean Square Error (MMSE) estimators for

Tagungsband Elektronische Sprachsignalverarbeitung, Bonn, 2001

the lost information. Our objective is to use the redundancy within a single speech frame to derive two or three descriptions of the spectral speech envelope and to restore lost information. The design of optimal estimators for spectral parameters is a challenging task in itself because these parameters are correlated over time and frequency, and the representation of prior densities using histograms [1] leads to prohibitively large memory requirements. We therefore employ an approach based on Gaussian Mixture Models (GMM) [2] and compare results for various encoding and estimation strategies. As the Adaptive Multi-Rate speech coder [3, 4] is a candidate for Voice over IP applications [5], the robust transmission of spectral parameters of the AMR coder will serve as an example throughout this paper.
The remainder of this paper is organized as follows: After a brief review of the quantization of LSF parameters and Multiple Description encoding in Sections 2 and 3, respectively, we turn to the development of MMSE estimators using Gaussian Mixture Models in Section 4. We conclude with a presentation of experimental results in Section 5.

## 2. Quantization of Spectral Parameters in the AMR Coder

The GSM AMR speech coder implements eight different source coding modes at bitrates between 4.75 kbit/s and 12.2 kbit/s [6]. All modes use a filter of order 10 for linear prediction (LP) analysis. The 12.2 kbit/s mode (which is identical to the GSM enhanced fullrate coder) uses two different windows to calculate two sets of LSF parameters. All other modes use a single window to calculate one set of coefficients for each speech frame of 20 ms. The LP coefficients are converted to the LSF domain and differentially encoded. To remove inter-frame correlation a predicted LSF vector is subtracted from each incoming LSF vector. The predicted LSF vector is computed using a first order linear prediction filter with fixed prediction coefficients. For modes below 12.2 kbit/s the residual LSF vectors are then Split Vector (SVQ) quantized. The 10-dimensional vectors are partitioned into subsets of 3, 3, and 4 residual LSF coefficients. Each of these subsets is then vector quantized with 7 to 9 bits. E.g., for the 10.2 kbit/s mode the first, the second, and the third subset are quantized with 8, 9, and 9 bits, respectively. The direct histogram-based implementation of an MMSE estimator for missing spectral information therefore requires large memory and computational resources. For the 10.2 kbit/s mode (which we used in this study) the joint statistics (histogram) of the quantized LSF coefficients would require the storage of $2^8 \cdot 2^9 \cdot 2^9 \approx 67 \cdot 10^6$ values. For a less complex but suboptimal Markov chain approach the transition matrices between subsets need to be stored. For the 10.2 kbit/s mode this still amounts to $2^8 \cdot 2^9 + 2^9 \cdot 2^9 = 393{,}216$ histogram values. We therefore propose to use Gaussian Mixture Models to represent the joint statistics of LSF vectors. The improvements which can be achieved by optimal estimators are directly related to the correlation within the parameter vectors. Figure 1 shows on the left hand side an intensity plot of the correlation coefficient matrix of the LSF vectors, computed on a data set of 20,000 LSF vectors. The right hand side of Fig. 1 depicts the inter-frame correlation of the LSF vectors. The inter-frame correlation can be used to conceal the loss of single speech frames. This is in fact exploited in the bad frame mechanism of the AMR coder. Our MD coding and estimation scheme, however, aims at using the intra-frame correlation.

## 3. Multiple Description Encoding

Multiple Descriptions (MD) source coding is a diversity based transmission scheme where information about the source is transmitted over multiple possibly independent channels.

