Robust GSM Speech Decoding Using the Channel Decoder's Soft Output

Tim Fingscheidt, Olaf Scheufen

Institute of Communication Systems and Data Processing
Aachen University of Technology, Templergraben 55, D – 52056 Aachen, Germany
Tel.: ++49 241 806963, Fax: ++49 241 8888186,
E-Mail: Tim.Fingscheidt@ind.rwth-aachen.de http://www.ind.rwth-aachen.de

ABSTRACT

In the digital mobile radio system GSM (Global System for Mobile Communications) there is a need for reducing the subjective effects of residual bit errors by error concealment techniques. Due to the fact that the standard does not specify these algorithms bit exactly, there is room for new solutions to improve the decoding process.

This contribution presents a new approach for optimum estimation of speech codec parameters [7] applied to the GSM system. It requires a soft-output channel decoder (e.g. soft-output Viterbi algorithm – SOVA [8]) providing a bit reliability information for the proposed parameter estimation process. Additionally, a priori knowledge about the residual redundancy in the sequence of codec parameters is exploited. The new method includes an inherent muting mechanism leading to a graceful degradation of speech quality in case of adverse transmission conditions. If the channel is error free, bit exactness as required by the GSM standard is preserved.

1. INTRODUCTION

The GSM recommendations [2] propose a simple solution to the problem of error concealment. The perceptually most significant 50 bits of any frame of 260 bits are checked by a 3-bit CRC check. A *bad frame indicator* (BFI) is set, if the CRC detects an error. Optionally, additional information can be regarded for the BFI decision, e.g. [4]. If a frame is marked as "bad", the last "good" frame is repeated. Consecutive bad frames are gradually muted resulting in complete silence after a period of 320 ms.

The single-bit BFI reliability information for a frame of 260 bits has a very coarse resolution. In contrast to that the proposed new speech decoding technique is able to use error probabilities for each bit. The simple frame repetition mechanism of the GSM recommendations enhances the speech quality significantly. However, our new approach which includes statistical modelling of codec parameters exploits the redundancy of speech codec parameters more effectively.

Let us consider a specific GSM speech codec parameter $\tilde{v} \in \mathbb{R}$ which is coded by M bits. In Fig. 1 the coding and transmission over a noisy channel as well as the conception of the proposed robust decoding process are depicted. The quantized parameter $\mathbf{Q}[\tilde{v}] = v$ with $v \in$ QT (QT: quantization table) is represented by the bit combination $\underline{x} = (x(0), ..., x(m), ..., x(M-1))$ consisting of M bits (bit mapping BM). The bits are assumed to be bipolar, i.e. $x(m) \in \{-1, +1\}$. Each 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 the conventional GSM decoder [1] the received bit combination $\hat{\underline{x}}$ denoting the channel decoder's hard output is input to an "inverse bit mapping" scheme, i.e. the appropriate parameter $\hat{v}$ is addressed in a quantization table.

The new robust speech decoding technique uses the channel decoder's soft output in terms of so-called log-likelihood values $\underline{L}^{(\hat{x})}$ 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 $\hat{\underline{x}}$ with $\hat{x}(m) = \mathrm{sign}[L^{(\hat{x})}(m)]$. The computation of the transition probabilities is discussed in section 2.1.

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* $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ of any transmitted $\underline{x}_0{}^{(i)}$ in the case of the received bit combinations $\hat{\underline{x}}_0, \hat{\underline{X}}_{-1}$ with $\hat{\underline{X}}_{-1} = (\hat{\underline{x}}_{-1}, \hat{\underline{x}}_{-2}, ...)$ and $\hat{\underline{x}}_{-n}$ denoting the bit combination n time instants[1] before the present one (section 2.2.).

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. All GSM speech codec parameters are discussed in this context in section 2.3.

[1] The term "time instant" denotes any moment when the regarded parameter is received. For LPC parameters this is a frame instant, whereas e.g. for LTP lags it is a subframe instant.

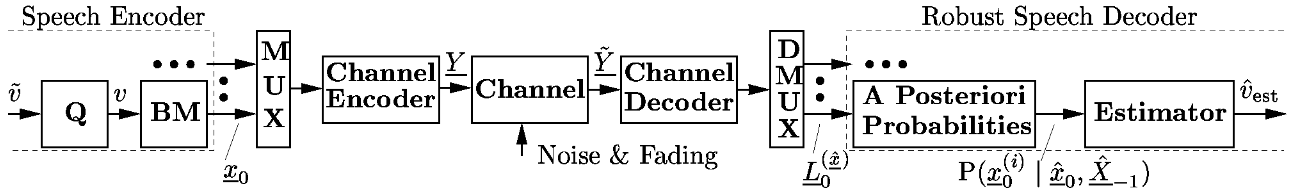

Figure 1: Conception of the new robust speech decoding technique

Finally, in section 3, simulations of a GSM transmission are presented to prove the capabilities of the proposed technique.

