ROBUST SPEECH DECODING: CAN ERROR CONCEALMENT BE BETTER THAN ERROR CORRECTION?

Tim Fingscheidt, Peter Vary, Jesús A. Andonegui

Aachen University of Technology
Institute of Communication Systems and Data Processing, D-52072 Aachen, Germany
Tim.Fingscheidt@ind.rwth-aachen.de   http://www.ind.rwth-aachen.de/~tim

ABSTRACT

Digital speech transmission systems use source coding to reduce the bit rate and channel coding to correct transmission errors. Furthermore, in periods of a very poor channel quality error concealment of residual bit errors becomes necessary as channel decoding fails. However, if the channel is clear, channel coding would not be required at all and the speech quality could be improved by allowing a higher bit rate for source encoding. Usually a compromise is taken between speech quality in case of clear channel and error robustness in case of poor channel quality. This paper addresses the problem of a joint optimization of error concealment and source/channel coding. Under the premise of a minimum mean square error criterion for signal reconstruction it turns out that error concealment instead of error correction may be the best choice if source coding leaves sufficient residual parameter correlations by less bit rate reduction.

1. INTRODUCTION

Error correction by means of channel coding provides powerful mechanisms to guarantee a low bit error rate over a wide range of channel qualities in digital speech communication systems. Nevertheless, under very poor transmission conditions the residual bit error rate after channel decoding becomes higher than that of an uncoded transmission scheme [10]. Usually, an error concealment technique is used after channel decoding to provide an acceptable quality of the reconstructed speech signal. On the other hand, at very good channel qualities channel coding could even be omitted to allow a higher bit rate of the source coding scheme providing an enhanced speech quality. These drawbacks of fixed channel coding schemes are motivating (besides other reasons) the current standardization effort of an *adaptive multi rate* (AMR) codec for GSM. AMR aims at an adaptive bit rate assignment to source and channel coding within the gross bit rate of the half-rate (TCH/HS) and full-rate (TCH/FS) speech channel [7].

In this paper we discuss an alternative approach without any channel coding that provides – if certain constraints are met – a higher quality both at good and very poor channel qualities without the need of switching between different

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

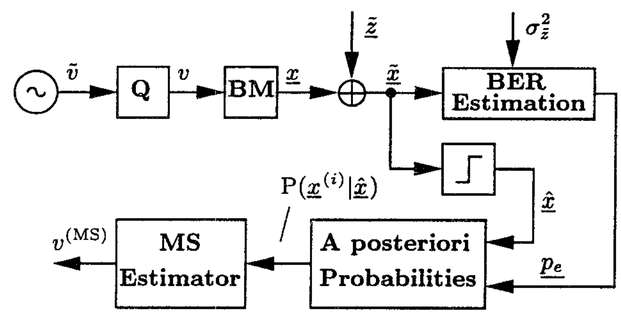

Figure 1: Transmission system with robust decoding (SD)

coding techniques. The largest challenge of this conception consists of reaching a performance comparable to a channel coded scheme at medium channel qualities. The system that we develop in the following focuses on error concealment based on the robust speech decoding technique described in [2–6] and is subsequently abbreviated by SD, i.e. *soft decoding*. It is to be compared to a system with conventional source coding and error correction by channel coding (abbreviated by CHCOD).

To carry out a fair comparison common system constraints and requirements must be defined. The transmission of a single speech codec parameter which is modelled as an uncorrelated Gaussian process is regarded. The source coding shall be carried out by using a scalar quantizer. As for many speech codec parameters the MMSE is the best quantization strategy, we define the optimum reconstruction quality to be the signal-to-noise ratio (SNR), although other criteria could be used (e.g. [3]). The available gross bit rate is $R_g = 4$ bit/parameter and the transmission scheme is BPSK over an AWGN channel. The required transmission "quality" minimum (judged as "good") is assumed to be $\text{SNR}_{\min} = 9$ dB.

