

SPEECH DECODING WITH ERROR CONCEALMENT USING RESIDUAL SOURCE REDUNDANCY

T. Fingscheidt and P. Vary

Institute of Communication Systems and Data Processing
Aachen University of Technology, Templergraben 55, D-52056 Aachen, Germany
Tim.Fingscheidt@ind.rwth-aachen.de, <http://www.ind.rwth-aachen.de/~tim>

ABSTRACT

In digital mobile communication systems there is a need for error concealment techniques to reduce the subjective effects of residual bit errors which have not been eliminated by channel decoding.

This contribution presents a new approach for optimum estimation of speech codec parameters. It can be applied to any speech codec standard if a reliability information about the channel decoded bits is available (e.g. soft-output Viterbi algorithm – SOVA [6]). The proposed method exploits the residual source redundancy and includes an inherent muting mechanism leading to a graceful degradation of speech quality in case of adverse transmission conditions. In the case of an error free channel, bit exactness as required by the standards can be preserved. This approach is applied here to the GSM full rate codec.

1. THE CONCEPTION OF SOFT INPUT SPEECH DECODING

In digital communication systems residual redundancy of the transmitted parameters can be observed even if source encoding techniques are used. This is due to the speech coding strategy, the limited processor resources, and the signal delay constraints. As already mentioned by Shannon [1] this source coding sub-optimality can be used at the receiver side to enhance the decoding process. Former publications on error concealment exploiting this residual source redundancy had been restricted to the MAP error criterion [2, 3], or they assumed that previously received parameters had to be known exactly, i.e. without errors [4].

This proposal focuses on the enhancement of the speech quality rather than on the minimization of a bit or symbol error rate [5]. Furthermore, the proposed error concealment technique is able to include parameter individual estimators without taking into consideration idealizing assumptions about previously received parameters.

Let us consider a specific codec parameter $\tilde{v} \in \mathbb{R}$ which is coded by M bits. The quantized parameter $Q[\tilde{v}] = v$ is represented by the bit combination $\underline{x} = (x(0), \dots, x(M-1))$ with $x(m) \in \{-1, +1\}$. 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 $i \in \{0, 1, \dots, 2^M - 1\}$ to denote the quantized parameter.

In Fig. 1 the soft input speech decoder is depicted. Assuming a channel decoding scheme such as the soft-output

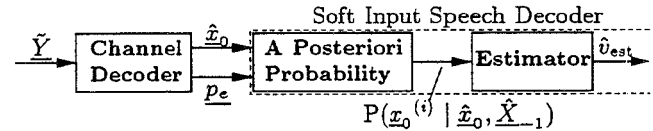


Figure 1: Conception of the soft input speech decoder

Viterbi algorithm (SOVA) in combination with interleaving as proposed in [6], the instantaneous bit error rate for the channel decoded bit $\hat{x}(m)$ is given by

$$p_e(m) = \frac{1}{1 + \exp |L(m)|} \quad (1)$$

with $L(m) = \ln \frac{P(x^{(i)}(m) = +1 | \underline{Y})}{P(x^{(i)}(m) = -1 | \underline{Y})}$

being the soft-output value whose sign $\hat{x}(m) = \text{sign}[L]$ equals the decoded hard-bit, $x^{(i)}(m)$ denoting the corresponding transmitted bit, and \underline{Y} being the received sequence of symbols that is input to the channel decoder. Using the instantaneous bit error rate $p_e(m)$ of the received bit $\hat{x}(m)$, a bit transition probability can be formulated as

$$P(\hat{x}(m) | x^{(i)}(m)) = \begin{cases} 1 - p_e(m) & \text{if } \hat{x}(m) = x^{(i)}(m) \\ p_e(m) & \text{if } \hat{x}(m) \neq x^{(i)}(m) \end{cases} \quad (2)$$

Because of the integrated interleaving scheme providing independent bit errors at least within any bit combination $\hat{\underline{x}}$, a set of 2^M transition probabilities from every transmitted parameter $\underline{x}^{(i)}$ to the received one $\hat{\underline{x}}$ can be computed by

$$P(\hat{\underline{x}} | \underline{x}^{(i)}) = \prod_{m=0}^{M-1} P(\hat{x}(m) | x^{(i)}(m)) \quad (3)$$

In the following, $\hat{\underline{x}}_{-n}$ denotes the bit combination n time instants¹ before the present bit combination $\hat{\underline{x}}_0$.

For the estimation of speech codec parameters at the receiver, *a posteriori probability* terms providing information about any transmitted parameter index i are required. The complete receiver knowledge comprising all previously received bit combinations $\hat{\underline{x}}_0, \hat{\underline{X}}_{-1}$ with $\hat{\underline{X}}_{-1} = \hat{\underline{x}}_{-1}, \hat{\underline{x}}_{-2}, \dots$ shall be exploited. To compute the *a posteriori* terms it is necessary to find a statistical model of the sequence of quantized parameters. In a first approach, the sequence of quantized parameters is modelled as a Markov process of 1st order, i.e. $P(\underline{x}_0 | \underline{x}_{-1}, \underline{x}_{-2}, \dots) = P(\underline{x}_0 | \underline{x}_{-1})$. Solutions for

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

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

higher order models can be derived. After some intermediate steps the solution can be given in terms of a recursion as

$$P(\underline{x}_0^{(i)} | \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = C \cdot P(\underline{x}_0 | \underline{x}_0^{(i)}) \cdot \sum_{j=0}^{2^M-1} P(\underline{x}_0^{(i)} | \underline{x}_{-1}^{(j)}) \cdot P(\underline{x}_{-1}^{(j)} | \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2}). \quad (4)$$

The constant C guarantees $\sum_{i=0}^{2^M-1} P(\underline{x}_0^{(i)} | \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = 1$. The terms $P(\underline{x}_0^{(i)} | \underline{x}_{-1}^{(j)})$ contain the a priori knowledge (redundancy information) which has to be measured once over a large data base. These measurements can be stored at the receiver in a ROM table of size $2^M \times 2^M$. Furthermore it should be noted that the term $P(\underline{x}_{-1}^{(j)} | \hat{\underline{x}}_{-1}, \hat{\underline{X}}_{-2})$ is nothing else but the resulting a posteriori probability $P(\underline{x}_0^{(j)} | \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ from the previous time instant.

