ROBUST SPEECH DECODING: A UNIVERSAL APPROACH TO BIT ERROR CONCEALMENT

Tim Fingscheidt, Peter Vary

Aachen University of Technology
Institute of Communication Systems and Data Processing, D-52072 Aachen, Germany
Tim.Fingscheidt@ind.rwth-aachen.de

ABSTRACT

In digital mobile communication systems there is the need for reducing the subjective effects of residual bit errors which have not been eliminated by channel decoding by the use of 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 speech quality.

This contribution develops a new approach for optimum estimation of speech codec parameters. It can be applied to any speech codec standard if a bit reliability information is provided by the demodulator (e.g. DECT), or by the channel decoder (e.g. soft-output Viterbi algorithm - SOVA [7] 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] e.g. describe a simple solution based on frame repetition. In [2] a Viterbi like decoder is used to find the codec parameters that provide the maximum a posteriori probability. Gerlach proposed a generalized extrapolation technique that is able to use parameter-individual estimators [3], but he assumed that previously received parameters are known exactly, i.e. without error. Recently, Hagenauer [4] 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 [5] is able to include parameter individual estimators without taking into consideration idealizing assumptions about previously received parameters. The Bayesian methods or alternatively linear prediction is 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 robust decoding process are depicted. The quantized parameter $\mathbf{Q}[\tilde{v}] = v$ with $v \in \mathrm{QT}$ (QT: quantization table) is represented by the bit combination $\underline{x} = (x(0), ..., x(m), ..., x(M-1))$ 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 receiver and transmitter values by a hat on the (possibly modified) received values. In a conventional decoding scheme the received bit combination $\hat{\underline{x}}$ 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 additionally exploits a reliability information $\underline{p_e}$ with $p_e(m)$ being the error probability of bit $\hat{x}(m)$, to compute a set of transition probabilities $\mathrm{P}(\hat{\underline{x}} \mid \underline{x}^{(i)})$, $i = 0, 1, ..., 2^M - 1$, of a transition from any bit combination $\underline{x}^{(i)}$ at the transmitter to the received bit combination $\hat{\underline{x}}$. The computation of the transition probabilities depends on the chosen channel model and is discussed in section 2.

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* $\mathrm{P}(\underline{x}^{(i)} \mid \hat{\underline{x}})$, with $i = 0, 1, ..., 2^M - 1$, denoting the probability that $\underline{x}^{(i)}$ had been transmitted in the case that $\hat{\underline{x}}$ 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 $\hat{v}_{\mathrm{est}}$ referring to a given criterion. Two widely used estimators are discussed in this context in section 4.

If a mean square estimator is used, section 5 gives an efficient alternative solution to the computation of the a posteriori probabilities based on linear prediction that provides good results.

Finally, in section 6, the application to PCM coded speech is presented to prove the capabilities of the proposed robust speech decoding technique.

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

0-8186-7919-0/97 $10.00

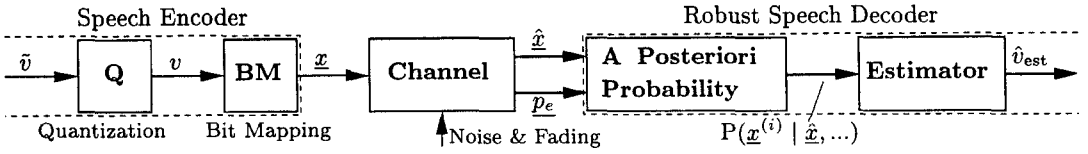

Figure 1: Conception of the new robust speech decoding technique

2. THE BIT RELIABILITY INFORMATION

2.1. The Channel Dependent Information

The transition probability from a transmitted bit $x^{(i)}(m)$ to a received bit $\hat{x}(m)$ can be written as

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

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

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

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

2.2. Channel Models and Their Bit Error Rates

For a simple fading channel with e.g. a BPSK modulation scheme the receiver output samples can be described by $\tilde{y}(m) = a \cdot x^{(i)}(m) + n(m)$ with $n(m)$ denoting the white Gaussian noise contribution and a being the fading factor. An instantaneous bit error rate for the detected bit $\hat{x}(m) = \mathrm{sign}[\tilde{y}(m)]$ is given in terms of log-likelihood values

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

assumed to be known at the receiver [6]. From (3) it can be seen that to any received value $\tilde{y}(m)$ 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*, whereas its mean value equals the well known BPSK bit error rate.