2. THE ROBUST DECODING SYSTEM

Let's regard the transmission system with robust decoding (SD) as given in Fig. 1. Due to the delay-free source coding, each Gaussian parameter $\tilde{v}$ is quantized by an M bit scalar Lloyd-Max quantizer (LMQ) to a value v [9]. The LMQ yields the maximum SNR in error-free channel situations. After applying a bit mapping scheme the bipolar bits

LMQ:	Lloyd-Max quantizer
MEQ:	*Minimum entropy* quantizer
SD,AKN:	Robust decoding, Nth order a priori knowledge
HD:	Hard decision by table lookup
CHCOD/S or /H:	Soft (or hard) input channel decoding

Table 1: Abbreviations

$x(m) \in \{-1, +1\}$, $m = 0, ..., M-1$, of the corresponding bit combination $\underline{x} = (x(0), ..., x(M-1))$ are BPSK modulated and transmitted over an AWGN channel with noise variance $\sigma_{\tilde{z}}^2 = N_0/2E_b$. In addition to a hard decision $\hat{x}(m)$ of the received values $\tilde{x}(m)$ the decoder provides instantaneous bit error probabilities $\underline{p_e} = (p_e(0), ..., p_e(m), ..., p_e(M-1))$ based on an ideal estimation of the noise variance $\sigma_{\tilde{z}}^2$. According to e.g. [2, 4] these bit error probabilities can be used to compute 2^M transition probabilities $\mathrm{P}(\hat{\underline{x}}|\underline{x}^{(i)})$ from every (possibly) transmitted bit combination $\underline{x}^{(i)}$ with the quantization table index $i = 0, 1, ..., 2^M - 1$, to the received one $\hat{\underline{x}}$. If a histogram measurement $\mathrm{P}(\underline{x}^{(i)})$ of the quantized parameter $v^{(i)}$ – in the following called a priori knowledge of *0th* order (AK0) – is available in the decoding process, then 2^M a posteriori probabilities

$$\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}}) = C \cdot \mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)}) \cdot \mathrm{P}(\underline{x}^{(i)}) \qquad (1)$$

referring to every valid parameter value $v^{(i)}$ of the transmitter can be computed. The a priori knowledge can be measured in advance, using representative speech material. Here and in the following the constant C normalizes the sum of the a posteriori probabilities to one.

Due to the SNR quality measure the decoder performs an MS estimation of the parameter according to

$$v^{(\mathrm{MS})} = \sum_{i=0}^{2^M-1} v^{(i)} \cdot \mathrm{P}(\underline{x}^{(i)}|\hat{\underline{x}}) \,. \qquad (2)$$

The superiority of the SD method in comparison to simple *hard decision* with a lookup table (HD) decoding can be seen in Fig. 2 for $M = R_g = 4$. Depending on the channel quality E_b/N_0 the SNR gain of SD against HD is quite considerable.

However, if the parameter contains residual sample-to-sample correlations that were not removed by the source encoding scheme, a higher decoding gain can be achieved by employing the recursively computed a posteriori probabilities [2, 4]

$$\mathrm{P}(\underline{x}_n{}^{(i)} \mid \hat{\underline{x}}_n, \hat{\underline{X}}_{n-1}) = C \cdot \mathrm{P}(\hat{\underline{x}}_n \mid \underline{x}_n{}^{(i)}) \cdot \sum_{j=0}^{2^M-1} \mathrm{P}(\underline{x}_n{}^{(i)} \mid \underline{x}_{n-1}{}^{(j)}) \cdot \mathrm{P}(\underline{x}_{n-1}{}^{(j)} \mid \hat{\underline{x}}_{n-1}, \hat{\underline{X}}_{n-2}) \qquad (3)$$

instead of $\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}})$ in eq. (2) with n denoting the present sampling instant. The history of received bit combinations $\hat{\underline{x}}_{n-1}, \hat{\underline{x}}_{n-2}, \ldots$ is summarized by $\hat{\underline{X}}_{n-1}$. The *1st order* a priori knowledge (AK1) used in (3) consists of the transition

