ERROR CONCEALMENT BY NEAR OPTIMUM MMSE-ESTIMATION OF SOURCE CODEC PARAMETERS

M. Adrat, J. Spittka, S. Heinen, and P. Vary

Institute of Communication Systems and Data Processing
Aachen University of Technology
{adrat,heinen,vary}@ind.rwth-aachen.de

ABSTRACT

In digital communications source coding is indispensable to achieve a high bandwidth efficiency in applications where bandwidth is a limited resource. Usually these source coding algorithms determine speech or audio parameters which are highly sensitive to transmission errors.

This paper deals with an error concealment technique that benefits from residual redundancy remaining after source coding. Due to delay and complexity constraints a certain amount of inter-frame as well as intra-frame correlation between source codec parameters remains which might be exploited to enhance the subjective audio quality.

In this contribution we propose an algorithm which is able to utilize both types of redundancy if the source parameters exhibit a two-dimensional Markov property. Actually, this approach is nearly optimal in the *Minimum Mean Square Error* (MMSE) sense but much less complex compared to the optimal estimator.

1. INTRODUCTION

Error concealment techniques are necessary if transmission errors may occur and if channel codes are not able to remove all of them. Usually, due to fading, multipath propagation and Doppler spread transmission errors are almost unavoidable in systems like digital audio broadcasting (DAB) or digital mobile communications (GSM). Speech or audio parameters determined by source coding algorithms are very vulnerable against channel noise and hence residual bit errors in the received parameters can result in extremely annoying artifacts.

In general, source codecs reduce redundancy, but due to delay and complexity constraints the source parameters will still exhibit considerable redundancy, either in terms of a non-uniform distribution or due to inter- and intra-frame correlation. As already indicated by Shannon [1] this residual redundancy can be exploited at the receiver to enhance the disturbed signal.

Recently, we have proposed an approach to error concealment by softbit parameter estimation [2, 3, 4]. Reliability information gained from a soft-output channel decoder [5, 6] is combined with *a-priori* knowledge about statistical properties of the source to estimate the codec parameters.

In the first approaches [2, 3] parameter estimation was exclusively based on the redundancy due to the non-uniform distribution and the inter-frame correlation. But measurements of the statistical properties of frame-oriented compression algorithms show that also significant intra-frame correlation is available [4]. For example in modern speech codecs like the GSM-AMR (*adaptive multi rate*) *split vector quantization* is applied to a vector of *line spectral frequencies* (LSF). Between subsequent subvectors high correlation is noticable. Another example are the scale factors of audio transform codecs like in DAB.

The paper is structured as follows. Firstly, we briefly describe a transmission system using softbit parameter estimation. In Section 3, we introduce the new near optimum MMSE approach which utilizes inter-frame as well as intra-frame redundancy. In Section 4 we compare our estimator to the optimal MMSE estimation algorithm [4] and to the approaches described in [2, 3]. The comparison will be done with respect to parameter SNR and complexity. Finally, we discuss some applications in Section 5.

2. PARAMETER ESTIMATION

For further considerations the transmission model depicted in Figure 1 is assumed.

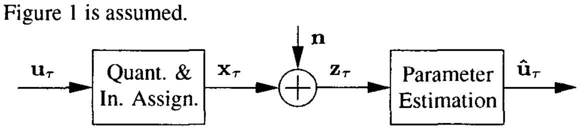

Figure 1: Transmission model

At time index τ a source encoder determines a set of N parameters $u_{\tau,\kappa}$ with $\kappa = 1, 2, \ldots, N$ representing the index of the vector element according to $\mathbf{u}_\tau = \{u_{\tau,1}, u_{\tau,2}, \ldots, u_{\tau,N}\}$. The continuous but time discrete values $u_{\tau,\kappa}$ are individually quantized by R_κ reproduction levels $\bar{u}_\kappa^{(i)}$ with $i = 1, 2, \ldots, R_\kappa$. The reproduction levels $\bar{u}_\kappa^{(i)}$ are dependent on κ but not on τ. A whole set of reproduction levels is given by $\mathbb{U}_\kappa = \{\bar{u}_\kappa^{(1)}, \bar{u}_\kappa^{(2)}, \ldots, \bar{u}_\kappa^{(R_\kappa)}\}$. To each $\bar{u}_\kappa^{(i)}$ a unique bit pattern $x_\kappa^{(i)}$, $i = 1, 2, \ldots, R_\kappa$, is assigned where the length w_κ of $x_\kappa^{(i)}$ is usually[1] given by $w_\kappa = \log_2(R_\kappa)$. The bit pattern determined from $u_{\tau,\kappa}$ at time index τ and position κ is denoted by $x_{\tau,\kappa}^{(i)}$. If two or more of these $u_{\tau,\kappa}$, $x_{\tau,\kappa}^{(i)}$ respectively, exhibit correlation properties in position κ then intra-frame redundancy is utilizable and otherwise if they exhibit correlation properties in time τ inter-frame redundancy is available.

