LOW COMPLEXITY SOURCE CONTROLLED CHANNEL DECODING IN A GSM SYSTEM

Paul Strauch, Carlo Luschi, Magnus Sandell and Ran Yan
Lucent Technologies, Bell Laboratories,
Windmill Hill Business Park, Swindon SN5 6PP, UK
Tel.: ++44 1793 883931, Fax: ++44 1793 883816
E-Mail: pstrauch@lucent.com

ABSTRACT

In this paper we investigate source controlled channel decoding with a hard output channel decoder. Various methods have been devised in the past for source controlled channel decoding, but most of them assume that a soft output channel decoder is used. Most receivers in mobile wireless communications have a standard Viterbi channel decoder which produces only hard outputs. It is shown that a simple sliding histogram is capable of improving the speech quality significantly. The ideas and methods in this paper are applied for the full-rate and enhanced full-rate speech codecs in the GSM system.

1. INTRODUCTION

Speech encoders in general are not ideal, in the sense that the bit streams produced by a speech encoder are not independent, identically distributed (i.i.d). Furthermore, the encoded bits are not equiprobable. This means that one could further compress the speech. Due to practical reasons, i.e. complexity and real-time issues of the speech encoder and due to the non-stationary nature of speech, the full-rate (FR) and the enhanced full-rate (EFR) speech encoders in GSM have been designed to realize a good trade-off between the following criteria: speech quality, low bit rate, delay and complexity. Therefore, the speech encoders produce residual redundancy which reflects in the residual correlation between bits and in the nonuniformity of the encoded bit streams. The work in this paper focuses on the GSM FR [1] and EFR [2] speech encoders. There is also some evidence in the literature that even half-rate encoders like the 4.8 kbit/s CELP encoder produces residual redundancy [3].

Although not all the encoded bits have residual redundancy, the most important bits for the speech reconstruction typically have the highest residual redundancy. If those bits are found to be in error e.g., using a cyclic redundancy check (CRC) code, a whole speech frame will be declared as bad. This is necessary to avoid annoying cracks or clicks at the speaker. The low complexity scheme in this paper will improve the most significant bits in a speech frame and therefore the speech quality at low SNR values. At higher SNR values the channel decoder makes reliable decisions and ignores the *a priori* information obtained from the source statistics.

There are two different types of correlation between bits in the encoded speech frames. One is the interframe correlation, which accounts for the correlation between bits in successive encoded speech frames. The other one is the intraframe correlation, which represents the correlation between bits inside an encoded speech frame.

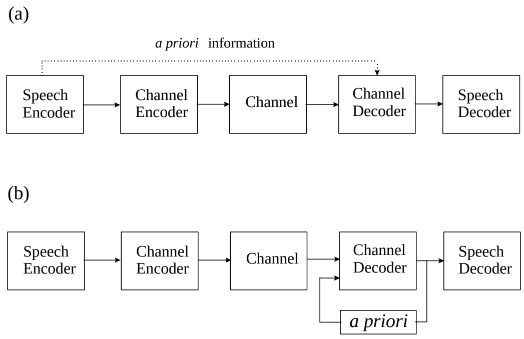

Figure 1. (a) Source controlled channel decoding in theory (b) Source controlled channel decoding in practice

In this paper, we are focusing on the improvement of single bits in the channel decoding process. This method is called source controlled channel decoding (SCCD) and was proposed by Hagenauer in 1995 [4]. The scheme is shown in Figure 1(a) and (b). In Figure 1(a) *a priori* information about the encoded bit statistics is passed to the channel decoder. In practice the scheme in Figure 1(b) is used. The statistics of the speech encoder are measured on the fly at the output of the channel decoder, the probabilities of the information bits is estimated and finally passed as *a priori* information to the channel decoder.

In Hagenauer's paper [4] the interframe correlation of the 10 most significant bits out of the 260 bits produced by the GSM FR encoder is exploited. Hagenauer devised a very simple and efficient scheme to calculate the *a priori* information for the 10 most significant bits. Hindelang *et al.* [5],[6] showed how to exploit the intraframe correlation by running the channel decoder twice. A complex multilevel approach by grouping the bits back into the speech parameters was proposed by Heinen *et al.* [7]. In [8] the speech quality was further improved by combining the *a priori* information derived from the inter- and intraframe correlation. All those aforementioned schemes, except [7], assume that a soft output channel decoder is used. The schemes influence the channel decoder to make less errors by supplying some additional *a priori* information about the source statistics to the channel decoder. We will show in this paper that by using a simple histogram method it is possible to use a hard output channel decoder.

