ERROR CONCEALMENT BY SOFTBIT SPEECH DECODING

Tim Fingscheidt, Peter Vary

Institute of Communication Systems and Data Processing (IND)
Aachen University of Technology, 52056 Aachen, Germany
e-mail: tim@ind.rwth-aachen.de

ABSTRACT

In digital mobile communication systems there is a need for reducing the subjective effects of residual bit errors by error concealment techniques. Due to the fact that most standards do not specify these algorithms bit exactly, there is room for new solutions to improve the decoding process.

This contribution develops a new approach for optimum estimation of speech codec parameters using softbits. It can be applied to any speech codec standard if channel state information is provided by the demodulator (e.g. DECT), or by the channel decoder (e.g. soft-output Viterbi algorithm - SOVA [6] in GSM). The proposed method includes an inherent muting mechanism leading to a graceful degradation of speech quality in case of adverse transmission conditions. Particularly the additional exploitation of residual source redundancy, i.e. some a priori knowledge about codec parameters gives a significant enhancement of the output speech quality. In the case of an error free channel, bit exactness as required by the standards can be preserved.

1. INTRODUCTION

There are some earlier publications that deal with error concealment using channel state information as well as a priori knowledge: The GSM recommendations [1, 2] e.g. describe a simple solution based on frame repetition. In [3] a Viterbi like decoder is used to find the codec parameter that provides the maximum a posteriori probability. Gerlach proposed a generalized extrapolation technique that is able to use parameter-individual estimators [4], but he assumed that previously received parameters are known exactly, i.e. without error. Recently, Hagenauer [7] introduced a channel decoding mechanism using a priori knowledge about bits to achieve a significantly reduced residual bit error rate before speech decoding.

In general terms the quality of the decoded speech under poor channel conditions depends on the proper estimation of codec parameters. For this reason, we focus on the estimation of codec parameters rather than on the detection of individual bits. Furthermore, the proposed error concealment technique is able to include parameter individual estimators without taking any assumption about previously received parameters. The Bayesian methods are applied to perform an optimum estimation of codec parameters.

Let us consider a specific codec parameter $\tilde{V} \in \mathbf{R}$ which is coded by M bits. In Fig. 1 the coding and transmission process via a noisy channel as well as the proposed decoding process by estimation are depicted. The quantized parameter $\mathbf{Q}[\tilde{V}] = V$ with $V \in$ QT (QT: quantization table) is represented by the bit combination $\underline{x} = (x(0), x(1), ..., x(m), ..., x(M-1))$ consisting of M bits. The bits are assumed to be bipolar, i.e. $x(m) \in \{-1, +1\}$. Any bit combination $\underline{x}$ is assigned to a quantization table index i, such that we can write $\underline{x} = \underline{x}^{(i)}$ as well as $V = V^{(i)}$ with index $i \in \{0, 1, ..., 2^M - 1\}$ to denote the quantized parameter. Furthermore, we distinguish the parameter value at the receiver from the value at the transmitter by a hat on the (possibly modified) received values. In a conventional decoding scheme the received bit combination $\underline{\hat{x}}$ is input to an "inverse bit mapping" or "inverse quantization" scheme, i.e. the appropriate parameter $\hat{V}$ is addressed in a quantization table.

The proposed error concealment technique needs additional information about the channel: the *channel state information* (CSI). It is required to compute a set of probabilities $\mathrm{P}(\underline{\hat{x}} \mid \underline{x}^{(i)})$, $i = 0, 1, ..., 2^M - 1$, of a transition from any bit combination $\underline{x}^{(i)}$ at the transmitter to the received bit combination $\underline{\hat{x}}$. The computation of the transition probabilities depends on the chosen channel model and is therefore discussed in section 2.

The joint information about the detected hardbit $\hat{x}(m)$ and the appropriate CSI is what we call a *softbit*. There are several alternative and equivalent representations of the softbit as shown in section 2. Thus the proposed error concealment technique is called a *softbit speech decoding*.

The next step is to exploit the transition probabilities as well as some *a priori knowledge* about the regarded parameter. Both types of information are combined in a set of *a posteriori probabilities* $\mathrm{P}(\underline{x}^{(i)} \mid \underline{\hat{x}}, ...)$, with $i = 0, 1, ..., 2^M - 1$, denoting the probability that $\underline{x}^{(i)}$ had been transmitted in the case that $\underline{\hat{x}}$ has been received (sec. 3).