At time index τ a set of bit combinations $\mathbf{x}_\tau = \{x_{\tau,1}^{(i)}, x_{\tau,2}^{(j)}, \ldots, x_{\tau,N}^{(k)}\}$, with $i = 1, 2, \ldots, R_1$, $j = 1, 2, \ldots, R_2$, $\ldots$, $k = 1, 2, \ldots, R_N$, is transmitted over a channel with additive noise $\mathbf{n}$ and a possibly disturbed set $\mathbf{z}_\tau$ is received. The task of the parameter estimation algorithm is to determine an estimate $\hat{\mathbf{u}}_\tau$ for the source encoded set $\mathbf{u}_\tau$. The MMSE optimality criterion is used individually for each parameter $E\{(u_{\tau,\kappa} - \hat{u}_{\tau,\kappa})^2\} \rightarrow \min$. $E\{\cdot\}$ denotes the expectation value.

[1]If R_κ is not a power of 2, w_κ might be given by the minimum integral number $w_\kappa = \lceil \log_2(R_\kappa) \rceil$. In this case *Source Optimized Channel Codes* [7] can benefit from the redundancy in bit mapping, $\bar{u}_\kappa^{(i)} \rightarrow x_\kappa^{(i)}$.

0-7803-6416-3/00/$10.00©2000 IEEE

Now, the estimation process is designed such that both types of redundancy, intra-frame as well as inter-frame, are exploited. It is obvious that due to the inter-frame redundancy the entire history of received sets $\mathbf{z}_\tau, \mathbf{z}_{\tau-1}, \ldots, \mathbf{z}_1$ must be taken into account. To simplify notation furtheron a sequence of sets is denoted by $\mathbf{z}_1^\tau = \{\mathbf{z}_\tau, \mathbf{z}_{\tau-1}, \ldots, \mathbf{z}_1\}$.

The well known MMSE estimation rule can be written as

$$\hat{u}_{\tau,\kappa} = \sum_{i=1}^{R_\kappa} \bar{u}_{\tau,\kappa}^{(i)} \cdot P(\bar{u}_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau) \quad . \tag{1}$$

$P(\bar{u}_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau)$ is the conditional probability for reproduction level $\bar{u}_{\tau,\kappa}^{(i)}$ at time index τ and position κ when the entire history of the received sets $\mathbf{z}_1^\tau$ is given. Due to the fixed index assignment $\bar{u}_\kappa^{(i)} \to x_\kappa^{(i)}$ $P(\bar{u}_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau)$ is equal to the conditional probability $P(x_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau)$. Hence, the problem of estimating the optimal parameter values $\hat{u}_{\tau,\kappa}$ can be reduced to the problem of determining the conditional probabilities $P(x_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau)$ with $i = 1, 2 \ldots R_\kappa$.

3. NEAR OPTIMUM MMSE ESTIMATION

The formal solution for determining $P(x_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau)$ is given by the marginal probability of $P(\mathbf{x}_1^\tau | \mathbf{z}_1^\tau)$ for $x_{t,k} = x_{\tau,\kappa}^{(i)}$,

$$P(x_{\tau,\kappa}^{(i)} | \mathbf{z}_1^\tau) = \sum_{\forall x_{t,k} \epsilon \mathbb{X}_{1,1}^{\tau,N},\ t,k \neq \tau,\kappa} P(\mathbf{x}_1^\tau | \mathbf{z}_1^\tau) \quad . \tag{2}$$

$\mathbb{X}_{1,1}^{\tau,N}$ denotes the set of $x_{t,k}$ forall possible combinations of t, k.