The paper is organized as follows. In Section 2 we give a brief description of the GSM EFR speech encoder and an analysis of the EFR encoded speech data. Section 3 shows the computation of the *a priori* information and in Section 4 we present some simulation results. Conclusions are drawn in Section 5.

2. GSM ENHANCED FULL-RATE SPEECH ENCODER

2.1. Description

There are numerous descriptions of the GSM FR speech codec available in the literature [1],[9]. Therefore, we only briefly review the newly introduced EFR speech codec for GSM. The coding scheme for the GSM EFR speech encoder is an Algebraic Code Excited Linear Prediction (ACELP) technique. The ACELP coder compresses the incoming 64 kbit/s speech signal to 12.2 kbit/s. The incoming speech signal is partitioned into 20 ms segments which contain 160 samples. The ACELP coder produces 244 bits at the output of every 20 ms speech frame. These 244 bits represent the LP filter parameters (twice per 20 ms frame), the adaptive codebook delay and gain, and the fixed codebook index and gain (every 5 ms subframe) as shown in Figure 2. A theoretical analysis of the ACELP speech codec is given in [2],[9].

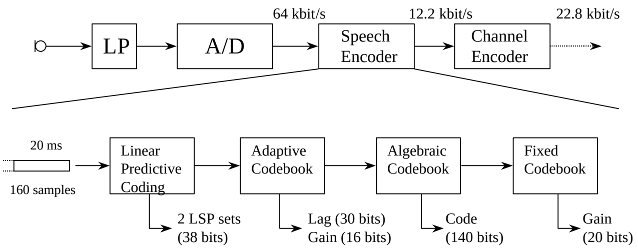

Figure 2. GSM enhanced full-rate speech encoder

As the EFR encoder only produces 12.2 kbit/s instead of 13 kbit/s like the FR encoder there is room for extra protection.

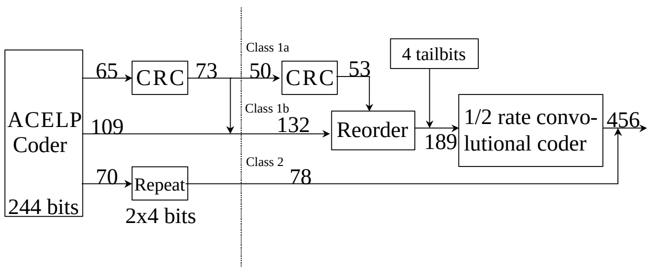

Figure 3. GSM channel coding for enhanced full-rate speech

In Figure 3 it can be seen how the spare 16 bits have been used to provide better bad frame indication (CRC (73,65)) and to protect 4 bits out of class 2. The resulting 260 bits (on the right side of the dotted line in Figure 3) are divided into 3 classes according to their subjective importance, exactly the same as for the FR encoded speech. For error protection the class 1a and 1b bits are channel encoded as shown in Figure 3.The 50 most important bits are put into class 1a which are very sensitive to errors and, therefore, get additional protection against bad transmission conditions. The next important 132 bits are put into class 1b and the 78 bits which are rather insensitive to subjective errors in the speech decoding process are grouped in class 2. The 50 bits of class 1a are protected by a weak CRC (53,50) which can detect one error and some two error patterns in class 1a. Both the CRC (53,50) and CRC (73,65) will be used to detect a bad frame at the receiver side. Together with the class 1b bits they are reordered and encoded by a 1/2 rate convolutional encoder with memory 4. The 4 tail bits are added for termination. The 78 bits in class 2 are not protected and this results in 456 bits to be transmitted at a bitrate of 22.8kbit/s.

2.2. Analysis

To verify how much residual redundancy the GSM EFR speech encoder produces we have analyzed inter- and intraframe correlation, and the probability distribution of each encoded bit. The GSM speech encoder takes 20 ms long speech frames and encodes them to 260 bits. For the analysis we used a 32.2 s long speech sequence and, therefore, the encoder produces 1610 frames with a total of 418.6 kbits.