Assuming a channel coding scheme such as the soft-output Viterbi algorithm (SOVA) in combination with interleaving as proposed in [7], the instantaneous bit error rate is given by

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

being the soft-output value whose sign $\hat{x}(m) = \mathrm{sign}[L]$ equals the decoded hard-bit, $x^{(i)}(m)$ denoting the corresponding transmitted bit, and $\underline{\tilde{Y}}$ 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

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 that

$$\mathrm{P}(\underline{x}^{(i)} \mid \underline{\hat{x}}) = C \cdot \mathrm{P}(\underline{\hat{x}} \mid \underline{x}^{(i)}) \cdot \mathrm{P}(\underline{x}^{(i)}). \quad (5)$$

Varying the a posteriori term over i, we get the probability of any transmitted $\underline{x}^{(i)}$ if $\underline{\hat{x}}$ had been received. Here and in the following, the normalizing constant C is chosen such that $\sum_{l=0}^{2^M-1} \mathrm{P}(\underline{x}^{(l)} \mid \underline{\hat{x}}, ...) = 1$. 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 $v^{(i)}$.

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

$$\mathrm{P}(\underline{x}^{(i)} \mid \underline{\hat{x}}) \approx C \cdot \mathrm{P}(\underline{\hat{x}} \mid \underline{x}^{(i)}) \quad . \quad (6)$$

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 (6) 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).

The classical approaches of speech coding aim at minimizing the residual redundancy of codec parameters. However, due to the coding strategy, limited 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 sub-optimality can be exploited at the receiver side in the parameter estimation process. The a posteriori term in (5) can easily be extended to include these parameter correlations: The maximum information that is available at the decoder consists of the complete sequence of already received bit combinations resulting in $\mathrm{P}(\underline{x}_0^{(i)} \mid \underline{\hat{x}}_0, \underline{\hat{X}}_{-1})$ with $\underline{\hat{X}}_{-1} = (\underline{\hat{x}}_{-1}, \underline{\hat{x}}_{-2}, ...)$ and $\underline{\hat{x}}_{-n}$ denoting the bit combination n time instants[1] before the present one.

To compute this a posteriori term it is necessary to find a statistical model of the sequence of quantized parameters v_{-n}. It seems reasonable to discuss the sequence of

[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.

quantized parameters as a Markov process of 1st order, i.e. $P(\underline{x}_0 \mid \underline{x}_{-1}, \underline{x}_{-2}, ...) = P(\underline{x}_0 \mid \underline{x}_{-1})$. Solutions for higher order models can be derived. After some intermediate steps the solution can be given in terms of a recursion as

$$P(\underline{x}_0^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot P(\hat{\underline{x}}_0 \mid \underline{x}_0^{(i)}) \cdot \sum_{j=0}^{2^M-1} P(\underline{x}_0^{(i)} \mid \underline{x}_{-1}^{(j)}) \cdot P(\underline{x}_{-1}^{(j)} \mid \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}). \quad (7)$$

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. (7) 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 $P(\underline{x}_0^{(j)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ 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.