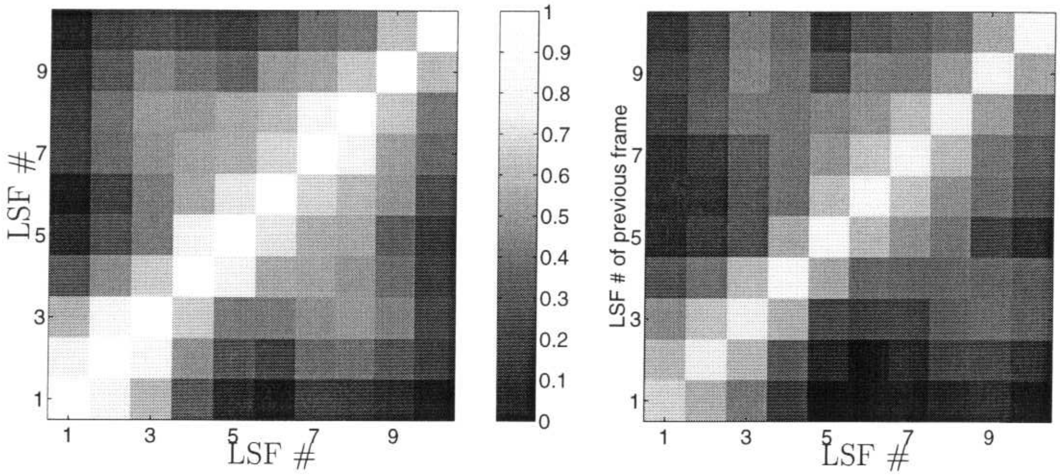


Figure 1: Intensity plot of intra-frame (left) and inter-frame (right) correlation coefficient matrices for LSF vectors.

From the perspective of information theory it can be considered as a generalization of the source coding problem subject to a fidelity criterion. Figure 2 depicts the basic idea for a two channel system. The source is encoded using encoder 1 with rate $R_1$ and encoder 2 with rate $R_2$. The encoded symbols of encoder 1 and encoder 2 are transmitted over channels 1 and 2, respectively. These channels can be physically independent channels like two different wireline or radio links but can be also two independent data streams over the same physical channel. At the receiver side there are three decoders: the central decoder 0 which has access to the output symbols of both channels and the two side decoders 1 and 2 which use the output symbols of just one channel. The two encoders and the three decoders are designed to meet distortion constraints $D_0$, $D_1$, and $D_2$.

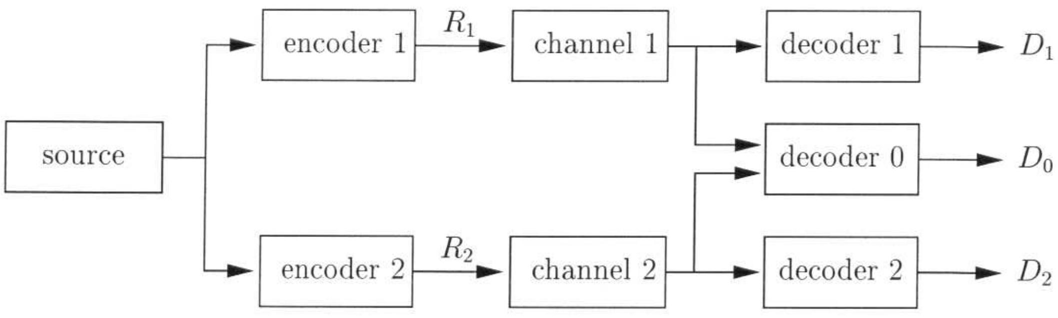


Figure 2: A two channel Multiple Description system.

If we send the same information on both channels half of the information is useless when both channels work. This is similar to a repetition code. There is no improvement with respect to the case when only one channel is received. The idea of MD encoding is to send different information on both channels such that a minimum fidelity is achieved when only one channel works and a higher fidelity is achieved when both channels work. This is in fact a joint source/channel coding problem and properly designed MD encoders/decoders can exhibit the "graceful degradation" feature known from joint source/channel coding systems.
The basic difficulty in the design of MD encoders/decoders lies in the fact that when a low distortion should be achieved for the single channel case both descriptions must be

correlated since they encode the same source. On the other hand, when the descriptions are correlated, the gain of combining both descriptions in the error free case is also limited, as little additional information is available. This dilemma has been quantified, e.g., in the theorem on the region of achievable rates by El Gamal and Cover [7].
The MD idea has been employed, e.g., in audio streaming over the Internet. The Robust Audio Tool (RAT) [8] piggybacks redundant information about the current packet onto a later scheduled packet. When a packet is lost this second description can then be used to restore the missing information at a lower fidelity. In our investigation we split the LSF parameters into two or three subsets and assume that all subsets are transmitted independently. If there are no packet losses the spectral information can be reassembled at the receiver to obtain full fidelity. If packets and spectral information is lost we estimate the missing information based on the received information and *a priori* knowledge.