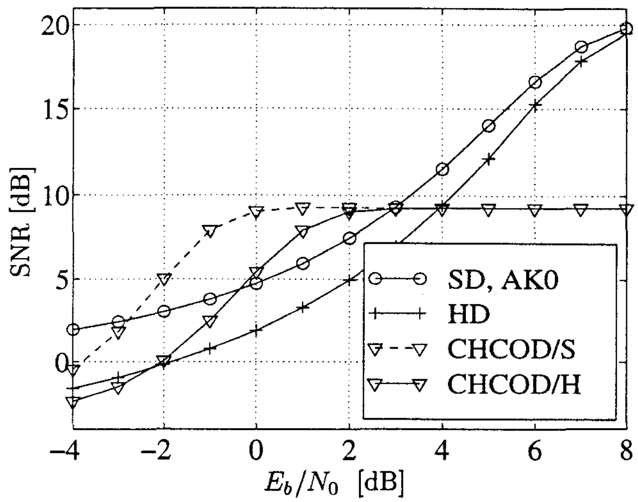

Figure 2: SNR performance: error concealment system (SD) with $M = 4$ bit Lloyd-Max quantizer (LMQ); rate $r = 1/2$ convolutional error correction system (CHCOD) with $M = 2$ bit LMQ

probability matrix $\mathrm{P}(\underline{x}_n{}^{(i)} \mid \underline{x}_{n-1}{}^{(j)})$. Dependent on the parameter correlations, this method can yield remarkable SNR gains compared to SD,AK0.

If a single-sample delay is allowed, interpolation (INT1) instead of extrapolation can be performed by using

$$\mathrm{P}(\underline{x}_n{}^{(i)}|\hat{\underline{x}}_{n+1}, \hat{\underline{x}}_n, \hat{\underline{X}}_{n-1}) = C \cdot \mathrm{P}(\underline{x}_n{}^{(i)}|\hat{\underline{x}}_n, \hat{\underline{X}}_{n-1}) \cdot \sum_{l=0}^{2^M-1} \mathrm{P}(\hat{\underline{x}}_{n+1}|\underline{x}_{n+1}{}^{(l)}) \cdot \mathrm{P}(\underline{x}_{n+1}{}^{(l)}|\underline{x}_n{}^{(i)}) \qquad (4)$$

in the estimation with $\mathrm{P}(\underline{x}_n{}^{(i)}|\hat{\underline{x}}_n, \hat{\underline{X}}_{n-1})$ taken from the recursion (3). This method provides a further SNR gain relative to SD,AK1. It should be noted that in a frame-oriented transmission like GSM the subframe speech codec parameters (except the last of the frame) can be decoded by the SD,AK1,INT1 method without introducing additional delay. It would be possible to switch to SD,AK1 to decode the last subframe parameter.

3. ERROR CONCEALMENT VERSUS ERROR CORRECTION

3.1. Transmission of an Uncorrelated Parameter

Concerning the system constraints and requirements given in section 1, the SD,AK0 curve in Fig. 2 shows the reference performance of an error concealment by robust decoding (SD). In the case of clear channel conditions a high SNR of 20.22 dB due to the LMQ is obtained.

What is the SNR performance of a system with channel coding and hard-input Viterbi decoding (CHCOD/H), and soft-input Viterbi decoding (CHCOD/S), respectively? In this case, the source coding can be performed by an $M = 2$ bit LMQ yielding a basic quality of SNR = 9.3 dB > $\mathrm{SNR}_{\min}$. To exploit the whole available gross bit rate of

