Channel Decoding Using Inter- and Intra-correlation of Source Encoded Frames

Christophe Veaux, Pascal Scalart, André Gilloire
France Telecom CNET, Speech Coding Group,
2 Ave Pierre Marzin, 22307 Lannion Cedex, France
Tel: 33 2 96 81 33 53   Fax: 33 2 96 81 35 30
e-mail: christophe.veaux@cnet.francetelecom.fr

Abstract

The goal of source controlled channel decoding is to improve the performance of the channel decoder by using the residual redundancy of the source encoded data. The original approach exploits only the interframe correlation between bits and few simple scheme have been devised in order to use the intraframe correlation. In this paper, we present an efficient method to exploit the intraframe correlation during the channel decoding process. We estimate an a priori soft value of the bit being decoded from the bits already decoded along the trellis path. This estimate is then used together with the estimate obtained from interframe correlation in the Viterbi metric. We also introduce an adaptive estimator of bit correlation that takes into account the reliability of the bits at the decoder output. Our channel decoding algorithm is applied to the full-rate and enhanced full-rate speech codecs in the GSM system and results in a noticeable improvement of the bit error rate.

1   Introduction

In most source coding scheme some redundancy remains at the output of the encoder. This residual redundancy corresponds to residual correlation and non-uniformity of the discrete symbol sequence produced by the encoder. Many researchers have proposed to capitalize on this redundancy in order to combat the effect of channel noise. One approach ([1],[2]) considers that the residual redundancy may be used alone to perform error correction without any explicit channel coding. Another approach ([3],[4]) consists in combining the residual redundancy with explicit redundancy inserted by a convolutional coder in order to provide better error protection. The latter approach is particularly suited if one want to ensure a low error rate over a wide range of channel qualities. In this contribution, we focus on such procedure which is usually referred to as *source controlled channel decoding*. The scheme is shown in Figure 1, the channel decoder is fed with additional a priori information based on source residual redundancy in order to make less errors.
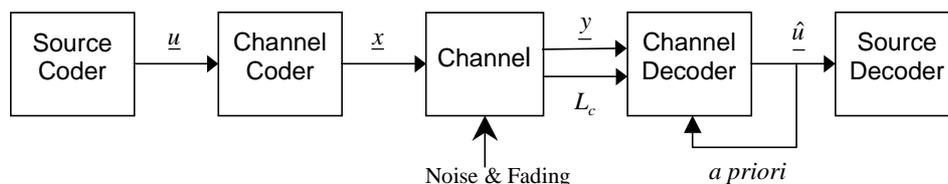

Figure 1: Source controlled channel decoding.

In his paper [4], Hagenauer exploits the correlation between individual bits of subsequent source encoded frames (*interframe* bit correlation). Yet it is obvious that the residual redundancy should be handle at the symbol level rather than at the bit level. That means we should use as a priori information the joint probability of bits representing a source coder symbol instead of single bit probabilities. Such scheme was applied in [3] to the particular case of the FS 4.8 kbit/s CELP coder. For most general case, Heinen [5] derives an approximate method to exploit joint a priori probabilities of bits groups within the channel decoding process. However, the computation of the joint probabilities requires large storage and computational burden. Furthermore, in the particular case of source coders using sub-frames, there is some dependency between source symbols within a frame and the method proposed in [5] can not easily exploit it.

A simpler but sub-optimal approach is to consider only the statistics of pairs of bits. Doing so, one can sub-optimally exploit the residual redundancy at the symbol level (i.e. non-uniformity and memory) by considering short-term *intraframe* correlation between bits. In this way, Ruscitto [6] exploits the correlation between the two most significant bits of a source parameter (non-uniformity) by running the channel decoder twice. This intraframe correlation is also used by Strauch in [7] who combines it with interframe correlation.

In this contribution, we present a much simpler and efficient method to directly exploit the intraframe correlation during the channel decoding process (running once). Our method is similar to the one introduced by Hagenauer for interframe correlation. Yet, we use an adaptive estimator of bit correlation that takes into account the reliability of the already decoded bits. Finally we combine intraframe and interframe correlation to get an a priori information about the bit being decoded by the channel decoder. This method is applied to the GSM full-rate (FR) and GSM enhanced full-rate (EFR) speech codecs.