The parameter estimator is the last block in the error concealment process. It uses the a posteriori probabilities to find the optimum parameter $\hat{V}_{est}$ referring to a given criterion. Two widely used estimators are discussed in this context in section 4.

Finally, in section 5, the application to PCM coded speech is presented to prove the capabilities of the proposed softbit speech decoding technique.

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) within the program "Mobilkommunikation".

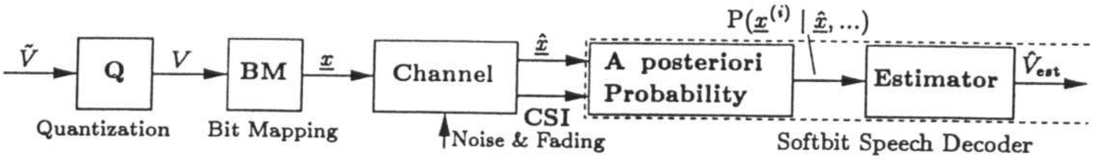

Figure 1: New Softbit Speech Decoding Technique

2. THE CHANNEL STATE INFORMATION

2.1 The Binary Symmetric Channel (BSC)

In Fig. 2a a BSC is depicted. The transition probability from a transmitted bit $x^{(i)}(m)$ to a received bit $\hat{x}(m)$ is then

$$\mathrm{P}(\hat{x}(m) \mid x^{(i)}(m)) = \begin{cases} 1 - p_e(m) & \text{if } \hat{x}(m) = x^{(i)}(m) \\ p_e(m) & \text{if } \hat{x}(m) \neq x^{(i)}(m) \end{cases} \quad (1)$$

where $p_e(m)$ denotes the instantaneous bit error rate. If the channel is assumed to be memoryless, the transition probability of a bit combination reads

$$\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)}) = \prod_{m=0}^{M-1} \mathrm{P}(\hat{x}(m) \mid x^{(i)}(m)) \quad . \qquad (2)$$

In the following, this term is called the *channel dependent information* referring to parameter index i. Assuming a memoryless channel any other symmetric channel model (like e.g. the fading channel in Fig. 2b) can be reduced to an estimate of $p_e(m)$ and thus (1) and (2) can be used.

2.2 The Fading Channel

The CSI of a fading channel as depicted in Fig. 2b consists of an estimate of the fading factor a. Additionally, we assume the noise variance $N_0/2E_b$ to be known at the receiver. In the case of an AWGN channel, a is set to one.

An instantaneous bit error rate for the received hardbit $\hat{x}(m)$ can be formulated in terms of log-likelihood values [5]

$$p_e(m) = \frac{1}{1 + \exp|L_c \cdot \tilde{x}(m)|} \quad \text{with} \quad L_c = 4a \cdot \frac{E_b}{N_0} \quad . \qquad (3)$$

From (3) it can be seen that to any received value $\tilde{x}(m)$ an individual bit error rate is assigned, even if the reliability value L_c of the channel remains constant. For this reason, we call the p_e-term in (3) an *instantaneous bit error rate.*

The mean value of the instantaneous bit error rate as defined in (3) equals the well known BPSK bit error rate. It can be used to compute estimates of the $\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)})$ terms.

2.3 The Soft-Output Viterbi Algorithm

If the soft-output Viterbi algorithm (SOVA) together with an interleaving scheme is used as proposed in [6], then the instantaneous bit error rate is computed by

$$p_e(m) = \frac{1}{1 + \exp|L(m)|} \qquad (4)$$
$$\text{with} \quad L(m) = \ln \frac{\mathrm{P}(x^{(i)}(m) = +1 \mid \tilde{\underline{X}})}{\mathrm{P}(x^{(i)}(m) = -1 \mid \tilde{\underline{X}})}$$

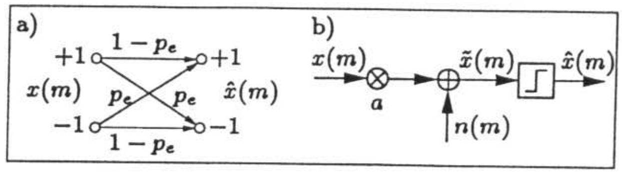

Figure 2: a) BSC channel, b) Fading channel with fading factor a and additive white Gaussian noise $n(m)$

being the soft-output value whose sign $\hat{x}(m) = \mathrm{sign}(L)$ equals the decoded hardbit, $x^{(i)}(m)$ denoting the corresponding transmitted bit, and $\tilde{\underline{X}}$ being the received sequence of symbols that is input to the channel decoder. Because of the integrated interleaving scheme, this bit error rate can be used in the same way as p_e in (3) to get the required channel dependent information.