The analysis starts with the investigation of interframe correlation. Interframe correlation means that bits with the same index position n in successive frames are correlated. Figure 4 shows two successive speech frames where the n^{th} bit of the k^{th} frame is denoted by $u_{k,n}$.

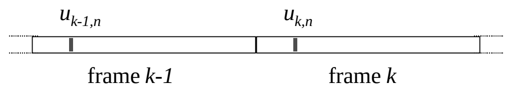

Figure 4. Interframe correlation

Therefore, we can write $u_{k,n} = f_1(u_{k-1,n})$. Figure 5 shows the normalized autocorrelation for lag = 1 for all 244 speech encoded bits.

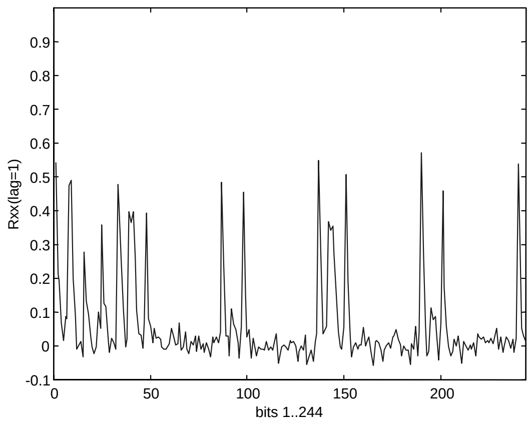

Figure 5. Normalised autocorrelation for lag = 1

The next type of correlation to investigate is the intraframe correlation. Intraframe correlation means that bits inside a speech frame are correlated. An example is shown in Figure 6. In this example $u_{k,n}$ and $u_{k,l}$ are correlated.

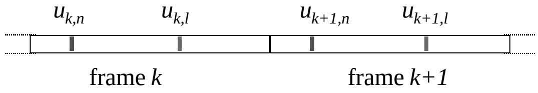

Figure 6. Intraframe correlation

Therefore, for each frame k, we can write $u_{k,n} = f_2(u_{k,l})$ and $u_{k,l} = f_3(u_{k,n})$. To verify this intraframe correlation between certain bits inside an encoded speech frame we have computed the crosscorrelations between all 244 bits in a speech frame and averaged the result over all available M speech frames:

$$\mathbf{C} = \frac{1}{M}\sum_{k=1}^{M} \mathbf{u}_k \mathbf{u}_k^T \qquad (1)$$

where $\mathbf{u}_k$ is the column vector containing the 244 bits of the k^{th} speech frame.

The normalised crosscorrelations averaged over 1610 speech frames for $|E\{u_{k,i}u_{k,j}\}| \geqq 0.4$ are shown in Table 1 .

$u_{k,i}$	$u_{k,j}$	$E\{u_{k,i}u_{k,j}\}$	$u_{k,i}$	$u_{k,j}$	$E\{u_{k,i}u_{k,j}\}$
1	9	0.54	33	35	0.41
1	2	0.52	48	98	0.59
1	8	0.50	87	88	-0.67
9	25	0.49	92	94	-0.44
33	34	0.59			

Table 1. Normalised crosscorrelations for certain bit combinations

A histogram in Figure 7 shows how often $u_{k,n}$ is -1 or +1.The relative frequencies were obtained again from 1610 speech frames. It is obvious that the EFR speech encoder

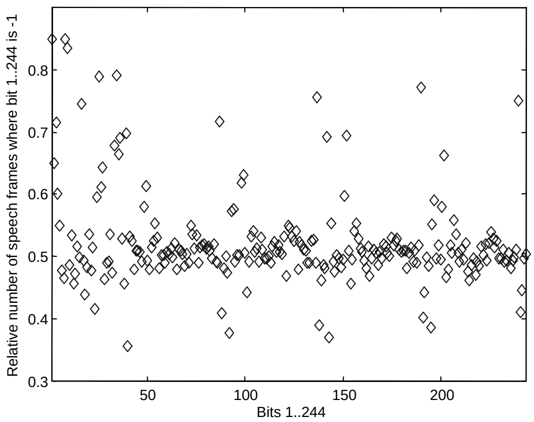

Figure 7. Relative frequencies of the EFR speech encoded 244 bits being -1 in 1610 speech frames

produces less redundancy compared to the results in [8], obtained with the FR speech encoder. However, there are bits mainly in the group of the 65 bits which are CRC (73,65) encoded which are correlated or not equally probable. The inter- and intraframe correlation could be exploited by a soft output channel decoder. As we are focussing on a hard output channel decoder in this paper we will only exploit the nonuniform distribution of the encoded bits. For comparison, we show the relative frequencies of the FR encoded speech bits in Figure 8.