2. THE ROBUST SPEECH DECODER

2.1. The Channel Decoder's Soft Output

The robust speech decoding scheme requires a bit reliability information from the channel decoder. Thus e.g. a soft-output Viterbi algorithm (SOVA) [8] or an algorithm after Bahl et al. [9] can be used. The soft output consists of a log-likelihood value

$$L^{(\hat{x})}(m) = \ln \frac{\mathrm{P}(x^{(i)}(m) = +1 \mid \underline{\tilde{Y}})}{\mathrm{P}(x^{(i)}(m) = -1 \mid \underline{\tilde{Y}})} \qquad (1)$$

with $x^{(i)}(m)$ denoting the transmitted bit, and $\underline{\tilde{Y}}$ being the received sequence of symbols that is input to the channel decoder. The instantaneous bit error rate to the decoded hard-bit $\hat{x}(m) = \mathrm{sign}[L^{(\hat{x})}(m)]$ is given by $p_e(m) = (1 + \exp|L^{(\hat{x})}(m)|)^{-1}$.

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

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

Because of bit shuffling and the interleaving scheme used in GSM, the cascade of the channel encoder, the discrete channel, and the channel decoder can approximately be modelled as memoryless. Thus the transition probability of a bit combination reads

$$\mathrm{P}(\underline{\hat{x}} \mid \underline{x}^{(i)}) = \prod_{m=0}^{M-1} \mathrm{P}(\hat{x}(m) \mid x^{(i)}(m)). \qquad (3)$$

2.2. A Posteriori Probabilities Computation

The robust speech decoding process aims at exploiting the maximum information that is available at the decoder. Accordingly, the a posteriori probabilities $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \underline{\hat{x}}_0, \underline{\hat{X}}_{-1})$ including the complete sequence of received bit combinations are to be computed. In order to achieve this, a statistical model of the sequence of quantized parameters v_{-n} is required. Table 1 shows the amount of redundancy for all parameters used in the GSM full-rate speech codec [6]. The entropy values were measured over 270 s of English and German speech spoken by male and female speakers. The first log-area ratio (LAR 1) e.g. is coded with $H_0 = 6$ bits, whereas its entropy is only 5.43 bit. The conditional entropy taking a first order Markov model into account is about one additional bit less, thus adjacent log-area ratios are relative strongly correlated. This observation holds for most GSM codec parameters except the RPE pulses and the RPE grid position. These can be modelled as a Markov process of 0th order [5] with $\mathrm{P}(\underline{x}_0 \mid \underline{x}_{-1}, \underline{x}_{-2}, \ldots) = \mathrm{P}(\underline{x}_0)$, while the other parameters profit from the 1st order model $\mathrm{P}(\underline{x}_0 \mid \underline{x}_{-1}, \underline{x}_{-2}, \ldots) = \mathrm{P}(\underline{x}_0 \mid \underline{x}_{-1})$.

The required a posteriori probabilities for a first order Markov model can recursively be computed [7] :

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

The first order a priori knowledge $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \underline{x}_{-1}{}^{(j)})$ of size $2^M \times 2^M$ has to be measured in advance and must be available to the robust speech decoder. The term $\mathrm{P}(\underline{x}_{-1}{}^{(j)} \mid \underline{\hat{x}}_{-1}, \underline{\hat{X}}_{-2})$ in eq. (4) is nothing else but the resulting a posteriori probability $\mathrm{P}(\underline{x}_0{}^{(j)} \mid \underline{\hat{x}}_0, \underline{\hat{X}}_{-1})$ from the previous time instant.

The set of a posteriori probabilities is now input to parameter individual estimators.

2.3. The Parameter Estimation

For a wide area of speech codec parameters the minimum mean square error criterion (MS) is appropriate. We found out that the non-integer GSM codec parameters can be well estimated using a MS estimator. In contrast to that, the estimation of a pitch period or the RPE grid position 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 the GSM speech codec parameter estimation.

