Journal of Southeast University (English Edition) June 2002 Vol.18 No.2 ISSN 1003—7985

Joint Source-Channel Decoding of EVRC Speech Encoder Using Residual Redundancy*

Ahmed J. Jameel** You Xiaohu Gao Xiqi

(National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

Abstract: The enhanced variable rate codec (EVRC) is a standard for the "Speech Service Option 3 for Wideband Spread Spectrum Digital System," which has been employed in both IS-95 cellular systems and ANSI J-STC-008 PCS (personal communications systems). This paper concentrates on channel decoders that exploit the residual redundancy inherent in the enhanced variable rate codec bitstream. This residual redundancy is quantified by modeling the parameters as first order Markov chains and computing the entropy rate based on the relative frequencies of transitions. Moreover, this residual redundancy can be exploited by an appropriately "tuned" channel decoder to provide substantial coding gain when compared with the decoders that do not exploit it. Channel coding schemes include convolutional codes, and iteratively decoded parallel concatenated convolutional "turbo" codes.

Key words: residual redundancy, convolutional codes, turbo codes

The main function of the source encoder is to transform the input speech signal into a more compact form. Ideally all of the redundant bits are removed in the source compression phase. The channel encoder then adds a certain amount of controlled redundancy to the input signal. This redundancy, under the form of an error-control code, is used to protect the information against the effects of the channel noise. Traditionally, source and channel coding have been treated as separate functions, resulting in what is known as a tandem source-channel coding system. This is justified by Shannon's separation principle[1], which states that the source and channel coding functions can be designed independently from each other without a loss in the optimality of the system. However, Shannon's theory is an asymptotic result that permits unlimited delay and complexity. Recently, systems with jointly designed source and channel coding operations have been shown to outperform tandem systems under practical limitations such as finite block lengths. For reasons of delay and complexity and because of highly nonstationary sources, many source-coding schemes still contain redundancy and output bits are of different significance and error sensitivity. One also notices that some bits are highly correlated, at least at times. Under these circumstances, source and channel coding as well as decoding cannot be treated separately. In fact, Shannon mentioned this possibility already in his paper[1]: "However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise." Consequently, if some redundancy*** or bit correlation is left by the source encoder this should be utilized jointly by the channel and source decoder[3].

In this work, we consider joint source-channel decoding for the reliable communication of EVRC parameters of Rate 1 encoded speech over very noisy binary phase-shift keying (BPSK) modulated additive white Gaussian noise (AWGN) and Rayleigh fading channels.

This paper is organized as follows. In section 1, the residual redundancy of the EVRC system is given. Exploiting residual redundancy with convolutional encoder and turbo decoder is presented in sections 2 and 3, respectively. In section 4, the channel models are briefly described. In section 5, the experimental results are introduced, and we conclude in section 6.

1 Analysis of EVRC Residual Redundancy

One frame of EVRC consists of 10 line spectral pair (LSP) parameters; which model the signal's shortterm

Received 2002-02-20.

* The project supported by the National Natural Science Foundation of China(69725001).

** Born in 1963, male, graduate, Iraq.

*** The term "redundancy" usually has three connotations: ① the source has memory but is uniformly distributed; ② the source distribution is nonuniform but it has no memory; and ③ the source has memory and a nonuniform distribution[2]. In this paper, "redundancy" implies cases ① or ③.

spectrum, is quantized with a weighted split vector LSP quantizer. EVRC coding also makes use of adaptive and fixed codebooks, which simulate the human speech's voiced and unvoiced excitations, respectively. The adaptive codebook is represented by 7-bit pitch delay, 5-bit delta delay, and three 3-bit adaptive codebook gains per frame. Similarly, the fixed codebook is represented by three 35-bit shapes and three 5-bit gain parameters. The bit allocation for each set of parameters is shown in Tab.1[4].