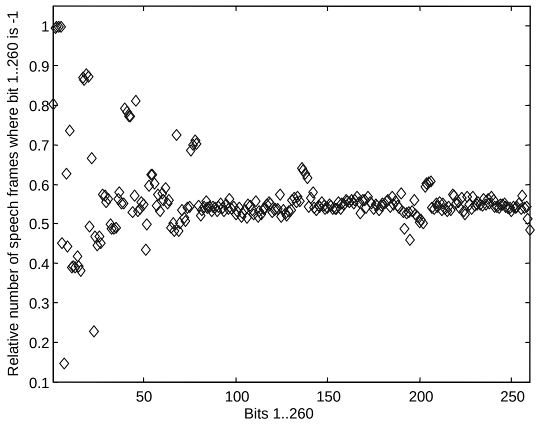

Figure 8. Relative frequencies of the FR speech encoded 260 bits being -1 in 1610 speech frames

3. CALCULATING THE *A PRIORI* INFORMATION

This section deals with the computation of the *a priori* information. The *a priori* information will be used in the channel decoder to provide additional information in the branch metric for the decoding process. Since we do not have a soft output channel decoder it is not beneficial to exploit the correlation between bits. The reason for this is given in the following example. In an interframe correlation exploiting procedure the current *a priori* information is predicted from a hard decided bit from the previous frame. If that bit is in error the *a priori* information will be wrong and will have a negative impact on the decoding process of the current frame. Therefore, we will only concentrate on the distribution of each bit. In order to exploit the *a priori* information in a Viterbi decoder the branch metric has to be modified. The branch metric $M_k^{(m)}$ of the hard output *a priori* Viterbi (APRI-VA) decoder [4] can be written as

$$M_k^{(m)} = M_{k-1}^{(m)} + \sum_{n=1}^{N} x_{k,n}^{(m)} L_{c_{k,n}} y_{k,n} + u_k^{(m)} L(u_k) \qquad (2)$$

where k denotes time, (m) is the state number, N the inverse convolutional code rate, x the coded bit, $L_{c_{k,n}} y_{k,n}$ the soft input to the decoder, u the information bit and $L(u_k)$ the *a priori* information. The *a priori* information is written as

$$L(u_k) = \log \frac{\Pr(u_k = +1)}{1 - \Pr(u_k = +1)} \qquad (3)$$

The *a priori* information will help the channel decoder in making a reliable decision in bad channel conditions. In good channel conditions the impact of the *a priori* information is rather small compared to the large confidence of the channel decoder's decision.. The probability $\Pr(u_k = +1)$

can be estimated by computing a histogram on the fly at the channel decoder. To start up the algorithm in an appropriate manner it is necessary to initialize the probabilities to be equal.

4. SIMULATION RESULTS

The following simulations use a typical urban channel model with a vehicle speed of 50 km/h. This model is named TU50 in the GSM specifications. The equalizer assumes a 5 tap channel model. A 2 minutes long English speech sequence, which contains 6000 speech frames is used to compute the BER at one specific SNR value. The BER curves for the FR speech codec are shown in Figure 9. For the complete class 1a *a priori* information (3) is estimated. The class 1a bits are improved by 0.8 dB at a BER level of 1%. The first 12 bits in class 1a are improved by 1.8 dB at the same BER level.

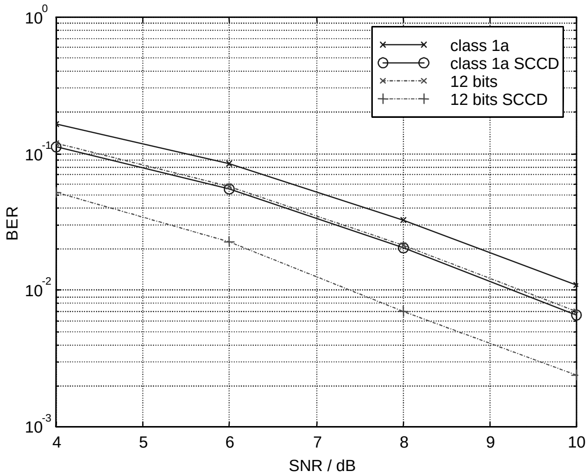

Figure 9. BER curves for the FR speech codec: Class 1a and the first 12 bits with and without SCCD

The results in Figure 9 are quite remarkable if one compares them with the results in [5]. In [5] the achieved performance increase for the 12 bits is about the same but note that a soft output (APP) channel decoder has to run twice.

