# QUALITY ENHANCEMENT OF CODED AND CORRUPTED SPEECHES IN GSM MOBILE SYSTEMS USING RESIDUAL REDUNDANCY

Thomas Hindelang[1], Wen Xu[2] and Christian Erben[2]

[1] Department of Communications Engineering, Technical University of Munich,
Arcisstr. 21, 80290 Munich, Germany
(e-mail: thomas@LNT.e-technik.tu-muenchen.de)

[2] Department of Cellular Product Development (PN KE CP T32), Siemens AG,
Hofmannstr. 51, 81359 Munich, Germany

## ABSTRACT

There is often residual redundancy remaining in coded speech data, even if a powerful speech codec (e.g. the full rate coder used in GSM mobile communications) is employed. By using such redundancy together with the information provided by the channel decoder, such as soft output (L-value), the number of channel bits inverted by the decoder, or a cyclic redundancy check, the bit error rate can be further reduced and a more graceful degradation of speech quality can be achieved, especially under bad channel conditions. In this paper, we report on the study with regard to this aspect for GSM full rate speech transmission and error concealment. The algorithms developed can beeasily implemented with a currently available DSP designed for GSM mobile phones.

## 1. INTRODUCTION

Traditionally, source and channel coding functions of a communication system have been designed independently of each other. This is justified by Shannon's principle which says that there is no performance loss if the two functions are separately treated. However Shannon's theory is an asymptotic result that permits unlimited delay and complexity. In Practice because of the incomplete knowledge on the source signal and transmission environment, even very complex coders may achieve a suboptimal performance. Therefore there may be some kind of residual redundancy left in data after source coding. Such residual redundancy has been found to be useful to combat the channel errors so that a more reliable decoding can be achieved [2, 3].

In GSM system, speech is coded in frames. Here the redundancy is characterized mainly by the statistical correlation among certain temporally neighboring coefficients or bits. In our investigations the redundancy among bits is first utilized to realize a source-controlled channel decoding [2], so that the bit error rate can be further reduced. Then based on the redundancy among coefficients an improved criterion for frame error detection is developed for GSM full rate speech transmission. As a result, both subjective and objective performance gains can be obtained, especially in a noisy transmission environment.

## 2. GSM FULL RATE SPEECH AND CHANNEL CODING

The GSM full rate speech and channel coding scheme are briefly explained here. Stress is placed on what is relevant for this work. The GSM speech coding is done in 3 steps (see Figure 1).

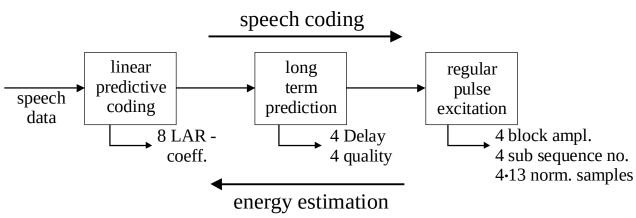


*Figure 1: Block diagram of the GSM full rate speech coding and the energy estimation.*

In the first step, a linear predictive coding (LPC) is conducted, where the spectral power density of a frame with a duration of 20 ms is rebuilt using a digital filter with 8 coefficients. The coefficients are transformed and quantized to 8 so-called LAR parameters.

The second step is a long term prediction (LTP). A subframe of 5 ms is correlated with the 3 previous subframes. The highest correlation and the delay at the current time are calculated. As this is done 4 times per frame, 4 delay and 4 quality parameters are obtained.

The third step is a regular pulse excitation (RPE). A low pass filtering is first carried out. Then, for each subframe of 5 ms 4 subsequences - each with 13 samples - are built. Only the subsequence with the largest amplitude is taken into account. So for each subframe the block amplitude, the number of the subsequence and 13 normalized samples are transmitted.