2   Channel decoding using a priori information

Let us consider the transmission process depicted in Figure 1. The source encoder produces frames of L bits $\underline{u} = \{u_1,...,u_L\}$ with $u_i = \pm 1$ that are channel encoded with rate $1/N$ resulting in frame $\underline{x} = \{x_1,...,x_{NL}\}$. The bits x_j are transmitted over a fully interleaved fading channel with reliability L_c.

Now, we focus on the channel decoder stage. An optimum metric to perform maximum-likelihood (ML) sequence estimation is

$$M_i^{(p)} = M_{i-1}^{(p)} + \sum_{n=1}^{N} x_{i,n}^{(p)} L_{c_{i,n}} y_{i,n} \tag{1}$$

where $x_{i,n}$ denotes the n-th bit of the i-th channel symbol coding the information bit u_i and p is the trellis path considered.

2.1 *Approximate MAP channel decoding*

In ML estimation (1), no assumptions are made about the distribution of the information bits. Hagenauer proposed to add a priori knowledge in form of an a priori *soft value* of the information bit u_i: $L_{prio}(u_i) = \log \dfrac{P_{prio}(u_i = +1)}{P_{prio}(u_i = -1)}$

$$M_i^{(p)} = M_{i-1}^{(p)} + \sum_{n=1}^{N} x_{i,n}^{(p)} L_{c_{i,n}} y_{i,n} + u_i^{(p)} L_{prio}(u_i) \qquad (2)$$

This modification of the Viterbi algorithm (VA) is referred to as APRI-VA. It performs maximum a posteriori (MAP) sequence estimation assuming independence of the relevant bits u_i within the decision region of the VA.

Hagenauer estimates the *a priori* soft value $L_{prio}(u_{k,i})$ of the *i*-th bit in the *k*-th frame $\underline{u}_k$ from the *a posteriori* soft value $L_{post}(u_{k-1,i})$ of the bit with same position *i* in the frame $\underline{u}_{k-1}$:

$$L_{prio}(u_{k,i}) = f\left(L_{post}(u_{k-1,i})\right) \qquad (3)$$

The soft value $L_{post}(u_{k-1,i})$ is already available at the output of the channel decoder if a Soft Output Viterbi Algorithm (SOVA) is used.

However, as already stated, the non-uniformity of the symbol sequence produced by the source encoder implies statistical dependence within group of bits coding a same quantization index. Furthermore, if the coder uses sub-frames, the bits coding similar parameters in different sub-frames may also be correlated. Therefore, let us consider a set of statistically dependent bits $\{u_{i_1}, ..., u_{i_m}\}$ within a frame (we omit the frame index for sake of simplicity). A proper way to exploit this dependency is to formulate the Viterbi metric in terms of the joint probability $P(u_{i_1}, ..., u_{i_m})$ instead of single bit probabilities $P(u_i)$. This is clearly impractical when the spreading range of the dependent bits exceeds the constraint length of the convolutional coder. This is the case of most coders, like GSM, that reorders the speech codec bits such that the most sensitive bits are positioned at the beginning and at the end of the frame for better error protection.

For this general case, Heinen derived in [5] an approximate MAP metric where only the already decided bits $\{u_{i_1}^{(p)}, ..., u_{i_l}^{(p)}; i_l < i\}$ along the current path p are considered:

$$M_i^{(p)} = M_{i-1}^{(p)} + \sum_{n=1}^{N} \log P\left(y_{i,n} \mid x_{i,n}^{(p)}\right) + \log P_{prio}\left(u_i = u_i^{(p)} \mid u_{i_1}^{(p)}, ..., u_{i_l}^{(p)}\right) \qquad (4)$$

The a priori probability used in (4) may be estimated with respect to the previously decoded frame $\underline{u}_{k-1}$ in order to exploit the interframe correlation [8]. Nevertheless, this estimation is quite complex and is based on fixed transition probabilities that can not deal

with speech non stationary. Furthermore, this approach turns out to be impractical when the set of bits $\{u_{i_1},...,u_{i_m}\}$ is extended to different source coder symbols in order to exploit the sub-frame correlation.

Therefore, we propose to restrict ourselves to pairs of correlated bits $\{u_{k,i};u_{k,j}\}$ within the frame $\underline{u}_k$. These bits may correspond to the same parameter as well as to correlated parameters. Doing so, we are also able to use a very simple procedure for adaptively estimate the correlation between $u_{k,i}$ and $u_{k,j}$ in order to track the speech non stationary. This procedure will be presented in section 3.

2.2 APRI-SOVA exploiting intraframe correlation

We exploit the *intraframe* correlation during the channel decoding process by using a an approach similar to Hagenauer for the *interframe* correlation.