Next we provide the results for the EFR speech codec. The *a priori* information is estimated for all the 65 bits which have the 8 bit CRC check. It can be seen in Figure 10 that the BER of class 1a has been improved by 0.3 dB and of class 1b only marginally.

5. CONCLUSIONS

In this paper we studied how to exploit the residual redundancy produced by the GSM FR and EFR speech encoder with a hard output channel decoder. The receiver is not part of the GSM standard and, therefore, it is possible to add features, such as source controlled channel decoding, in order to improve the performance. A brief description of the EFR speech encoder was given in Section 2. The analysis in Section 2.2 of the encoded speech data showed that certain bits have residual redundancy. The source statistics can be computed on the fly at the channel decoder output, as described in Section 3. The results in Section 4. show that

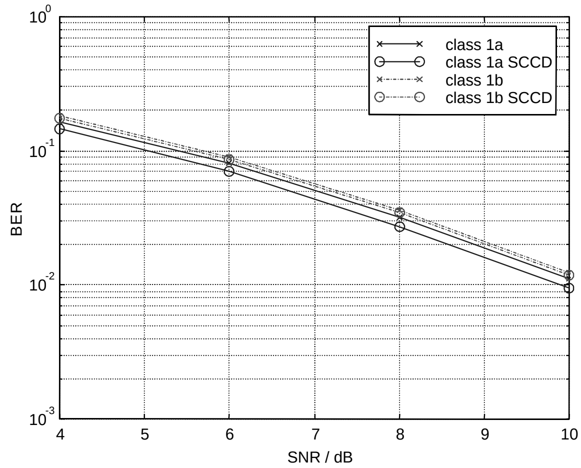

Figure 10. BER curves for the EFR speech codec: Class 1a and 1b with and without SCCD

it is possible to improve the performance by using a simple histogram method in conjunction with a APRI-VA decoder. The method is also well suited for DSP implementation as for example the quantization effects of soft output values are avoided. Listening tests revealed that the FR encoded speech quality was greatly improved. The improvement for EFR encoded speech is more subtle. Generally one can say, that there is less distortion in the EFR decoded speech with SCCD. The reasons why source controlled channel decoding is not as successful as with the FR speech codec are: 1. Less residual redundancy 2. Better bad frame indication (8 bit CRC) and 3. Better error concealment procedure.

REFERENCES

[1] GSM-06.10, *GSM Full Rate Speech Transcoding.* ETSI, 1995.

[2] GSM-06.60, *Enhanced Full Rate EFR Speech Transcoding.* ETSI, 1997.

[3] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," in *IEEE Trans. Speech and Audio Proc.*, vol. 4, pp. 325–336, Sept. 1996.

[4] J. Hagenauer, "Source Controlled Channel Decoding," in *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.

[5] A. Ruscitto and T. Hindelang, "Channel decoding using the residual intraframe correlation in GSM system," *IEE Electronics Letters*, vol. 33, pp. 1754–1755, Oct. 1997.

[6] T. Hindelang and A. Ruscitto, "Kanaldekodierung mit Apriori Wissen bei nicht binaeren Quellensymbolen," *ITG-Fachbericht 146*, pp. 163–167, Mar. 1998.

[7] S. Heinen, A. Geiler, and P. Vary, "MAP Channel Decoding by exploiting Multilevel Source A Priori Knowledge," *ITG-Fachbericht 146*, pp. 89–94, Mar. 1998.

[8] P. Strauch, C. Luschi, M. Sandell, and R. Yan, "Improved source controlled channel decoding in a GSM system," in *ISPACS'98*, (Melbourne, Australia), Nov. 1998.

[9] R. Steele, ed., *Mobile Radio Communications.* John Wiley & Sons, 1996.