## 4. MMSE Estimation of Missing LSF Parameters

We assume that some components of the current LSF vector $\underline{L}_k$ are lost in several successive frames and that all other components are received without error. The aim of the estimation procedure is to restore the missing components using MMSE estimation. The approach taken here is also known as "Missing Feature Theory" which has been employed in robust speech recognition [9].
We partition the current LSF vector $\underline{L}_k$ into a received (or present) part $\underline{L}_k^{(p)}$ and a lost (or missing) part $\underline{L}_k^{(m)}$,

$$\underline{L}_k = \begin{pmatrix} \underline{L}_k^{(m)} \\ \underline{L}_k^{(p)} \end{pmatrix} . \tag{1}$$

If quantization errors are negligible, the MMSE estimate of the missing components is given by the conditional expectation

$$\widehat{\underline{L}}_k^{(m)} = E\{\underline{L}_k^{(m)} \mid \underline{L}_k^{(p)}, \underline{Z}_{k-1}\} \tag{2}$$

as a function of the present components and the sequence of all previously received LSF vectors $\underline{Z}_{k-1}$. If we assume that no information is available from previous vectors (burst errors) the expression can be further simplified

$$\widehat{\underline{L}}_k^{(m)} = \int_{\underline{L}_k^{(m)}} \underline{L}_k^{(m)} p(\underline{L}_k^{(m)} \mid \underline{L}_k^{(p)}) d\underline{L}_k^{(m)} . \tag{3}$$

To compute the optimal estimate the conditional probability density $p(\underline{L}_k^{(m)} \mid \underline{L}_k^{(p)})$ must be known.

### 4.1 Gaussian Mixture Models

Mixture models are frequently used in data classification and clustering problems with the aim of fitting the probability density function (pdf) of some given data. We approximate the joint probability density of the LSF vectors by a Gaussian Mixture Model [10], i.e., by a sum of $M$ multivariate Gaussian densities

$$p(\underline{L}_k) = \sum_{i=1}^{M} \alpha_i \mathfrak{N}(\underline{L}_k, \underline{\mu}_i, \mathbf{C}_i) \tag{4}$$

where each $N$-dimensional mixture density is given by

$$\mathfrak{N}(\underline{L}_k, \underline{\mu}_i, \mathbf{C}_i) = \frac{1}{\sqrt{(2\pi)^N \mid \mathbf{C}_i \mid}} \cdot \exp(-\frac{1}{2}(\underline{L}_k - \underline{\mu}_i)^T \mathbf{C}_i^{-1}(\underline{L}_k - \underline{\mu}_i)) \tag{5}$$

and $\alpha_i$ denotes the *a priori* probability of the mixture components $\mathfrak{N}_i = \mathfrak{N}(\underline{L}_k, \underline{\mu}_i, \mathbf{C}_i)$, i.e. $P(\mathfrak{N}_i) = \alpha_i$.
The mixture probabilities $\alpha_i$, the mixture mean vectors $\underline{\mu}_i$, and the covariance matrices $\mathbf{C}_i$ are determined from training the model by means of the well known Estimate-Maximize (EM) algorithm [11]. To reduce the number of free parameters it is common to use covariance matrices with non-zero elements on the main diagonal only [12, 10].

### 4.2 MMSE Estimation Using GMM

In order to compute an approximate MMSE estimate using the GMM, we partition all parameters of the GMM with respect to present and missing components. Analogous to the LSF parameter vector in (1) the mean vectors $\underline{\mu}_i$ and the covariances $\mathbf{C}_i$ of all mixture components can then be written as follows

$$\underline{\mu}_i = \begin{pmatrix} \underline{\mu}_i^{(m)} \\ \underline{\mu}_i^{(p)} \end{pmatrix} , \tag{6}$$

$$\mathbf{C}_i = \begin{pmatrix} \mathbf{C}_i^{(m,m)} & \mathbf{C}_i^{(m,p)} \\ \mathbf{C}_i^{(p,m)} & \mathbf{C}_i^{(p,p)} \end{pmatrix} . \tag{7}$$

The conditional pdf of present and missing components can now be expressed in terms of a GMM

$$p(\underline{L}_k^{(m)} \mid \underline{L}_k^{(p)}) = \frac{p(\underline{L}_k^{(m)}, \underline{L}_k^{(p)})}{p(\underline{L}_k^{(p)})} = \sum_{i=1}^{M} \frac{\alpha_i}{p(\underline{L}_k^{(p)})} \mathfrak{N}(\underline{L}_k, \underline{\mu}_i, \mathbf{C}_i) . \tag{8}$$