Of course this formal solution is by far too complex. To reduce the complexity we consider the entire history of transmitted bits $x_{1,\kappa}^\tau$ and received bits $z_{1,\kappa}^\tau$ and we take only the entire histories of adjacent parameters $x_{1,\kappa}^\tau, x_{1,\kappa-1}^\tau$ and $x_{1,\kappa+1}^\tau$ into account

$$\begin{aligned} &P(x_{\tau,\kappa}^{(i)} | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau) \\ &\quad = \sum_{\forall x_{t,k} \epsilon \mathbb{X}_{1,\kappa-1}^{\tau,\kappa+1},\ t,k \neq \tau,\kappa} P(x_{1,\kappa}^\tau, x_{1,\kappa-1}^\tau, x_{1,\kappa+1}^\tau | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau). \end{aligned} \tag{3}$$

Furthermore, the marginal probability given by (3) can immediately be evaluated for considerable parts of the entire histories. If Markov properties in time τ and position κ are assumed eq. (3) reduces to

$$\begin{aligned} &P(x_{\tau,\kappa}^{(i)} | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau) = \\ &\sum_{l,k,r=1}^{R_{\kappa-1}, R_\kappa, R_{\kappa+1}} P(x_{\tau,\kappa}^{(i)}, x_{\tau-1,\kappa}^{(k)}, x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)} | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau). \end{aligned} \tag{4}$$

An efficient computation of the marginal probability (4) is possible by using a recursive description. Hence, we apply Bayes' Theorem as well as the chain rule to the conditional probability under the sum and get

$$\begin{aligned} &P(x_{\tau,\kappa}^{(i)}, x_{\tau-1,\kappa}^{(k)}, x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)} | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau) \\ &= C \cdot p(z_{\tau,\kappa}, z_{\tau,\kappa-1}, z_{\tau,\kappa+1} | x_{\tau,\kappa}^{(i)}, x_{\tau-1,\kappa}^{(k)}, \\ &\qquad x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\quad \cdot P(x_{\tau,\kappa}^{(i)}, x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)} | x_{\tau-1,\kappa}^{(k)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\quad \cdot P(x_{\tau-1,\kappa}^{(k)} | z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \end{aligned} \tag{5}$$

with the normalization constant

$$C = \frac{p(z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1})}{p(z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau)} \quad . \tag{6}$$

The recursive part is given by the last term in (5) which is the same as the conditional probability on the left hand side of (4) if τ is replaced by $\tau - 1$. If we consider that the channel is memoryless then the power density function (pdf) in the second line of (5) can be expressed as product of independent pdfs

$$\begin{aligned} &p(z_{\tau,\kappa}, z_{\tau,\kappa-1}, z_{\tau,\kappa+1} | x_{\tau,\kappa}^{(i)}, x_{\tau-1,\kappa}^{(k)}, \\ &\qquad x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &= p(z_{\tau,\kappa} | x_{\tau,\kappa}^{(i)}) \cdot p(z_{\tau,\kappa-1} | x_{\tau,\kappa-1}^{(l)}) \cdot p(z_{\tau,\kappa+1} | x_{\tau,\kappa+1}^{(r)}). \end{aligned} \tag{7}$$

Each term in (7) comprises information about the channel quality in adjacent positions $\kappa - 1$, κ and $\kappa + 1$ at time instance τ.

If we apply the chain rule to the second term in (5) and if we assume independence of $x_{\tau-1,\kappa}^{(k)}, x_{\tau,\kappa-1}^{(l)}$ and $x_{\tau,\kappa+1}^{(r)}$ as well as a memoryless channel then we get

$$\begin{aligned} &P(x_{\tau,\kappa}^{(i)}, x_{\tau,\kappa-1}^{(l)}, x_{\tau,\kappa+1}^{(r)} | x_{\tau-1,\kappa}^{(k)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &= P(x_{\tau,\kappa-1}^{(l)} | x_{\tau,\kappa}^{(i)}, x_{\tau,\kappa+1}^{(r)}, x_{\tau-1,\kappa}^{(k)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\quad \cdot P(x_{\tau,\kappa+1}^{(r)} | x_{\tau,\kappa}^{(i)}, x_{\tau-1,\kappa}^{(k)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\quad \cdot P(x_{\tau,\kappa}^{(i)} | x_{\tau-1,\kappa}^{(k)}, z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\approx P(x_{\tau,\kappa-1}^{(l)} | x_{\tau,\kappa}^{(i)}) \cdot P(x_{\tau,\kappa+1}^{(r)} | x_{\tau,\kappa}^{(i)}) \cdot P(x_{\tau,\kappa}^{(i)} | x_{\tau-1,\kappa}^{(k)}) \end{aligned} \tag{8}$$