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

$$v_{MAP} = v^{(\nu)} \quad \text{with } P(\underline{x}_0^{(\nu)} \mid \hat{\underline{x}}_0, ...) = \max_i P(\underline{x}_0^{(i)} \mid \hat{\underline{x}}_0, ...),$$

while $P(\underline{x}_0^{(i)} \mid \hat{\underline{x}}_0, ...)$ denotes any of the a posteriori probabilities given in (6), (5), or (7) dependent on the chosen order of the model 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 minimizing 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

$$v_{MS} = \sum_{i=0}^{2^M-1} v^{(i)} \cdot P(\underline{x}_0^{(i)} \mid \hat{\underline{x}}_0, ...) . \quad (8)$$

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 $\sigma_{e_{MS}}^2 = \sigma_v^2 - \sigma_{v_{MS}}^2$. Because $\sigma_{e_{MS}}^2 \geq 0$ we can state that the variance $\sigma_{v_{MS}}^2$ of the estimated parameter v_{MS} is smaller than or equal to the variance σ_v^2 of the error free parameter v. 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 consequence, the MS estimated parameter according to eq. (8) is completely attenuated to zero if v has a zero mean. This is e.g. the case 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 a major advantage of the proposed robust speech decoding technique.

5. AN ALTERNATIVE SOLUTION: LINEAR PREDICTION

If a MS estimator is used, linear prediction can provide an alternative approximation of the a posteriori probabilities efficiently because it uses the same error criterion. The idea is to estimate a "predictive" a posteriori probability $P(\underline{x}_0^{(i)} \mid \hat{\underline{X}}_{-1})$ and finally to merge it with the channel dependent term $P(\hat{\underline{x}}_0 \mid \underline{x}_0^{(i)})$ to get the required probability

$$P(\underline{x}_0^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot P(\hat{\underline{x}}_0 \mid \underline{x}_0^{(i)}) \cdot P(\underline{x}_0^{(i)} \mid \hat{\underline{X}}_{-1}). \quad (9)$$

Let's model the unquantized parameter $\tilde{v}$ as an autoregressive process of order N following $\tilde{V}(z) = E(z)/(1-A(z))$ with $A(z) = \sum_{n=1}^{N} a_n \cdot z^{-n}$ and the zero mean innovation $E(z)$ having a symmetrical pdf $p_E(e)$. The pdf $p_E(e)$ as well as the prediction coefficients a_n have to be determined once and must be stored as a priori knowledge in the decoder. Alternatively, the coefficients a_n can be framewised updated requiring an LPC analysis of the MS estimated parameters $v_{-n_{MS}}$ located at the decoder side.

Knowing previous samples $\tilde{v}_{-1}, ..., \tilde{v}_{-N}$, the decoder has to perform a linear prediction:

$$v_0' = \sum_{n=1}^{N} a_n \cdot \tilde{v}_{-n} = \int_{-\infty}^{+\infty} \tilde{v}_0 \cdot p_{\tilde{V}}(\tilde{v}_0 \mid \tilde{v}_{-1}, ..., \tilde{v}_{-N})\, d\tilde{v}_0 \quad (10)$$

What we need to compute (9) is not a single predicted value v_0' but the pdf of $\tilde{v}_0$. Regarding v_0' as a deterministic constant, we can write $p_{\tilde{V}}(\tilde{v}_0 \mid \tilde{v}_{-1}, ..., \tilde{v}_{-N}) = p_E(\tilde{v}_0 - v_0')$ using $\tilde{v}_0 = e_0 + v_0'$ with e_0 being the innovation at time $n = 0$. The previous samples $\tilde{v}_{-1}, ..., \tilde{v}_{-N}$ are not available at the decoder side, thus they are approximated by the already MS-estimated parameters $v_{-n_{MS}}$. The resulting pdf is quantized leading to the approximation

$$P(\underline{x}_0^{(i)} \mid \hat{\underline{X}}_{-1}) \approx \int_{I_i} p_E(\tilde{v}_0 - v_0')\, d\tilde{v}_0 \quad (11)$$

with I_i being the i-th quantization interval. Thus the complete algorithm consists of linear prediction (10), shifting of $p_E()$ by v_0', evaluating (11) by numerical integration and finally using the result in the calculation of (9).