2.3.1. LTP lag and RPE grid position

The simplest estimator that can be applied to the LTP lag as well as the RPE grid position is the MAP estimator. It is the one requiring the least computational complexity following the rule

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

$[\frac{\text{bit}}{\text{parameter}}]$	LAR No.								LTP		RPE		
	1	2	3	4	5	6	7	8	Lag	Gain	Grid	Max.	Pulse
$H_0(\underline{x}_0)$	6	6	5	5	4	4	3	3	7	2	2	6	3
$H(\underline{x}_0)$	5.43	4.88	4.75	4.53	3.73	3.76	2.84	2.88	6.31	1.88	1.96	5.39	2.86
$H(\underline{x}_0 \mid \underline{x}_{-1})$	4.46	4.29	4.18	4.09	3.37	3.39	2.49	2.46	5.75	1.74	1.96	4.29	2.86

Table 1: Perfect information content $H_0(\underline{x}_0)$, entropy $H(\underline{x}_0)$, conditional entropy $H(\underline{x}_0 \mid \underline{x}_{-1})$

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

2.3.2. The other parameters

The other parameters can be sufficiently well estimated by a MS estimator. The optimum decoded parameter v_{MS} in a mean square sense equals

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

According to the well-known orthogonality principle of the linear mean square (MS) estimation the variance of the estimation error $e_{MS} = v_{MS} - v$ is simply $\sigma^2_{e_{MS}} = \sigma^2_v - \sigma^2_{v_{MS}}$. Because $\sigma^2_{e_{MS}} \geq 0$ we can state that the variance $\sigma^2_{v_{MS}}$ of the estimated parameter v_{MS} is smaller than or equal to the variance σ^2_v of the error free parameter v. Assuming a worst case channel with $p_e = 0.5$ the a posteriori probabilities degrade to $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) = \mathrm{P}(\underline{x}_0{}^{(i)})$. As a consequence, the MS estimated parameter according to eq. (6) is completely attenuated to zero if v has a zero mean. This is exactly the case for the RPE block maximum with sign. Thus the MS estimation of the RPE block maximum results in an inherent muting mechanism providing a graceful degradation of speech. This is a major advantage of the proposed robust speech decoding technique.

2.3.3. Alternative: The LTP lag

As an alternative to the MAP estimation, we tried to estimate the LPC residual d_n itself rather than the integer lag parameter L. Using the LTP synthesis equation $d_n = e_n + b_{n_{MS}} \cdot d_{n-L}$ with e_n denoting the RPE signal and $b_{n_{MS}}$ being the already estimated LTP gain we performed a MS estimation of the LPC residual by

$$d_{n_{MS}} = e_n + b_{n_{MS}} \cdot \sum_{i=40}^{120} (d_{n-i})_{MS} \cdot \mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$$

with $\mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1})$ being the a posteriori probabilities of the LTP lag in the valid range 40 ... 120. It turns out that this estimation scheme performs slightly better than the MAP estimator, but not justifying the additional computational load.

2.3.4. Alternative: The log-area ratios

To prove whether the MS estimation of log-area ratios is optimum, we performed a minimization of the spectral distance [10]

$$D^2 = 200 \cdot (\log_{10} e)^2 \cdot \sum_{k=1}^{\infty} [c_0(k) - \hat{c}_0(k)]^2 \quad [dB] \qquad (7)$$

with $c_0(k)$ and $\hat{c}_0(k)$ denoting cepstral coefficients. Minimizing the spectral distance is thus equivalent to MS estimation of the cepstral coefficients. Since any cepstral coefficient depends on any log-area ratio, we assumed LAR3 ... LAR8 for simplicity to be already MS estimated. Accordingly, we varied over all combinations of LAR1 and LAR2 to compute $2^6 \times 2^6$ sets of cepstral coefficients $c_0^{(i,j)}$ each truncated to the length of 16 coefficients. Assuming LAR1 (i.e. $\underline{x}_0{}^{(i)}$) being independent from LAR2 (i.e. $\underline{z}_0{}^{(j)}$), the optimum set of cepstral coefficients equals

$$\begin{aligned} c_0(k)_{MS} = \sum_i \sum_j c_0^{(i,j)}(k) \cdot \mathrm{P}(\underline{x}_0{}^{(i)} \mid \hat{\underline{x}}_0, \hat{\underline{X}}_{-1}) \\ \cdot \mathrm{P}(\underline{z}_0{}^{(j)} \mid \hat{\underline{z}}_0, \hat{\underline{Z}}_{-1}) \qquad k = 1, ..., 16. \end{aligned} \qquad (8)$$

Thus the LPC synthesis filter runs using 16 reflection coefficients. It turns out that 0.1 ... 0.2 dB of speech SNR can be gained by this method resulting in some cases in a slightly better speech quality. Nevertheless, its complexity is very high and thus the MS estimation of LAR's can be seen as near-optimum.