Using (3) an approximate MMSE estimate of $\underline{L}_k^{(m)}$ can be written as [2]

$$\widehat{\underline{L}}_k^{(m)} = E\{\underline{L}_k^{(m)} \mid \underline{L}_k^{(p)}\} = \sum_{i=1}^{M} \alpha_i^{(m|p)} \underline{\mu}_i^{(m|p)} \tag{9}$$

where $\alpha_i^{(m|p)}$ and $\underline{\mu}_i^{(m|p)}$ are given by:

$$\alpha_i^{(m|p)} = \frac{\alpha_i \mathfrak{N}(\underline{L}_k^{(p)}, \underline{\mu}_i^{(p)}, \mathbf{C}_i^{(p,p)})}{\sum_{\ell=1}^{M} \alpha_\ell \mathfrak{N}(\underline{L}_k^{(p)}, \underline{\mu}_\ell^{(p)}, \mathbf{C}_\ell^{(p,p)})} \qquad \underline{\mu}_i^{(m|p)} = \underline{\mu}_i^{(m)} + \mathbf{C}_i^{(p,m)}(\mathbf{C}_i^{(p,p)})^{-1}(\underline{L}_k^{(p)} - \underline{\mu}_i^{(p)}) \tag{10}$$

If the GMM models the pdf of the current LSF vector $\underline{L}_k$ by means of *diagonal* covariance matrices the off-diagonal matrices $\mathbf{C}_i^{(m,p)}$ and $\mathbf{C}_i^{(p,m)}$ are all zero and the MMSE estimate is further simplified

$$\widehat{\underline{L}}_k^{(m)} = \sum_{i=1}^{M} \alpha_i^{(m|p)} \underline{\mu}_i^{(m)} . \tag{11}$$

The *a posteriori* probabilities can now be computed using products of univariate normal densities

$$\alpha_i^{(m|p)} = \frac{\alpha_i \exp(-0.5 \sum_j (L_{k,j}^{(p)} - \mu_{i,j}^{(p)})^2/(\sigma_{i,j}^{(p,p)})^2) \prod_j 1/\sigma_{i,j}^{(p,p)}}{\sum_{\ell=1}^{M} \alpha_i \exp(-0.5 \sum_j (L_{k,j}^{(p)} - \mu_{\ell,j}^{(p)})^2/(\sigma_{\ell,j}^{(p,p)})^2) \prod_j 1/\sigma_{\ell,j}^{(p,p)}} \; . \tag{12}$$

$L_{k,j}^{(p)}$ denotes the $j$-th component of the $k$-th vector $\underline{L}_k^{(p)}$ of present components, and $\mu_{i,j}^{(p)}$ and $(\sigma_{i,j}^{(p,p)})^2$ are the mean and the variance of the $j$-th vector component and the $i$-th mixture component, respectively. $\sum_j = \sum_{j=1}^{N_p}$ and $\prod_j = \prod_{j=1}^{N_p}$ denote sums and products over the present components where $N_p$ denotes the number of present components.
The memory requirements of the GMM approach are directly proportional to the dimension of the LSF vector and the GMM order, i.e for LSF vectors of size 10 and GMM's with diagonal covariance matrices $(10 + 10 + 1) \cdot M$ values must be stored. Since $M$ is typically much smaller than 100 the GMM approach offers significant memory advantages with respect to the histogram approach.

## 5. Experimental Results

The AMR coder groups the LSF parameters into three subsets. Each of these subsets is then vector quantized and transmitted independently. In our experiments we assume that one or two of these data streams are disrupted. More specifically, we consider three different scenarios:

- One of the three subsets is permanently lost;
- Two of the three subsets are permanently lost;
- Even or odd numbered LSF parameters are permanently lost.