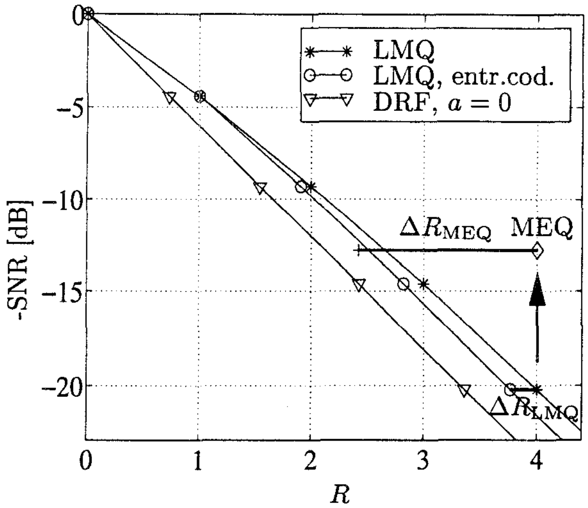

Figure 3: Gaussian distortion rate function (DRF); LMQ and entropy coded LMQ; *minimum entropy quantizer* MEQ at $M = 4$ bit, $\text{SNR}_{\text{MEQ}} = 12.8$ dB;
horizontal lines: redundancy used by the SD,AK0 scheme

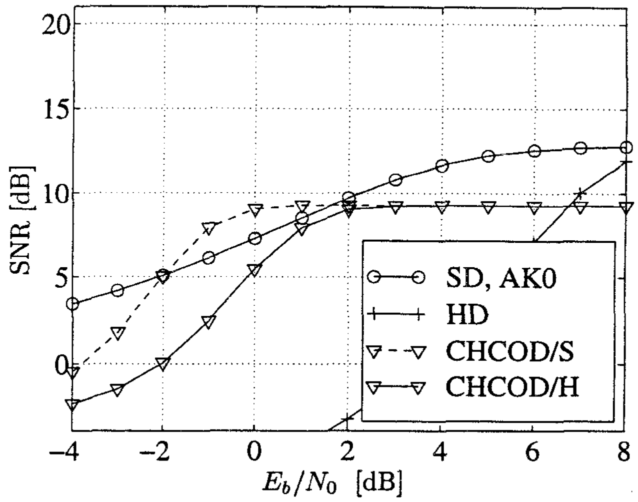

Figure 4: SNR performance: error concealment system (SD) with $M = 4$ bit MEQ; rate $r = 1/2$ error correction system (CHCOD) with $M = 2$ bit LMQ

$R_g = 4$ bit/parameter a rate 1/2 convolutional channel coding scheme with constraint length $L = 5$ is applied. Source decoding is done by table lookup.

To allow a comparison to the SD technique, the results are also depicted in Fig. 2 with E_b denoting the energy of the BPSK symbols. The SD mechanism provides a very good quality at good and a respectable quality at very bad channel qualities, but in between the widely used CHCOD/S scheme remains superior.

3.2. Optimized Source Encoding with Respect to Robust Decoding

Is there a possibility to enhance the performance of the SD,AK0 scheme with regard to channel coding especially in the region around $E_b/N_0 = 0$ dB at the price of a lower clear channel quality? The answer can be given by the rate distortion theory [1]. In Fig. 3 the so-called *distortion-rate function* (DRF) is depicted by triangles indicating the minimum distortion (i.e. the negative SNR) that can be achieved by optimum quantization of a white Gaussian process at a given data rate R. Because of the fixed rate $R = M$ for any quantized symbol and the delay-free coding requirement the LMQ is quite suboptimum as can be seen by the horizontal distance of the stars to the triangles indicating the theoretical limit. Some of this redundancy (not all!) is given by $\Delta R = M - H(\underline{x})$ with

$$H(\underline{x}) = -\sum_{i=0}^{2^M-1} \mathrm{P}(\underline{x}^{(i)}) \cdot \log_2 \mathrm{P}(\underline{x}^{(i)}) \tag{5}$$

being the entropy of the LMQ outputs. For orientation, the performance of the entropy coded LMQ output is depicted in Fig. 3 by circles. Because the SD,AK0 method uses the histogram $\mathrm{P}(\underline{x}^{(i)})$ for enhanced decoding, the amount of used redundancy exactly equals ΔR. This redundancy used by the SD,AK0 scheme is depicted as an horizontal line from the $M = R_g = 4$ bit Lloyd-Max quantizer star in Fig. 3 to the left. The value of $\Delta R_{\text{LMQ}} = 0.235$ bit seems to be much too less to compete with channel coding schemes especially around $E_b/N_0 = 0$ dB.