After the full rate coding there are a total of 76 coefficients. They are quantized - dependent on their importance - into 2 to 7 bits. This results in 260 bits per frame or 13 kbit/s. According to the subjective importance the 260 bits are divided into 3 classes. 50 most important bits are put into the class 1a which are very sensitive to errors, 132 bits into the class 1b and 78 bits into class 2 (insensitive to errors).

In further investigations it was found that errors in the 12 most significant bits of the class 1a (Bit No. 1 to 12 as specified in GSM Recommendation 06.10 [3]) often lead to loud and noisy cracks in speech reproduction. These 12 bits are namely the first three bits of the first LAR coefficient, the first and second bits of the second and third LAR coefficient and the first (most significant) bit of the fourth LAR coefficient, and the first bit of each of the four block amplitudes.

For error protection, the GSM channel coding is done as shown in Figure 2.

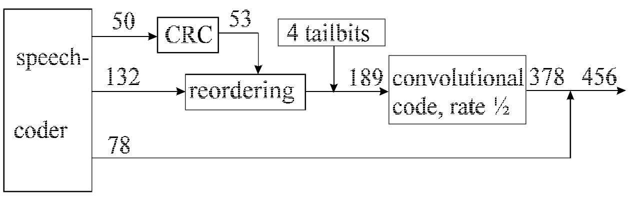


*Figure 2: Channel coding in the GSM Full Rate system*

For the 50 class 1a bits a cyclic redundancy check (CRC) with 3 bits is calculated, which will be used for bad frame detection at the receiver side. Together with the class 1b bits they are reordered and encoded by a convolutional code with memory 4 and rate ½. For termination 4 tail bits are added. The class 2 bits are transmitted without channel coding. This results in 456 bits per frame, i.e. a bit rate of 22.8 kbit/s.

## 3. ERROR CORRECTION DECODING

In the source-controlled channel decoding method, the *a priori* information L(u) is used to modify the metric of the Viterbi algorithm (VA). To build the L-value from the *a posteriori* decoding soft output is required. Therefore the so-called "soft output VA" introduced in [1] is employed here.

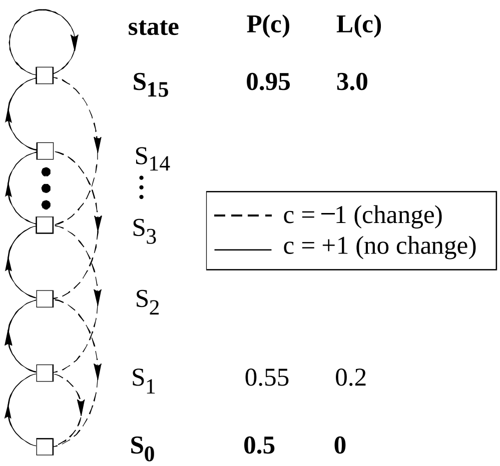


*Figure 3: Generation of a priori information L(c) for bits with low probability of bit changing.*

We implemented a simple algorithm [2] to build L(u) from the soft outputs of the previous frame $L(\hat{u}_{k-1})$ and the estimated bit changing probability L(c)

$$L(u_k) = L(\hat{u}_{k-1}) \boxplus L(c_k) = L(\hat{u}_{k-1} \oplus c_k)$$
$$= \log \frac{1 + e^{L(\hat{u}_{k-1})} \cdot e^{L(c_k)}}{e^{L(\hat{u}_{k-1})} + e^{L(c_k)}} \qquad (1)$$

In this work the following approximation was used

$$L(u_k) \approx \mathrm{sign}(L(\hat{u}_{k-1})) \cdot \mathrm{sign}(L(c_k)) \cdot \min(|L(\hat{u}_{k-1})| \cdot |L(c_k)|) \quad (2)$$

Since speech is a highly non-stationary process the L-values L(c) have to be calculated newly for every frame. In comparison to the original HUK-algorithm presented in [2] some modifications were made as shown in Figure 3. 22 of the important bits (21 of which belong to class 1a) have been found to possess an average bit changing probability between 0.05 and 0.30. Since all the correlated bits have a low probability of bit changing only positive L-values need to be calculated.