All these loss scenarios assume a burst error situation, i.e., that the parameters of more than one consecutive frames are lost. The third scenario is not compatible with the AMR coder since it assumes that the LSF parameters are grouped into only two subsets. The first subset contains LSF # 1,3,5,7,9 and the other subset contains LSF # 2,4,6,8,10. If one of these subsets is lost the missing information is highly correlated with the present information and can be thus recovered with high accuracy.
The experimental results were obtained using GMM's with $M = 1, 2, 4, 8, 16, 32$, and 64 mixture components. If only one component is used the estimate is equivalent to the *a priori* mean of the missing LSF components. The use of the *a priori* mean is also termed *mean imputation*. The GMM's were trained by means of the EM algorithm and 900,000 LSF vectors. A data base of male and female speech and the AMR coder (10.2 kbit/s mode) was used to generate the LSF vectors. The estimation algorithm was evaluated using 20,000 LSF vectors which were not part of the training data base.

### 5.1 Performance Measures

The performance of the estimators is measured by evaluating the parameter SNR and the spectral distortion. The SNR of the $j$-th parameter is given by

$$SNR_j = 10 \log_{10} \frac{\sum_k (L_{k,j}^{(m)})^2}{\sum_k (\widehat{L}_{k,j}^{(m)} - L_{k,j}^{(m)})^2} \; . \tag{13}$$

The resulting SNR value is then normalized on the SNR which is achieved for mean imputation. The spectral distortion is defined as

$$SD = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[10\log_{10}\left(\frac{|H(\Omega)|^2}{|\widehat{H}(\Omega)|^2}\right)\right]^2 d\Omega} \quad . \tag{14}$$

The latter can be efficiently computed using cepstral coefficients [13]. The reference frequency response $H(\Omega)$ was computed using unquantized LSF vectors.

### 5.2 Loss of a single LSF subset

The improvement in parameter SNR with respect to *a priori* mean imputation is shown in Table 1. The gain depends mainly on the position within the LSF vector and the correlation to adjacent LSF's. For $M = 64$ an average gain of about 3 dB is obtained. Further doubling the GMM order did result in small improvements but also in a significantly increased computational complexity.
Table 2 summarizes the spectral distortion which is achieved for GMM order $M$. We note that a substantial improvement with respect to mean imputation $M = 1$ is obtained.

| lost | LSF | GMM Model Order $M$ | | | | | |
|---|---|---|---|---|---|---|---|
| set | # | 2 | 4 | 8 | 16 | 32 | 64 |
| | 1 | -0.09 | 0.30 | 0.81 | 1.31 | 1.55 | 1.83 |
| 1 | 2 | 0.28 | 0.97 | 1.72 | 2.14 | 2.48 | 2.66 |
| | 3 | 1.17 | 2.09 | 3.24 | 3.83 | 4.46 | 4.64 |
| | 4 | 2.20 | 2.75 | 3.24 | 3.31 | 3.84 | 4.08 |
| 2 | 5 | 0.82 | 1.03 | 1.95 | 2.75 | 3.10 | 3.43 |
| | 6 | 0.97 | 1.94 | 3.12 | 3.38 | 3.81 | 4.11 |
| | 7 | 1.24 | 1.42 | 2.53 | 2.52 | 2.50 | 2.71 |
| 3 | 8 | 0.97 | 1.19 | 1.86 | 1.73 | 1.59 | 1.80 |
| | 9 | 1.19 | 1.60 | 2.18 | 2.30 | 2.64 | 2.90 |
| | 10 | 0.44 | 0.80 | 1.27 | 1.84 | 1.99 | 2.14 |

Table 1: Improvement of parameter SNR in dB with respect to *a priori* mean imputation for the estimation of a single lost LSF subset using a GMM of order $M$.

| lost | GMM Model Order $M$ | | | | | | |
|---|---|---|---|---|---|---|---|
| set | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
| 1 | 3.47 | 3.28 | 2.89 | 2.58 | 2.42 | 2.33 | 2.27 |
| 2 | 4.26 | 3.56 | 3.26 | 3.00 | 2.83 | 2.71 | 2.62 |
| 3 | 3.44 | 3.07 | 2.95 | 2.67 | 2.60 | 2.57 | 2.50 |

Table 2: Spectral distortion (SD) in dB for the estimation of a single lost LSF subset using a GMM of order $M$.