Starting from the $M = 4$ bit LMQ star a system optimization is possible by moving up in Fig. 3, i.e. designing a *worse* quantizer with a lower SNR but a larger horizontal distance to the rate distortion bound. Because the SD technique performs an estimation with graceful but early degradation the SNR of a *useful worse* quantizer should be chosen a few dB's above the minimum required transmission quality of $\text{SNR}_{\min} = 9.0$ dB that is kept constant over a wide range of E_b/N_0. We found a clear channel quality of SNR = 12.8 dB to be a good compromise (diamond in Fig. 3). Nevertheless, it should be kept in mind that it is not the distance to the rate distortion bound that can be used by the SD,AK0 method, instead it is always $\Delta R = M - H(\underline{x})$. Thus amongst the infinite number of quantization schemes that provide an SNR of 12.8 dB at $M = R_g = 4$ bit we have to find the one with the maximum ΔR, or equivalently the minimum entropy. It is known that (at least for large rates $R = M$) uniform quantization minimizes the output entropy $H(\underline{x})$ for a given distortion [8]. So we designed a 4 bit uniform quantizer providing an SNR of 12.8 dB and got a step size $d = 0.79$ for $\sigma_{\bar{v}}^2 = 1$. Its usable redundancy equals $\Delta R_{\text{MEQ}} = 1.58$ bit (!) and is again depicted in Fig. 3 as an horizontal line to the left. The amount of unusable redundancy for this quantizer is only a 0.29 bit distance to the theoretical limit. In the following we call this quantizer the *minimum entropy quantizer* (MEQ).

Additionally to the new quantizer design we searched for the optimum bit mapping scheme. The bit mapping has influence both on the HD and the SD techniques und hence we applied an iterative (suboptimum) algorithm to find a good bit mapping for SD,AK0.

In Fig. 4 the performance of SD,AK0 applying an MEQ is depicted. It outperforms the CHCOD/H scheme for all values of E_b/N_0. The SNR gain becomes significant in good and in very bad channel quality situations, thus the method in Fig. 1 using an MEQ is not only an error concealment technique but it combines the advantage of an "excellent" quality (i.e. SNR > 9 dB) at clear channel conditions with the robustness during poor channel situations.

Nevertheless, concerning the soft decision Viterbi decoding there is still a region of E_b/N_0 values where the SNR of SD,AK0 is up to 2 dB less than that of CHCOD/S. If the 9 dB quality requirement is a hard limit, CHCOD/S is still the preferable transmission scheme because it fits the requirements down to $E_b/N_0 = 0$ dB. On the other hand there might be a tolerated quality less than 9 dB below $E_b/N_0 = -2$ dB. In this case the SD system instead of channel coding remains an interesting option.

3.3. Transmission of a Parameter with Residual Correlations

We assume now that the Gaussian process modelling a speech codec parameter is not uncorrelated. This means that the source encoder could not remove all parameter correlations. In the following, let these residual correlations be expressed by $\tilde{v}$ being a sample of a Gaussian autoregressive process (AR(1)) with correlation coefficient $\rho_{\tilde{v}\tilde{v}}$. In this case the decoder can easily be extended to perform the AK1 method using eq. (3). If these residual correlations can be expressed by a correlation factor of $\rho_{\tilde{v}\tilde{v}} = 0.73$ or higher, the SD,AK1 technique becomes superior (or at least comparable) even to the soft decision Viterbi decoding at all values of E_b/N_0, as can be seen in Fig. 5. The interpolation technique INT1 applying eq. (4) provides an additional SNR gain of up to 1 dB. Using INT1, the residual correlations after speech coding need only to be about $\rho_{\tilde{v}\tilde{v}} = 0.6$ to outperform CHCOD/S at any channel quality. Even the concatenation of soft-input/soft-output channel decoding and the SD scheme using an $M = 2$ bit LMQ does not seem to be a real alternative: It turns out that its performance in the critical regions at $E_b/N_0 = 0$ dB is quite comparable and at other channel conditions is significantly worse than that of SD,AK0 with an $M = 4$ bit MEQ.