3. THE PROBABILITY OF A RECEIVED PARAMETER

3.1 Simple Approximations

For the estimation of speech codec parameters at the receiver, *a posteriori probability* terms providing information about any transmitted parameter index i are required. It can be shown

$$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}) = C \cdot \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)}) \cdot \mathrm{P}(\underline{x}^{(i)}) \qquad (5)$$
$$\text{with} \quad C = \frac{1}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(l)}) \cdot \mathrm{P}(\underline{x}^{(l)})} \quad .$$

Varying the a posteriori term over i, we get the probability of any transmitted channel index if $\hat{\underline{x}}$ had been received. The term $\mathrm{P}(\underline{x}^{(i)})$ provides a *source dependent information* and is called the *0th order a priori knowledge* about the source, because it is provided by a simple histogram of $V^{(i)}$ regardless any residual correlations between successive parameters.

If there is no knowledge available about the source statistics, one can only exploit the channel dependent information assuming that the parameters $V^{(i)}$ being equally likely. In this case (5) is simplified to

$$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}) = \frac{\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)})}{\sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(l)})} \quad . \qquad (7)$$

In practice, this simplification does not hold very well because e.g. optimum Lloyd-Max quantizers yield identical quantization error variance contributions of any quantization interval i rather than identical probabilities $\mathrm{P}(\underline{x}^{(i)})$.

We can summarize that equation (7) is based on a coarse approximation to compute the a posteriori probabilities of codec parameters. A significantly better solution is given by the exact formula (5).

$$P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}) = C \cdot P(\hat{\underline{x}}_0 \mid \underline{x}_0{}^{(i)}) \cdot \sum_{j=0}^{2^M-1} P(\underline{x}_0{}^{(i)} \mid \underline{x}_{-1}{}^{(j)}) \cdot P(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}) \tag{6}$$

$$\text{with} \quad C = \frac{1}{\sum_{l=0}^{2^M-1} P(\hat{\underline{x}}_0 \mid \underline{x}_0{}^{(l)}) \cdot \sum_{j=0}^{2^M-1} P(\underline{x}_0{}^{(l)} \mid \underline{x}_{-1}{}^{(j)}) \cdot P(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})} \; .$$

3.2 Approximations Exploiting Correlations

The classical approaches of speech coding aim at minimizing the residual redundancy of codec parameters. However, due to the coding strategy or at least to the limited amount of processor resources and the maximum of the allowed signal delay, in most applications residual correlations between successive speech codec parameters can be observed. As already mentioned by Shannon [8] this source coding suboptimality can be exploited at the receiver side in the parameter estimation process. In contrast to [7], we will exploit the residual redundancy not per bit but per parameter, thus being independent on the specific mapping of bits $\underline{x}^{(i)}$ to quantization table entries $V^{(i)}$ in the encoder. Our a posteriori term can easily be extended to regard as well parameter correlations: The maximum information that is available at the decoder consists in the complete path of already received bit combinations

$$P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{x}}_{-2}, ..., \hat{\underline{x}}_{-n}, ..., \hat{\underline{x}}_{-N}) = P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}) \tag{8}$$

with $\hat{\underline{x}}_{-n}$ denoting the bit combination n time instants[1] before the present one and N being the number of already received bit combinations in the past. The term $\hat{\underline{X}}_{-2}$ includes the complete history up to the received bit combination two time instants before the actually regarded one.

To compute the a posteriori term (8) it is necessary to find a statistical model of the sequence of quantized parameters $V_{-n}^{(i)}$. It seems reasonable to discuss the sequence of quantized parameters modelled as a Markov process of 1st order. Solutions for higher order models can be derived in a similar manner.