### 5.3 Loss of two LSF subsets

When two subsets are lost the improvements which can be obtained by our estimation procedure are significantly lower. The SNR measurements are summarized in Table 3. Good results are obtained only for those LSF coefficients which are adjacent to the present components. This is also reflected in the spectral distortion measurements (see Table 4).

| lost set | LSF # | GMM Model Order $M$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 8 | 16 | 32 | 64 |
| | 1 | -0.06 | -0.24 | 0.09 | 0.19 | 0.09 | 0.07 |
| 1 | 2 | 0.37 | 0.35 | 0.72 | 0.79 | 0.76 | 0.71 |
| | 3 | 1.07 | 1.20 | 1.71 | 1.82 | 1.87 | 1.87 |
| | 4 | 1.38 | 1.69 | 1.96 | 1.97 | 2.03 | 2.10 |
| 2 | 5 | 1.31 | 1.74 | 1.94 | 2.22 | 2.25 | 2.37 |
| | 6 | 1.55 | 2.04 | 2.33 | 2.22 | 2.42 | 2.58 |
| | 4 | 2.02 | 2.65 | 2.81 | 2.92 | 3.25 | 3.38 |
| 2 | 5 | -0.04 | 0.52 | 0.75 | 0.94 | 1.11 | 1.20 |
| | 6 | -0.27 | 0.69 | 1.48 | 1.84 | 2.08 | 2.15 |
| | 7 | 0.29 | 0.55 | 0.85 | 0.96 | 1.01 | 1.11 |
| 3 | 8 | 0.14 | 0.47 | 0.80 | 0.96 | 0.92 | 1.05 |
| | 9 | 0.21 | 0.70 | 1.13 | 1.30 | 1.42 | 1.52 |
| | 10 | 0.22 | 0.57 | 0.90 | 1.16 | 1.15 | 1.15 |
| | 1 | 0.07 | 0.84 | 1.26 | 1.55 | 1.64 | 1.73 |
| 1 | 2 | 0.37 | 1.41 | 1.80 | 2.08 | 2.31 | 2.29 |
| | 3 | 1.25 | 2.53 | 3.00 | 3.55 | 4.01 | 4.06 |
| | 7 | 1.82 | 2.16 | 2.83 | 2.74 | 2.64 | 2.68 |
| 3 | 8 | 1.59 | 1.69 | 2.00 | 1.83 | 1.78 | 1.75 |
| | 9 | 1.54 | 1.63 | 2.15 | 2.18 | 2.15 | 2.37 |
| | 10 | 0.42 | 0.64 | 1.00 | 1.28 | 1.37 | 1.53 |

Table 3: Improvement of parameter SNR in dB with respect to *a priori* mean imputation for the estimation of two lost LSF subsets using a GMM of order $M$.

| lost sets | GMM Model Order $M$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
| 1,2 | 5.56 | 4.99 | 4.79 | 4.57 | 4.54 | 4.54 | 4.51 |
| 2,3 | 5.64 | 5.28 | 4.94 | 4.72 | 4.60 | 4.52 | 4.48 |
| 1,3 | 5.10 | 4.58 | 4.19 | 3.95 | 3.83 | 3.78 | 3.75 |

Table 4: Spectral distortion (SD) in dB for the estimation of two lost LSF subsets using a GMM of order $M$.

### 5.4 Loss of even or odd numbered LSF coefficients

The success of the GMM based estimator clearly depends on the correlation between present and missing LSF components. An interesting variation of the above schemes is therefore to group the LSF parameters into two interleaved subsets. The first subset (denoted as "even" in Table 5) contains LSF # 2, 4, 6, 8, 10, and the second subset

| lost set | LSF # | GMM Model Order $M$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2 | 4 | 8 | 16 | 32 | 64 |
| even | 2 | 1.05 | 4.36 | 5.57 | 6.52 | 7.75 | 8.14 |
| | 4 | 3.20 | 4.00 | 4.75 | 5.22 | 5.68 | 5.87 |
| | 6 | 2.26 | 3.16 | 5.02 | 5.35 | 5.89 | 6.44 |
| | 8 | 2.06 | 1.71 | 2.88 | 3.31 | 3.68 | 4.13 |
| | 10 | 0.73 | 0.98 | 1.42 | 2.40 | 2.59 | 2.83 |
| odd | 1 | 0.25 | 2.41 | 3.54 | 4.38 | 5.16 | 5.61 |
| | 3 | 2.30 | 5.16 | 6.06 | 6.68 | 7.58 | 7.59 |
| | 5 | 2.24 | 2.72 | 4.31 | 6.22 | 6.52 | 6.91 |
| | 7 | 1.78 | 2.06 | 3.22 | 3.53 | 3.98 | 4.33 |
| | 9 | 2.14 | 2.11 | 2.75 | 3.47 | 3.89 | 4.34 |