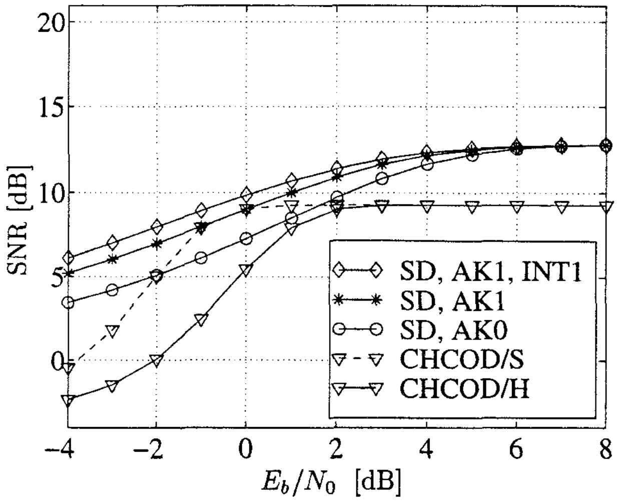

Figure 5: SNR performance for a correlated ($\rho_{\tilde{v}\tilde{v}} = 0.73$) parameter: error concealment system (SD) with $M = 4$ bit MEQ; rate $r = 1/2$ error correction system (CHCOD) with $M = 2$ bit LMQ

4. SUMMARY

In this paper we discussed a transmission scheme based on error concealment as an alternative to error correction. Using SNR-suboptimum and thus sufficiently high-redundant uniform quantizers, at any quality of an AWGN channel the proposed technique becomes superior to a comparable channel coding scheme with hard decision Viterbi decoding. Dependent on the amount of residual parameter correlations of a non-ideal source coding scheme, our approach may even be better than a soft decision Viterbi decoding scheme, especially at good and very poor channel qualities. The aim of this contribution was not to declare channel coding to be superfluous, but to initiate a discussion about the system constraints, where on the one hand combined channel and source coding is the preferable scheme, or where on the other hand intelligent high bit rate source coding with robust i.e. error concealing decoding is superior, respectively.

5. REFERENCES

[1] T. Berger, "Rate Distortion Theory", *Prentice-Hall, Inc.*, 1971.

[2] T. Fingscheidt and P. Vary "Error Concealment by Softbit Speech Decoding", *Proc. of ITG-Fachtagung Sprachkommunikation*, pp. 7–10, September 1996.

[3] T. Fingscheidt and O. Scheufen, "Error Concealment in the GSM System by Softbit Speech Decoding", *Proc. of 9th Aachen Symposium "Signaltheorie"*, pp. 229–232, March 1997.

[4] T. Fingscheidt and P. Vary, "Robust Speech Decoding: A Universal Approach to Bit Error Concealment", *Proc. of ICASSP'97*, vol. 3, pp. 1667–1670, April 1997.

[5] T. Fingscheidt and P. Vary, "Speech Decoding With Error Concealment Using Residual Source Redundancy", *Proc. of IEEE Workshop on Speech Coding*, pp. 91–92, September 1997.

[6] T. Fingscheidt and O. Scheufen, "Robust GSM Speech Decoding Using the Channel Decoder's Soft Output", *Proc. of EUROSPEECH'97*, pp. 1315–1318, September 1997.

[7] K. Hellwig, "Zukünftige Sprachdienste in GSM", *Proc. of 9th Aachen Symposium "Signaltheorie"*, pp. 195–198, March 1997.

[8] N.S. Jayant and P. Noll, "Digital Coding of Waveforms", *Prentice-Hall International*, 1984.

[9] J. Max, "Quantizing for Minimum Distortion", *IRE Transactions on Information Theory*, vol 6, pp. 7–12, March 1960.

[10] J.G. Proakis, "Digital Communications", *McGraw-Hill*, 1989.