## 4. BAD FRAME CONCEALMENT

After channel decoding a data frame may still have too many errors so that it becomes useless (bad frame). To detect such a bad frame effectively we make again use of the residual redundancy in speech data to extend the CRC criterion used in the GSM standard. The aim is to detect a frame which possesses good quality as good and a frame of bad quality as bad (the bad frame indication (BFI) flag will be set). The probability that the decoded bit have errors can be calculated using the CRC specified in [4] and the reliability information which is generated e.g. by a soft output Viterbi algorithm (SOVA) [1]. Moreover, the probability of a combination of speech coefficients can be computed based on the coefficient distribution. Together with the bad frame statistic of the last n frames (e.g. n=20) the current frame can be indicated as good or bad.

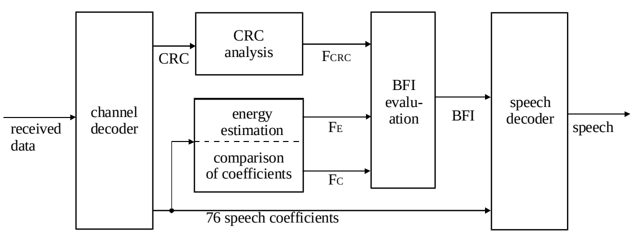


*Figure 4: Scheme of the error concealment unit.*

As it is complicated to calculate these probabilities we derived a simple algorithm especially for the GSM full rate system. The block diagram of the new error concealment is shown in Figure 4 and consists of the following three factors:

- The CRC as specified in GSM standard.
- The LTP-Delay which can only take a value between 40 and 120. As the coefficient is quantized with 7 bit, it is clear that if it takes values between 0 and 39 or 121 and 127 an error has occurred. All bits in this coefficient are found to be very important for speech quality. Especially in the case of slow fading channels which are typical in mobile communications a bit error in this coefficient usually indicates errors in other coefficients and may lead to a bad speech quality. Thus, if the value is outside the limit, a frame is declared as bad.

| estimated energy [dB] | <10 | <15 | <20 | <25 | <30 | <35 | <40 | <45 | <50 | <55 | >55 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| max. difference [dB] | 39 | 36 | 30 | 27 | 24 | 20 | 16 | 12 | 9 | 6 | 2 |

*Table 1: Maximally allowed difference of the energy between two frames dependent on the estimated energy of the current frame.*

- The energy difference between two neighboring frames is limited to certain values as found in our investigations (see Table 1). If the difference is bigger than an allowed value or if the absolute energy is greater than a maximum value, e.g. 57 dB, the frame will be classified as bad. Here an energy of 0 dB means a very low noise on the mobile phone (it can be chosen arbitrarily). The energy is estimated from some of the received 76 speech coefficients, namely the coefficient $X_{max}$ in the regular pulse excitation (RPE), the quality $b_c$ and the delay N in the long term prediction (LTP) and the reflection coefficients $k_i$ in the linear predictive coding (LPC). It is in fact an inverse process of the GSM full rate encoding as shown in Figure 1. A detailed derivation of the energy estimation has been submitted for patent.

## 5. SIMULATION RESULTS

With the method described in section 3 it is possible to correct more than 50% of the corrupted bits which belong to the 12 most important bits as described above. 10 of these 12 bits have an average changing probability less than 0.25 and 5 of these bits even less than 0.05. The bit error rate of all class 1a bits can be reduced around 30%. The gain in the 12 most significant bits is shown in figure 5. The signal-to-noise ratio can be lowered by 0.8 dB to obtain the same bit error rate.