Assuming $j < i$, we use the metric (2) with an *a priori* soft value $L_{prio}(u_{k,i})$ estimated from the current estimate of the *a posteriori* soft value $L_{post}(u_{k,j})$:

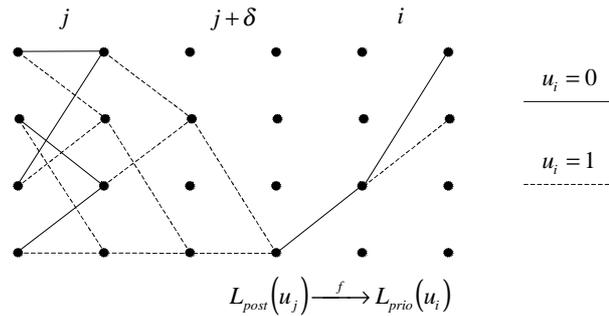$$L_{prio}(u_{k,i}) = f\left(L_{post}(u_{k,j})\right) \tag{5}$$

Figure 2: Estimation of the a priori soft value $L_{prio}(u_{k,i})$.

This scheme is illustrated in Figure 2. The soft value $L_{post}(u_{k,j})$ is available if we have $(i-j) > \delta$, δ being large enough so that all the different path at time j have merged with high probability. If not, $L_{post}(u_{k,j})$ will be overestimated. That means the estimation of $L_{prio}(u_{k,i})$ tends to rely on the hard bit value of $u_{k,j}$ when j tends toward i. However, it may be noticed that in such case, our metric turns out to be a restriction of equation (4) to the pair $\{u_{k,i};u_{k,j}\}$.

A problem arises when $(i-j)$ is less or equal to the constraint length $(v+1)$. Nevertheless, we may notice that $L_{prio}(u_{k,i})$ can be equivalently added to the metric $M_{i+v}^{(p)}$ rather than to $M_i^{(p)}$ thus its estimation may be delayed.

It should be obvious that we need to store the soft value of the bit $u_{k,j}^{(p)}$ already decoded along the path p in order to compute our metric. Actually, if we use a SOVA implemented in the register exchange mode [9], the path memory contains this information.

Finally, in order to exploit the maximum knowledge available on the source, we can combine the a priori information from intraframe and interframe correlation using the same adding relation as in [7]:

$$L_{prio}\left(u_{k,i}\right) = f\left(L_{post}\left(u_{k,j}\right)\right) + f\left(L_{post}\left(u_{k-1,i}\right)\right) \tag{6}$$

3 Estimation of the a priori information

We now deal with the estimation of $L_{prio}\left(u_l\right)$ from $L_{post}\left(u_q\right)$ where (q,l) denotes either the frame index (interframe correlation) or the bit position within a frame (intraframe correlation). Similarly to [4], we model the correlation between bits u_q and u_l with a change bit c_{ql} :

$$u_l = u_q \oplus c_{ql} \tag{7}$$

with *known* a priori soft value $L_{prio}\left(c_{ql}\right) = \log \dfrac{1 - p_{ql}^{prio}}{p_{ql}^{prio}} \tag{8}$

where $p_{ql}^{prio} = P\left(u_l = -1 \middle| u_q = +1\right) = P\left(u_l = +1 \middle| u_q = -1\right)$ is the *a priori* probability of change. This model which corresponds to symmetric transition probabilities is more relevant for interframe correlation than for intraframe correlation. However, it has the main advantage that a very simple relation between the soft values can be derived from equation (7):

$$\begin{aligned}
L_{prio}\left(u_l\right) &= f\left(L_{post}\left(u_q\right)\right) \\
&\approx \mathrm{sign}\left(L_{prio}\left(c_{ql}\right)\right).\mathrm{sign}\left(L_{post}\left(u_q\right)\right).\min\left(\left|L_{prio}\left(c_{ql}\right)\right|, \left|L_{post}\left(u_q\right)\right|\right)
\end{aligned} \tag{9}$$

Hagenauer has proposed an ad hoc method named HUK that *adaptively* estimates the a priori knowledge $L_{prio}\left(c_{ql}\right)$. In this algorithm $L_{prio}\left(c_{ql}\right)$ is increased or depreciated depending on the hard values (± 1) of the already decoded bits u_q and u_l. Nevertheless, it seems more convenient to take into account the reliability of the already decoded bits when estimating $L_{prio}\left(c_{ql}\right)$. Therefore, we present a new method to adaptively estimate $L_{prio}\left(c_{ql}\right)$ from the soft values observed at the output of the channel decoder:

Let us consider the estimates $\hat{u}_q$ and $\hat{u}_l$, the associated change bit is:

$$c_{ql} = \hat{u}_q \oplus \hat{u}_l \tag{10}$$

According to (9), its soft value $L_{post}(c_{ql})$ is given by:

$$L_{post}(c_{ql}) \approx \text{sign}(L_{post}(u_l)).\text{sign}(L_{post}(u_q)).\min(|L_{post}(u_l)|,|L_{post}(u_q)|)$$

Therefore, we can express the a posteriori probability of change:

$$p_{ql}^{post} = 1/\left(1 + e^{L_{post}(c_{ql})}\right) \tag{11}$$

and obtain a new estimate of $L_{prio}(c_{ql})$ by updating the a priori probability p_{ql}^{prio} according to:

$$p_{ql}^{prio} = \lambda.p_{ql}^{prio} + (1-\lambda).p_{ql}^{post} \tag{12}$$

with forgetting factor λ.

4 Redundancy in GSM speech codecs

Before applying our method to the GSM FR and EFR, we analyze the residual interframe and intraframe correlation produced by these coders. For this analysis, we compute the average interframe and intraframe correlation over 1500 frames of silence-removed speech. In the following, we show theses correlations only for the first 80 most sensitive bits of the speech encoded frames.

4.1 *Analysis of the GSM Full-Rate*

The full-rate is a Regular Pulse Excited-Long Term Prediction (RPE-LTP) coder [10]. It produces 260 bits for every 20 ms speech frame. These 260 bits represent one set of 8 LP filter coefficients converted to log-area ratio (LAR), 4 sets of LTP filter coefficients and RPE excitation (updated every 5 ms subframe). All these bits are *scalar quantized*.

All the 260 bits not have the same subjective importance. The 50 most sensitive bits (*Class 1a*) must not be used if corrupted and in that case the whole speech frame is declared as bad. However, it is shown in Figures 3 and 4 that the most subjectively important bits are highly correlated. Therefore one can expect to greatly improve the speech quality by exploiting this redundancy in the channel decoding process. The positions of the most significant bits (MSB) for some of those parameters are given as follows:

LAR1 (1,6)	Xmax1 (2,17)	LTP lag1 (13,24)	LTP gain1 (53)
LAR2 (7,10)	Xmax2 (3,18)	LTP lag2 (14,25)	LTP gain2 (54)
LAR3 (8,11)	Xmax3 (4,19)	LTP lag3 (15,26)	LTP gain1 (55)
LAR4 (12,45)	Xmax4 (5,20)	LTP lag4 (16,27)	LTP gain1 (56)
LAR5 (22,51)			
LAR6 (23,52)			

We may notice that for some parameters, like the block amplitudes Xmax, the intraframe correlation is much higher between the corresponding bits of successive sub-frames than between the 2 MSB of each sub-frame parameter.
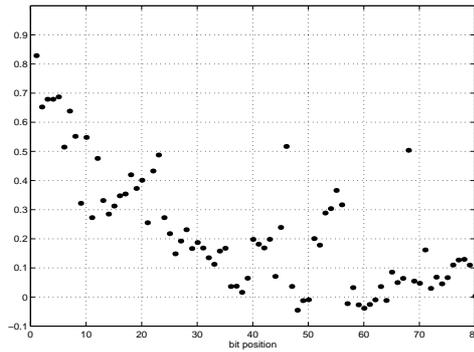
Figure 3: Normalized inter-correlation between bits of successive frames for the GSM FR.
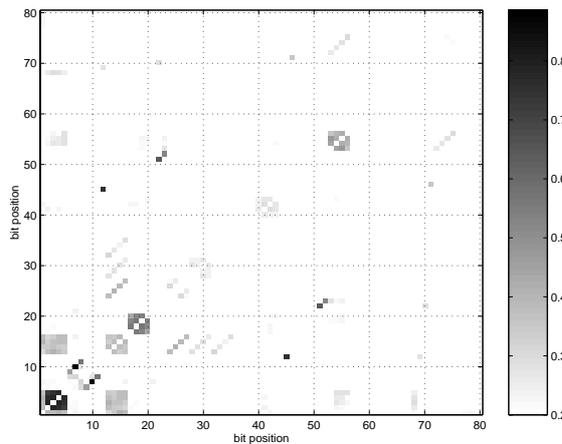

Figure 4: Normalized cross-correlation (absolute value) between bits of a same frame for the GSM FR.