3. GSM SIMULATIONS

We applied the above described conception of robust speech decoding to the GSM full-rate speech decoder [1]. For simplicity, any codec parameter has been modelled as a Markov process of 1st order, although Table 1 shows that for two parameters a model order of zero would be sufficient. The parameter estimators are chosen as described in sections 2.3.1. and 2.3.2.

In Fig. 2 the results of a complete GSM simulation using the COSSAP GSM library [11] with speech and channel coding, interleaving, modulation, a channel model, demodulation and equalization are depicted. The channel model represents a typical case for an urban area (TU) regarding 6 characteristic propagation paths [3]. The reference conventional GSM decoder performs its error concealment by a frame repetition algorithm as proposed in [2].

Although the SNR surely is not the optimum measure for speech quality, informal listening tests as well as SNR measurements show a significant superiority

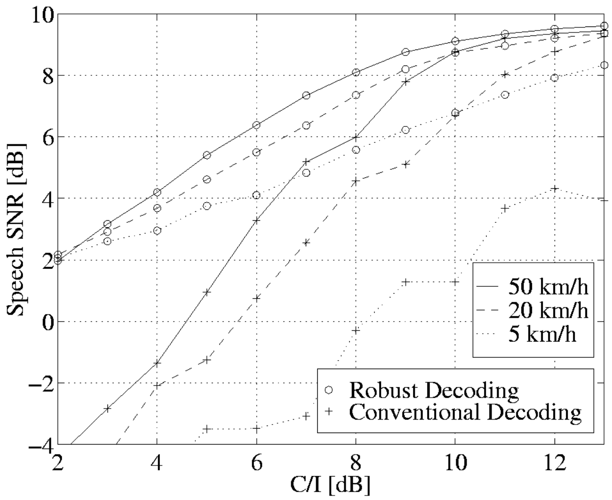

Figure 2: Robust speech decoding in GSM

of the robust decoder in comparison to the conventional decoding scheme in all situations of vehicle speed and C/I ratio. The new error concealment technique provides a quite good speech quality up to C/I = 6 dB, whereas the conventional frame repetition produces severe distortions already at C/I = 7 dB. Even long error bursts caused by a low vehicle speed can be decoded sufficiently by the new technique. In the robust decoding simulation, the hard annoying clicks caused by CRC failures and the synthetic sounds of the frame repetition disappeared completely and turned into a slightly noisy or modulated speech. This leads to a significant enhancement of speech quality. The conference CD provides two audio samples referring to Fig. 2 at $C/I = 4\,dB$ and 20 km/h user speed: Conventional decoding (1) and robust speech decoding (2).

4. CONCLUSION

In this paper we proposed a new error concealment technique for the GSM system. It is able to exploit a bit reliability information provided by a soft-output channel decoder. The amount of residual redundancy in the case of a simple 1st order Markov modelling of the codec parameters justified the additional usage of a priori knowledge about codec parameters.

We derived the optimum a posteriori probability of a bit combination to be used in parameter individual estimators. Four different estimators were discussed concerning the GSM codec parameters. It turns out that the mean square estimation of the RPE block maximum is able to perform a graceful degradation of speech in case of decreasing quality of the transmission link because of its inherent muting mechanism. On the other hand, bit exactness in the case of an error free channel is preserved. The complexity of the proposed robust speech decoder is fully scalable by the order of the Markov models used, and by the amount of bits involved in the parameter estimation. The additional speech decoder complexity of our implementation estimating complete parameters is about 3.3 MOPS.

Finally, GSM simulations showed that the subjective speech quality could be enhanced significantly in comparison to a conventional frame repetition algorithm.

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) within the program "Mobilkommunikation". The authors would like to thank Prof. P. Vary and S. Heinen for a lot of helpful discussions.

5. REFERENCES

[1] "Recommendation GSM 06.10, GSM Full Rate Speech Transcoding", *ETSI/TC SMG*, February 1992.

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

[3] "Recommendation GSM 05.05, Radio Transmission and Reception", *ETSI/TC SMG*, February 1992.

[4] D. Sereno, "Frame Substitution and Adaptive Post-Filtering in Speech Coding", *Proc. of EUROSPEECH'91*, pp. 595–598, September 1991.

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

[6] 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.

[7] 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.

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

[9] L.R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate", *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[10] R. Hagen, "Spectral Quantization of Cepstral Coefficients", *Proc. of ICASSP'94*, vol. 1, pp. 509–512, April 1994.

[11] Synopsys Inc., Mountain View, CA, USA, "COSSAP — Model Libraries Volume 3", 1995.