Modelling of a parameter as a 1st order Markov process means $P(\underline{x}_0 \mid \underline{x}_{-1}, \underline{x}_{-2}, ..., \underline{x}_{-N}) = P(\underline{x}_0 \mid \underline{x}_{-1})$. After some intermediate steps the solution can be given in terms of a recursion according to (6). To emphasize that correlations between adjacent parameters are regarded, we call $P(\underline{x}_0{}^{(i)} \mid \underline{x}_{-1}{}^{(j)})$ a *1st order* a priori knowledge. In eq. (6) the term $P(\underline{x}_{-1}{}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})$ is nothing else but the resulting a posteriori probability $P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})$ from the previous time instant. Thus a recursion could be found computing the a posteriori probabilities of all 2^M possibly transmitted bit combinations at any time instant exploiting the maximum knowledge that is available at the decoder.

[1] The term "time instant" denotes any moment when the regarded parameter is received. In the ADPCM codec e.g. it equals a sample instant, in CELP coders it may be a frame or a subframe instant.

4. INDIVIDUAL PARAMETER ESTIMATION USING THE A POSTERIORI PROBABILITIES

For a wide area of speech codec parameters the minimum mean square error criterion (MS) is appropriate. These parameters may be PCM speech samples, spectral coefficients, gain factors, etc. In contrast to that the estimation of a pitch period from an unreliable received bit combination must be performed according to a different error criterion. The simplest is the MAP (maximum a posteriori) estimator. In the following we discuss these two well known estimators in the context of speech codec parameter estimation.

4.1 The MAP Estimation

The MAP estimator is the one requiring the least additional computational complexity. It follows the criterion

$$V_{MAP} = V^{(\nu)} \quad \text{with} \tag{9}$$
$$P(\underline{x}_0{}^{(\nu)} \mid \hat{\underline{x}}_0, ...) = \max_i P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...), \quad i = 0, 1, ..., 2^M-1$$

while $P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...)$ denotes any of the a posteriori probabilities given in (7), (5), or (6) dependent on the chosen model order and the availability of a priori knowledge. The optimum decoded parameter in a MAP sense V_{MAP} always equals one of the codebook/ quantization table entries. Regardless the type of the speech codec parameter, a MAP estimation always minimizes the decoding error probability [9]. Nevertheless, a wide area of parameters can be reconstructed much better using the mean square estimator.

4.2 The Mean Square Estimation

The optimum decoded parameter V_{MS} in a mean square sense equals

$$V_{MS} = \sum_{i=0}^{2^M-1} V^{(i)} \cdot P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...) \; . \tag{10}$$

According to the well known orthogonality principle of the linear mean square (MS) estimation (see e.g. [9]) the variance of the estimation error $e_{MS} = V_{MS} - V$ is simply

$$\sigma^2_{e_{MS}} = \sigma^2_V - \sigma^2_{V_{MS}} \tag{11}$$

with σ^2_V being the variance of the undisturbed parameter V and $\sigma^2_{V_{MS}}$ denoting the variance of the mean square estimated parameter V_{MS}. Because $\sigma^2_{e_{MS}} \geq 0$ we can state that the variance of the estimated parameter is smaller than or equals the variance of the error free parameter. In the case of a worst case channel with $p_e = 0.5$ the a posteriori probability degrades to $P(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, ...) = P(\underline{x}_0{}^{(i)})$. As a

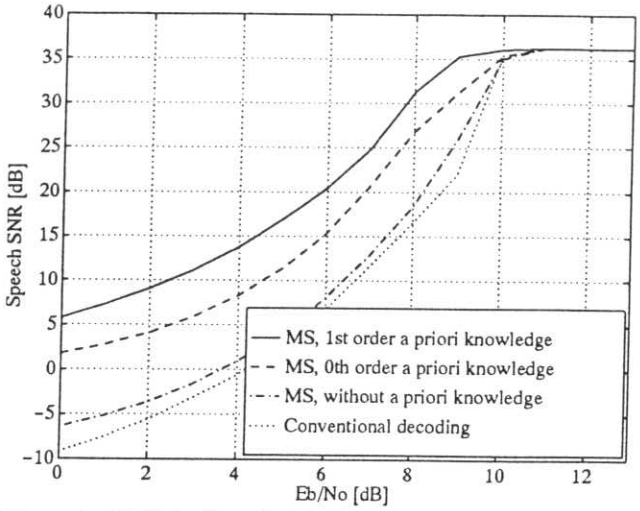

Figure 3: Softbit decoding: A-law PCM over an AWGN channel using coherently detected BPSK (Fig. 2b, $a = 1$): High bit error rate resolution according to eq. (3).

consequence, the MS estimated parameter according to eq. (10) is completely attenuated to zero if the quantization table entries $V^{(i)}$ as well as $P(\underline{x}_0{}^{(i)})$ are distributed symmetrically about zero. These symmetries are often found for gain factors in CELP coders. Thus the MS estimation of the gain factors results in an inherent muting mechanism providing a graceful degradation of speech. This is one of the main results of our proposed softbit speech decoding technique.