2. INDIVIDUAL PARAMETER ESTIMATION AND SIMULATION RESULTS

The maximum a posteriori probability (MAP) estimator follows the simple criterion

$$v_{MAP} = v^{(\nu)} \text{ with } \nu = \arg \max_i P(\underline{x}_0^{(i)} | \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}). \quad (5)$$

The optimum decoded parameter in a MAP sense v_{MAP} always equals one of the codebook/ quantization table entries minimizing the decoding error probability.

In Fig. 2 the results of a complete GSM simulation using the COSSAP GSM library [7] with speech and channel coding, interleaving, modulation, a channel model, demodulation and equalization are depicted for different carrier-to-interferer ratios C/I and user speeds. The channel model represents a typical case for an urban area (TU) taking into account 6 characteristic propagation paths. Although the SNR surely is not the optimum measure for speech quality, informal listening tests as well as SNR measurements show a significant superiority of MAP estimation in comparison to the conventional GSM decoder with a frame repetition algorithm [8] as error concealment method.

For a wide area of speech codec parameters such as spectral coefficients, gain factors, etc., the minimum mean square error (MS) estimation criterion performs even better than the MAP one. The optimum decoded parameter v_{MS} in a mean square sense equals

$$v_{MS} = \sum_{i=0}^{2^M-1} v^{(i)} \cdot P(\underline{x}_0^{(i)} | \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}). \quad (6)$$

According to the well known orthogonality principle of the linear mean square estimation the variance of the estimation error $e_{MS} = v_{MS} - v$ is $\sigma_{e_{MS}}^2 = \sigma_v^2 - \sigma_{v_{MS}}^2$ with $\sigma_{e_{MS}}^2 \geq 0$. Therefore, we can state that $\sigma_{v_{MS}}^2 \leq \sigma_v^2$. In the case of a worst case channel with $p_e = 0.5$, the MS estimated parameter according to eq. (6) 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. By MS estimation of all non-integer parameters of the GSM full rate speech codec (curves marked by "o" in Fig. 2), the speech

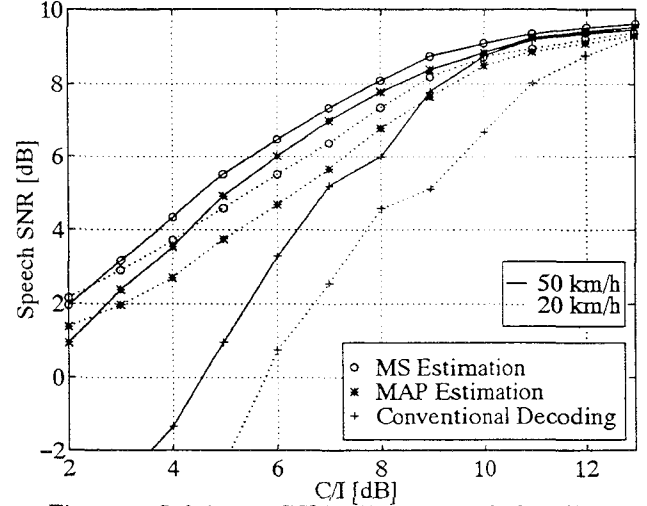


Figure 2: Soft input GSM full rate speech decoding

quality can be enhanced further in comparison to the MAP estimation. Hard annoying effects disappear and are turned to a slightly modulated and noisy speech quality including the inherent muting effect. This is subjectively judged to be much more pleasant than the typical click noise and the synthetic frame repetition effects of digital transmission systems. It should be noted that this approach can be applied to different source coding schemes such as ADPCM and CELP using individual parameter estimators and Markov model orders. Simulated speech files can be found in the web at location <http://www.ind.rwth-aachen.de/~tim> and will be presented at the workshop.

ACKNOWLEDGEMENTS

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

3. REFERENCES

- [1] C.E. Shannon, "A Mathematical Theory of Communication", *Bell Systems Technical Journal*, vol. 27, pp. 379-423, July 1948.
- [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] J. Hagenauer, "Source- Controlled Channel Decoding", *IEEE Transactions on Communications*, vol. 43, no. 9, pp. 2449-2457, September 1995.
- [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] T. Fingscheidt and P. Vary "Error Concealment by Soft-bit Speech Decoding", *Proc. of ITG-Fachtagung Sprachkommunikation*, pp. 7-10, September 1996.
- [6] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and its Applications", *Proc. of GLOBE-COM*, pp. 1680-1686, 1989.
- [7] Synopsys Inc., Mountain View, CA, USA, "COSSAP — Model Libraries Volume 3", 1995.
- [8] "Recommendation GSM 06.11, Substitution and Muting of Lost Frames for Full Rate Speech Traffic Channels", *ETSI/TC SMG*, February 1992.