Table 5: Improvement of parameter SNR in dB with respect to *a priori* mean imputation for the estimation of "even" or "odd" LSF subsets using a GMM of order $M$.

| lost set | GMM Model Order $M$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
| even | 6.15 | 4.36 | 3.74 | 3.11 | 2.80 | 2.61 | 2.46 |
| odd | 5.33 | 3.94 | 3.36 | 2.96 | 2.62 | 2.48 | 2.39 |

Table 6: Spectral distortion (SD) in dB for the estimation of "even" or "odd" LSF subsets using a GMM of order $M$.

(denoted as "odd" in Table 5) contains LSF # 1, 3, 5, 7, 9. This scheme is not compatible with the AMR coder but leads, as is shown in Tables 5 and 6, to significant improvements.

## 6. Conclusions

The proposed LSF reconstruction scheme was implemented in the AMR coder. It enhances the quality of the received speech signals when bursty frame losses occur. The scheme is especially useful when no more than one LSF subset gets lost. Compared to histogram based approaches the GMM based approach consumes significantly less memory.
The improvements obtained for missing LSF subsets can be increased to the values given in Table 6 when we deviate from the AMR transmission scheme and group LSF # 1, 3, 5, 7, 9 into subset one and LSF # 2, 4, 6, 8, 10 into subset two. Such a transmission scheme will reduce the efficiency of the vector quantizer but will significantly improve the robustness in the presence of frame erasures. The Multiple Description problem discussed here is another example of joint source/channel coding where we have to find a good balance between the redundancy in the source parameters and the redundancy added by the error control mechanism. It is not quite obvious whether the best possible source coding (vector quantization in this case) combined with less effective error concealment techniques or suboptimal vector quantization combined with more effective parameter estimation techniques will achieve the better subjective quality. These issues are currently studied in more detail.

## References


[1] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Trans. Speech and Audio Processing*, vol. 9, pp. 240–251, mar 2001.

[2] R. Martin, C. Hoelper, and I. Wittke, "Estimation of Missing LSF Parameters Using Gaussian Mixture Models," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 2, 2001.

[3] ETSI, *EN 301 703, V7.0.2: Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR) speech processing functions; General Description*, December 1999.

[4] S. Bruhn, E. Ekudden, and K. Hellwig, "Adaptive Multi-Rate: A New Speech Service for GSM and Beyond," in *Proc. 3rd ITG Conference, Source and Channel Coding*, vol. 159, (Munich), pp. 319–324, VDE Verlag, 2000.

[5] J. Sjoberg et al., *RTP Payload Format and File Storage Format for AMR and AMR-WB Audio.* Draft, Internet Engineering Task Force, June 11, 2001.

[6] ETSI, *GSM 06.71: Digital cellular telecomunications system (Phase 2+); Adaptive Multi-Rate (AMR); Speech Processing Functions; General Description.* European Telecommunications Standards Institute, 1998.

[7] A. A. El Gamal and T. Cover, "Achievable Rates for Multiple Descriptions," *IEEE Trans. Information Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[8] V. Hardman, M. Sasse, and I. Kouvelas, "Successful Multiparty Audio Communication over the Internet," *Communications of the ACM*, vol. 41, no. 5, pp. 74–80, 1998.

[9] A. Morris, M. Cooke, and P. Green, "Some Solutions to the Missing Feature Problem in Data Classification, with Application to Noise Robust ASR," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pp. 737–740, 1998.

[10] P. Hedelin and J. Skoglund, "Vector Quantization Based on Gaussian Mixture Models," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 4, pp. 385–401, 2000.

[11] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc.*, vol. 39, pp. 1–38, 1977.

[12] D. Reynolds and R. Rose, "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models," *SAP*, vol. 3, no. 1, pp. 72–83, 1995.

[13] R. Hagen, "Spectral quantization of cepstral coefficients," in *Proc. Int. Conf. Acoust., Speech, Signal Processing, ICASSP*, pp. (I) 509–512, IEEE, 1994.