4.2 *Analysis of the GSM Enhanced Full-Rate*

The coding scheme of the GSM EFR is an Algebraic Code Excited Linear Prediction (ACELP) technique [11]. It produces 244 bits for every 20 ms speech frame. The 10 LP filter coefficients (twice per frame) are converted to line spectrum pairs (LSP) and *jointly quantized* after moving average (MA) prediction. The pitch delay, the gains of adaptive and fixed codebook are *scalar quantized* (with moving average prediction applied to the fixed codebook gain). It seems obvious that the EFR produces less redundancy than the GSM FR because the intraframe correlation is partly removed by the joint quantization and the residual of MA prediction is rather decorrelated. Nevertheless, it is shown in Figures 5 and 6 that some significant redundancy remains. The correspondence between the most correlated bits and the coder parameters are given as follows (we denote LPC1 the 1st quantization index of LSP):

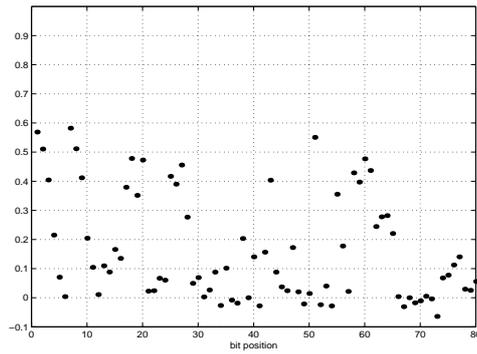|			
LPC1 (25,26,51)	LTP gain1 (17)	FCB gain1 (18,62)	LTP lag1 (1,2,3,4,5,6)
	LTP gain2 (19)	FCB gain2 (20,63)	LTP lag2 (13,14)
	LTP gain3 (58)	FCB gain3 (60,64)	LTP lag3 (7,8,9,10,11,12)
	LTP gain4 (59)	FCB gain4 (61,65)	LTP lag4 (15,16)

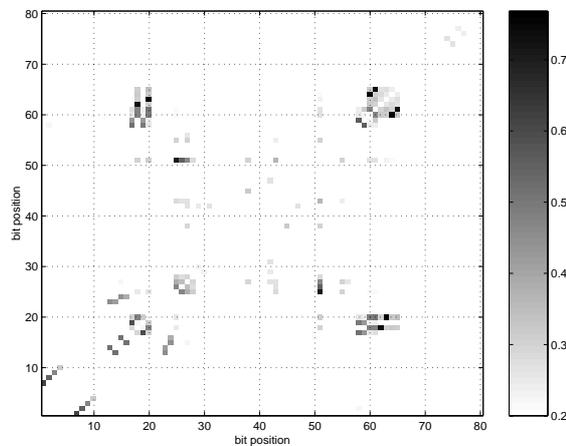*Figure 5: Normalized inter-correlation between bits of successive frames for the EFR.*

Figure 6: Normalized cross-correlation (absolute value) between bits of a same frame for the EFR.

5 Simulation results

In implementing our algorithm, we consider only the bits whose normalized interframe or intraframe correlation (Figure 3 to 6) exceeds a threshold $\lambda = 0.2$. In the case of intraframe correlation, for each selected bit, the associated bit chosen to estimate its a priori soft value must satisfy an *order constraint* (he must precede the "predicted" bit after the reordering of the encoded frame). We give in following the main bit combinations used for the GSM FR and EFR respectively (see section 4):

FR: 10–7, 2–3, 4-3, 5-3, 12-45, 22-51, 8-11, 18-19, 20-19, 19-17, 52-23, 55-53

EFR: 20-63, 60-64, 65-61, 18-62, 51-25, 7-1, 19-17, 58-59, 26-51, 2-8, 14-13, 16-15

We simulate the channel transmission with static error patterns corresponding to a typical urban channel model with a vehicle speed of 50 km/h and carrier to interference ratio C/I from 2 dB to 8 dB. A speech sequence of 7500 frames was used to compute the bit error