Eq. (8) exclusively consists of *a-priori* informations representing the statistical properties of the source. Substituting eqs. (5)-(8), into (4) and re-ordering the sums finally results in

$$\begin{aligned} &P(x_{\tau,\kappa}^{(i)} | z_{1,\kappa}^\tau, z_{1,\kappa-1}^\tau, z_{1,\kappa+1}^\tau) = C \cdot p(z_{\tau,\kappa} | x_{\tau,\kappa}^{(i)}) \\ &\cdot \sum_{k=1}^{R_\kappa} P(x_{\tau,\kappa}^{(i)} | x_{\tau-1,\kappa}^{(k)}) \cdot P(x_{\tau-1,\kappa}^{(k)} | z_{1,\kappa}^{\tau-1}, z_{1,\kappa-1}^{\tau-1}, z_{1,\kappa+1}^{\tau-1}) \\ &\cdot \sum_{l=1}^{R_{\kappa-1}} p(z_{\tau,\kappa-1} | x_{\tau,\kappa-1}^{(l)}) \cdot P(x_{\tau,\kappa-1}^{(l)} | x_{\tau,\kappa}^{(i)}) \\ &\cdot \sum_{r=1}^{R_{\kappa+1}} p(z_{\tau,\kappa+1} | x_{\tau,\kappa+1}^{(r)}) \cdot P(x_{\tau,\kappa+1}^{(r)} | x_{\tau,\kappa}^{(i)}). \end{aligned} \tag{9}$$

Eq. (9) is an extension of the approach given in [2, 3]. The first two lines exploit the inter-frame correlation as well as the non-uniform distribution. Furthermore there are two additional sums which utilize the intra-frame redundancy. Each term combines source dependent and channel dependent informations for one adjacent parameter $x_{\tau,\kappa-1}^{(l)}$ and $x_{\tau,\kappa+1}^{(r)}$ respectively to enhance the estimation process.

4. SIMULATION RESULTS

For simulations we used the source model proposed in [4]. This model allows to adjust correlation properties in time τ by the *auto correlation factor* ρ and in position κ by the *cross correlation factor* δ. Furthermore, in our experiments we set $N = 10$. The parameters are quantized by an 16-level Lloyd-Max Quantizer using 4 bits, i.e. $R = R_\kappa = 16\,\forall\,\kappa$. The index assignment is optimized with respect to the MMSE. As transmission channel serves an AWGN channel with known E_s/N_0.

To demonstrate the performance of the proposed algorithm it is compared with known estimation techniques under noisy channel conditions:

- **MS** *Mean Square* estimator exploits *a-priori* information due to non-uniform distribution of the parameters.
- **SBSD** *Softbit Source Decoding* according to [2, 3] exploits *a-priori* information due to non-uniform distribution and inter-frame correlation.

OPT Optimal estimator according to [4] exploits *a-priori* information due to non-uniform distribution, intra-frame and inter-frame correlation. OPT can be interpreted as a combination of SBSD [2, 3] and Bahl's algorithm [6].

N-OPT Near optimum estimation approach according to (9) exploits *a-priori* information due to non-uniform distribution, intra-frame and inter-frame correlation. N-OPT is an extension of SBSD [2, 3].

Fig. 2 depicts the parameter SNR as a function of E_s/N_0 for a simulation with $\rho = 0$ and $\delta = 0.8$, i.e. there is intra-frame correlation but no inter-frame correlation.

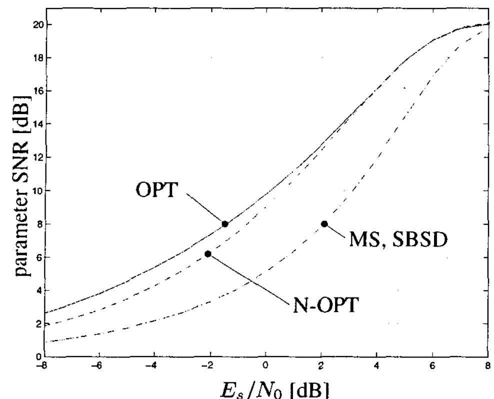

Figure 2: Source with intra-frame correlation $\delta = 0.8$ but without inter-frame correlation $\rho = 0$

The curves of MS and SBSD coincide, because there is no inter-frame redundancy which can be utilized by SBSD. N-OPT as well as OPT benefit from the given intra-frame correlation and therefore outperform MS and SBSD by up to 5 dB parameter SNR. N-OPT comes close to the performance bounds of the optimal MMSE estimator in a wide range of noisy channel conditions. While the channel quality decreases the loss compared to OPT slightly increases but N-OPT is still much better than SBSD.