5. AN APPLICATION EXAMPLE: PCM

In principle, the proposed algorithms are able to be applied to any speech codec. We simulated a simple PCM transmission over an AWGN channel assuming a coherent BPSK demodulation. For this case the channel model in Fig. 2b fits when setting $a = 1$. Fig. 3 shows four different simulation results in terms of speech SNR as a function of the E_b/N_0 ratio. At first the SNR of the conventionally decoded speech is drawn having applied simply the hard decision mechanism as depicted in Fig. 2b. In addition three different cases of a mean square estimation are shown: A simple version that does not require any a priori knowledge (i.e. the a posteriori probability from eq. (7)), secondly an estimation using 0th order a priori knowledge (eq. (5)), and finally an estimation using 1st order a priori knowledge (eq. (6)).

In any case, the MS estimated speech degrades asymptotically to 0 dB with decreasing E_b/N_0, i.e. the inherent muting mechanism of MS estimation. The shape of the curves strongly depends on the order of exploited a priori knowledge. While a MS estimation without a priori knowledge just leads to a small gain of about 1 ... 2 dB (speech SNR), the exploitation of a priori knowledge allows gains of up to 10 dB (0th order), and up to 15 dB (1st order), respectively. This leads to a significant enhancement of speech quality although the model of the speech as Markov process of 1st order is still a very simple one and surely can be refined further.

6. SUMMARY

In this paper we proposed a new error concealment technique that is able to exploit different amounts of a priori knowledge about the source. It uses channel state information to compute transition probabilities from one bit combination to another bit combination each representing a speech codec parameter. For two well known channel models as well as the soft-output Viterbi algorithm (SOVA) channel decoder we gave expressions to compute these probabilities.

We derived the optimum a posteriori probability of a bit combination as well as different approximations to be used in parameter individual estimators. Two common estimators were discussed showing that the mean square estimator is able to perform a graceful degradation of speech in case of decreasing quality of the transmission link because of its inherent muting mechanism.

We applied the mean square estimator to PCM coded speech over an AWGN channel gaining up to 15 dB in the speech SNR. The subjective speech quality could be enhanced drastically.

This approach can be applied to different source coding schemes such as ADPCM and CELP.

ACKNOWLEDGEMENTS

The authors would like to thank D. Herzberg as well as W. Papen and C. Gerlach for a lot of inspiring discussions.

7. REFERENCES

[1] "Recommendation GSM 06.11, Substitution and Muting of Lost Frames for Full Rate Speech Traffic Channels", *ETSI/TC SMG*, February 1992.

[2] "Recommendation GSM 06.21, Substitution and Muting of Lost Frames for Half Rate Speech Traffic Channels", *ETSI/TC SMG*, January 1995.

[3] K. Sayood and J.C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/ Channel Coders", *IEEE Transactions on Communications*, vol. 39, no. 6, pp. 838–846, June 1991.

[4] C.G. Gerlach, "A Probabilistic Framework for Optimum Speech Extrapolation in Digital Mobile Radio", *Proc. of ICASSP'93*, pp. II-419–II-422, April 1993.

[5] J. Hagenauer, "Viterbi Decoding of Convolutional Codes for Fading- and Burst-Channels", *Proc. of the 1980 Zurich Seminar on Digital Communications*, pp. G2.1-G2.7, 1980.

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

[7] J. Hagenauer, "Source- Controlled Channel Decoding", *IEEE Transactions on Communications*, vol. 43, no. 9, pp. 2449–2457, September 1995.

[8] C.E. Shannon, "A Mathematical Theory of Communication", *Bell Systems Technical Journal*, vol. 27, pp. 379–423, July 1948.

[9] J.L. Melsa and D.L. Cohn, "Decision and Estimation Theory", *McGraw-Hill Kogakusha*, 1978.