rate (BER) at each C/I value. The improvement of the BER *within class 1a* is represented in Figure 7 and 8 for the GSM FR and EFR respectively. We compare the BER curves obtained without source controlled channel decoding to the results of our algorithm using the interframe correlation alone, the intraframe correlation alone or their combination according to (6).
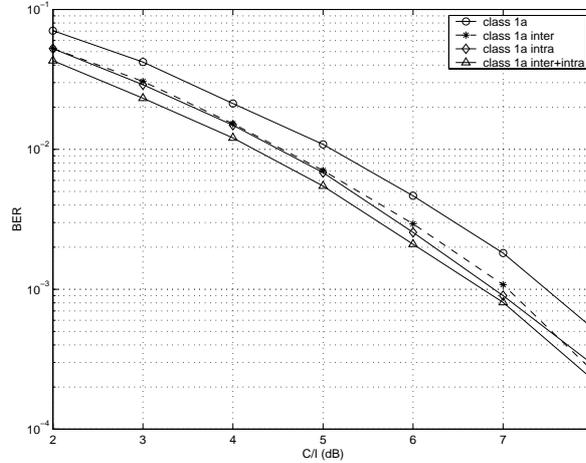

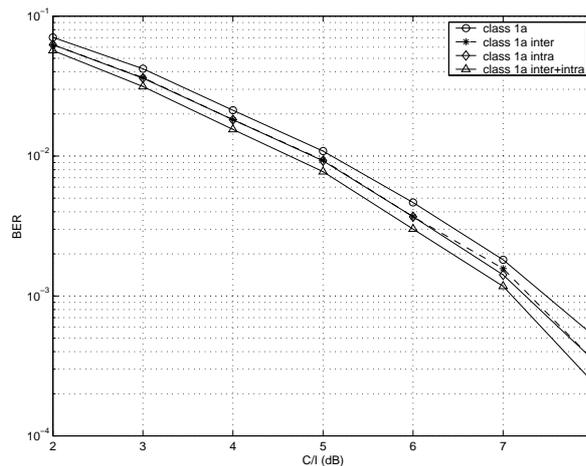*Figure 7: Bit error rates for class 1a in GSM FR.*

Figure 8: Bit error rates for class 1a in GSM EFR.

The performance obtained with either the interframe or the intraframe correlation alone are very similar whereas one should expect better results in the case of intraframe correlation as in [6]. This is a consequence of the constraint imposed when selecting the bit combination. However, this constraint allows us to run the channel decoding process only once whereas it is run twice in [6].

When combining inter- and intraframe correlation, it is possible to achieve a total gain of ~0.9 dB for the GSM FR and ~0.5 dB for the EFR at 10^{-2} BER. In the case of the GSM FR, the improvement compared to the inter- or intraframe correlation alone is rather small. This may be because the a priori information obtained from either the interframe or the intraframe correlation is rather the same. This may be due also to a poor combination of these two information by the "ad hoc" relation (6).

6 Conclusion

The analysis of the encoded speech frames for the GSM FR and EFR have shown that there is still some redundancy remaining and that the correlation between parameters of different sub-frames takes a significant part in it. In this contribution, we have presented a new method to exploit the intraframe correlation together with the interframe correlation during the channel decoding process. Such approach is sub-optimal but allows to take easily into account the correlation between different source coder symbol. Furthermore, our estimator of bit correlation is able to track the speech non stationary and is less sensitive to error propagation since it relies on the bit reliability. The results obtained with our channel decoding algorithm show the benefit of using interframe and intraframe correlation although a better way to combine them may be found.

References

[1] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source/channel coders", *IEEE Trans. Commun.*, vol. 39, pp. 838-846, June 1991.

[2] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels -Application to combined source-channel coding", *IEEE Trans. Inform. Theory*, vol. 40, pp. 186-193, 1994.

[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] S. Heinen, A. Geiler, and P. Vary, "MAP Channel Decoding by exploiting Multilevel Source A Priori Knowledge", *ITG-Fachbericht 146*, pp. 163-167, Mars 1998.

[6] 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.

[7] P. Strauch, C. Luschi, M. Sandell, and R. Yan, "Improved source controlled channel decoding in a GSM system", in *Proc. ISPACS'99*, (Melbourne, Australia), Nov. 1998.

[8] T. Fingscheidt and O. Scheufen, "Robust GSM Speech Decoding Using the Channel Decoder's Soft Output", in *Proc. Eurospeech'97*, pp. 1315-1318, Sept. 1997.

[9] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", in *Proc. GLOBECOM'89*, Dallas, Nov. 1989, pp. 47.1.1-47.1.7.

[10] *GSM Recommendation 06.10*, "Full Rate speech transcoding".

[11] *GSM Recommendation 06.11*, "Enhanced Full Rate speech transcoding".