In a second experiment we assumed inter-frame as well as intra-frame correlation with $\rho = 0.74$ and $\delta = 0.7$.

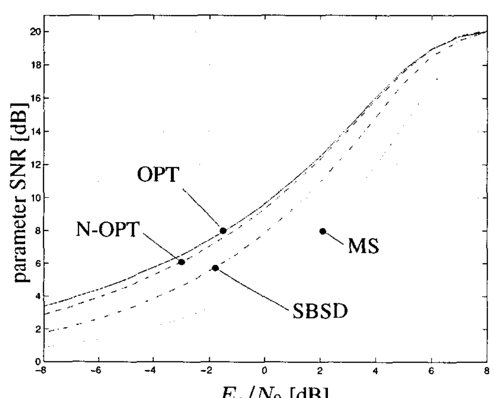

Figure 3: Source with inter-frame correlation $\rho = 0.74$ and intra-frame correlation $\delta = 0.7$

Fig. 3 shows that there is still a significant gain of about 1.51 dB parameter SNR between SBSD and N-OPT. The loss compared to OPT is negligible small.

Furthermore, we compare all methods with respect to the computational complexity which is required to determine a set of $P(x_{\tau,\kappa}^{(i)}|\mathbf{z}_1^{\tau})$ with $\kappa = 1, 2, \ldots, N$.

MS	SBSD	OPT	N-OPT
$N \cdot R$	$\approx N \cdot R^2$	$\approx 6(N-1) \cdot R^3$	$\approx 3N \cdot R^2$

Table 1: Estimates for the computational complexity

Complexity will be counted in terms of arithmetic operations such as multiplications, additions as well as multiply-accumulates. For the complexity estimates listed in Tab. 1 we just consider terms which differ between the methods, e.g. the normalization by C is neglected. Indeed due to the additional sums in (9) N-OPT demands three times the operations compared to SBSD but compared to OPT the power of R can be decreased by one.

5. APPLICATIONS

The proposed parameter estimator has been applied to the differentially encoded LSF coefficients of the GSM-AMR codec at the rate of 10.2 kbit/s and to the scale factors of the DAB transform codec. In case of the GSM-AMR codec inter-frame correlation of about $\rho = 0.74$ and intra-frame correlation of about $\delta = 0.7$ was measured for a considerable selection of speech samples. The settings of the second experiment from Section 4 were aligned to these values. For the DAB codec correlation of about $\rho = 0.95$ and $\delta = 0.8$ was determined [4]. The relative gain of parameter SNR of SBSD compared to OPT, N-OPT respectively, is similar to the results presented for the GSM-AMR codec.

6. CONCLUSION

We developed a new parameter estimator for source coding schemes which generate parameters with inter-frame as well as intra-frame redundancy. The proposed parameter estimator is able to utilize both types of redundancy and therefore outperforms techniques which neglect either one of them. The new approach is suboptimal in the MMSE sense but it is able to reach the performance bounds of the optimal estimator in a wide range of noisy channel conditions with much less complexity.

REFERENCES

[1] C. E. Shannon, *Collected Papers*. IEEE Press, N.J.A. Sloane and A. Wyner, New York, 1993.

[2] T. Fingscheidt and P. Vary, "Error Concealment by Softbit Speech Decoding," in *Proc. of ITG-Fachtagung "Sprachkommunikation"*, (Frankfurt a. M., Germany), pp. 7–10, Sep 1996.

[3] T. Fingscheidt and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *Accepted for publication in IEEE Trans. on Speech and Audio Processing*, 2000.

[4] S. Heinen, P. Vary, and J. Spittka, "An Optimal MMSE-Estimator For Source Codec Parameters Using Intra-Frame and Inter-Frame Correlation," *Submitted to GLOBECOM*, (San Francisco, USA), 2000.

[5] J. Hagenauer and P. Höher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications," in *Proc. of GLOBECOM*, pp. 1680–1686, Nov. 1989.

[6] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. on Information Theory*, pp. 284–287, Mar. 1974.

[7] S. Heinen and P. Vary, "Source Optimized Channel Codes (SOCCS) for Parameter Protection," in *Proc. of Int. Symp. on Information Theory, ISIT*, (Sorrento, Italy), IEEE, June 2000.