If a fixed set of coefficients a_n is used, the algorithmic complexity and the amount of required data ROM hardly depend on the AR model order. For an $M = 8$ bit parameter and an $L = 12$ bit resolution of $p_E()$, the linear predictive approach is about $2^{2 \cdot M}/2^L = 16$ times less complex than the 1st order Markov recursion (7), showing the main advantage in comparison to the Bayesian approach.

A further refinement to this method is motivated by the fact, that the process $\tilde{V}$ is mostly not a stationary one.

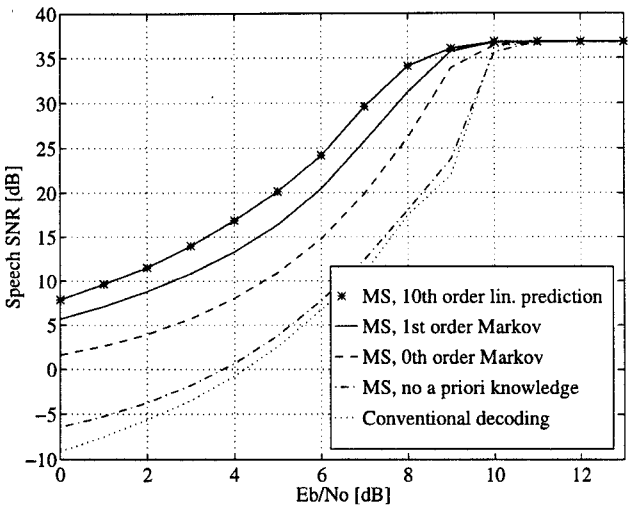

Figure 2: Robust speech decoding: A-law PCM over an AWGN channel using coherently detected BPSK; bit reliability information according to eq. (3) with $a = 1$.

Dependent on the estimated variance of the prediction error $v_{0_{MS}} - v'_0$, one out of a small set of pdf's $p_E()$ with different variances and/or shapes is chosen in eq. (11) leading to an improved performance especially in speech pauses.

6. AN APPLICATION EXAMPLE: PCM

In principle, the proposed algorithms can be applied to any speech codec. In a first experiment we simulated a simple PCM transmission over an AWGN channel assuming a coherent BPSK demodulation without channel coding. For this case we use the fading channel model discussed in 2.2 with $a = 1$. Fig. 2 shows five different simulation results in terms of speech SNR as a function of the E_b/N_0 ratio. The reference is the SNR of the conventionally decoded speech with the hard decision mechanism at the channel output. In comparison to that four different cases of mean square estimation are shown: Three orders of a priori knowledge are used according to eqns. (6), (5), and (7), respectively. Furthermore, the results of a 10th order linear prediction according to eq. (9) with coefficient update every 20 ms and a simple variance estimation are depicted. Four different pdf's $p_E()$ were used with respect to a different prediction error variance behaviour of speech segments.

In any case, the MS estimated speech degrades asymptotically to 0 dB with decreasing E_b/N_0, even in the case without a priori knowledge. Thus the algorithms provide an inherent muting mechanism. The shape of the curves strongly depends on the order of 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. Linear prediction of 10th order performs still better with gains of up to 17 dB. This leads to a significant enhancement of speech quality although e.g. long time correlations are yet unexploited and could refine the model of the speech further.

7. 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 a simple fading channel 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. For the mean square estimation, alternatively an efficient and well performing linear prediction technique was evaluated to provide estimates of the a posteriori probabilities. We applied the mean square estimator to PCM coded speech over an AWGN channel gaining up to 17 dB in the speech SNR. The subjective speech quality could be enhanced significantly.

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

ACKNOWLEDGEMENTS

The authors would like to thank C.G. Gerlach, W. Papen, and S. Heinen for a lot of inspiring discussions.

8. REFERENCES

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

[2] 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.

[3] 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.

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

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

[6] 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.

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

[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.