A received speech frame will be declared as bad (i.e. BFI = 1) if one or a proper combination of the three factors indicates that the frame is bad. In a case study we implemented the methods described above for the GSM full rate speech coding where the speech decoding was done as described in GSM Recommendations [4, 5]. The simulations show that the number of undetected and incorrectly detected bad frames can be reduced by about 35%.

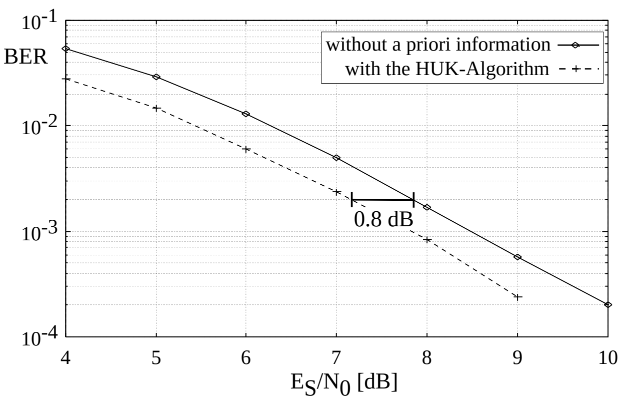


*Figure 5: Bit error rates of the 12 most significant bit in the GSM Full Rate System.*

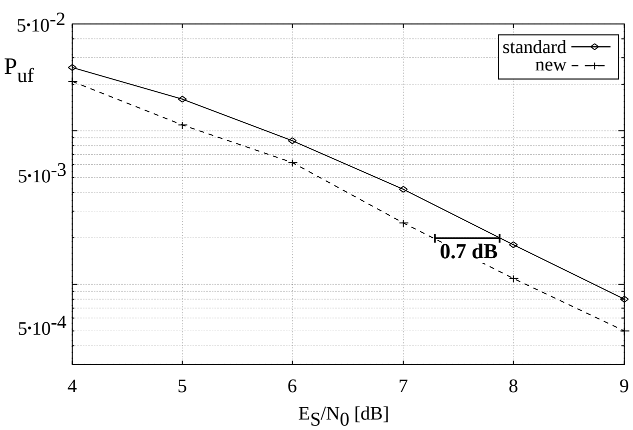


*Figure 6: Probability of undetected bad frames $P_{uf}$ with source controlled channel decoding and new error concealment (new) in comparison to GSM standard.*

Figure 6 shows the improvement of the undetected bad frames, dependent on the signal-to-noise ratio of a speech signal transmitted over a GSM traffic channel (typical urban area at a vehicle speed of 50 km/h). Here a corrupted frame is defined as a frame which has at least one error in the class 1a bits [4].

With the above measures based on the use of residual redundancy the signal-to-noise ratio can be lowered typically by 0.7 dB to obtain the same number of undetected bad frames without changing the GSM standard. Such an improvement can also be heard easily in a subjective test, because undetected frames of bad quality often lead to loud and noisy cracks on mobile phones and this can almost always be reduced using the new error concealment.

## REFERENCES

[1] Joachim Hagenauer and Peter Hoeher, " A Viterbi Algorithm with Soft-Decision Outputs and its Applications", *Proc. GLOBECOM*, Dallas, Texas, pp. 47.1.1-47.1.7, Nov. 1989.

[2] Joachim Hagenauer, "Source controlled channel decoding", *IEEE Trans. on Comm.*, vol. 43, no. 9, pp. 2449 - 2457, Sept. 1995.

[3] Wen Xu, Joachim Hagenauer and Juergen Hollmann, "Joint source-channel decoding using the residual redundancy in compressed images", *Proc. ICC'96*, Dallas, Texas, pp. 142 - 148, Jun. 1996.

[4] *GSM Recommendation* 06.10, "GSM Full Rate Speech Transcoding", 1995.

[5] *GSM Recommendation* 06.11, "Substitution and Muting of Lost Frames for Full Rate Speech Traffic Channels", 1995.

[6] *GSM Recommendation* 05.03, "Channel Coding", 1995.