Tab.1 Bit allocations of Rate 1

Parameter	Number of bits
LPCFLAG	1
LSP	6 + 6 + 9 + 7 = 28
Pitch delay	7
Delta delay	5
ACB gain	3 × 3 = 9
FCB shape	3 × 35 = 105
FCB gain	3 × 5 = 15
(Reserved)	1
Total	171

Our goal is to quantify the residual redundancy in the CELP parameters that are present in every frame: the LSP's, the pitch gains, the pitch delays, the adaptive codebook gains, and fixed codebook gains; to do this we will estimate the entropy rate of these parameters. All the parameters are of different bit lengths. For consistency we chose to quantify the redundancy in the 3 most significant bits (MSB's) of each EVRC parameter.

For each set of parameters, let the random process, $\{U_{i,j}\}$, represent the three most significant bits of the ith (quantized) EVRC parameter in frame j, and let $\boldsymbol{U}_j = [U_{1,j}, U_{2,j}, \cdots, U_{l,j}]$, where l denotes the number of parameters per frame. We assume that the process, $\{\boldsymbol{U}_j\}_{j=1}^{\infty}$, is block stationary, then the per frame entropy rate of this process is given by

$$H(U) = \lim_{j\to\infty} H(\boldsymbol{U}_j \mid \boldsymbol{U}_{j-1}, \boldsymbol{U}_{j-2}, \cdots, \boldsymbol{U}_1) \qquad (1)$$

where $H(\boldsymbol{U}_j \mid \boldsymbol{U}_{j-1}, \boldsymbol{U}_{j-2}, \cdots, \boldsymbol{U}_1) = -\sum_{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j} P(\boldsymbol{U}_j = \boldsymbol{u}_j, \boldsymbol{U}_{j-1} = \boldsymbol{u}_{j-1}, \cdots, \boldsymbol{U}_1 = \boldsymbol{u}_1) \log_2 [P(\boldsymbol{U}_j = \boldsymbol{u}_j \mid \boldsymbol{U}_{j-1} = \boldsymbol{u}_{j-1}, \boldsymbol{U}_{j-2} = \boldsymbol{u}_{j-2}, \cdots, \boldsymbol{U}_1 = \boldsymbol{u}_1)]$.

Here, $H(U)$ represents the minimum number of bits per frame required to describe $\{\boldsymbol{U}_j\}_{j=1,2,\cdots}$ without incurring distortion; alternately, if the process is encoded at a rate R, then the quantity $\rho = R - H(\boldsymbol{U})$ is the residual redundancy incurred by the encoding.

To estimate the residual redundancy of the EVRC parameters, it is necessary to provide a probabilistic model for their generation. We assume that the parameters can be modeled by a stationary first-order Markov chain and computing the entropy rate based on the relative frequencies of transition[5]. We assume that the parameters in different frames are independent:

$$P(\boldsymbol{U}_j = \boldsymbol{u}_j \mid \boldsymbol{U}_{j-1} = \boldsymbol{u}_{j-1}, \boldsymbol{U}_{j-2} = \boldsymbol{u}_{j-2}, \cdots, \boldsymbol{U}_1 = \boldsymbol{u}_1) = P(\boldsymbol{U}_j = \boldsymbol{u}_j) \qquad (2)$$

and within a frame as

$$P(U_{i,j} = u_{i,j} \mid U_{i-1,j} = u_{i-1,j}, U_{i-2,j} = u_{i-2,j}, \cdots, U_{1,j} = u_{1,j}) = P_A^{(i)}(u_{i,j} \mid u_{i-1,j}) \qquad (3)$$

for $i = 2, 3, \cdots, l$ and $j = 1, 2, \cdots$. Note that for $i = 1$, Eq.(3) becomes $P_A^{(1)}(u_{1,j})$. This assumption is supported by a number of observations.

- In practice, detected errors are often masked by repeating or interpolating parameter values from adjacent frames.
- A second feature is the ordered nature of the LSP parameters within one EVRC frame ($\mathrm{LSP}_1 < \mathrm{LSP}_2 < \cdots < \mathrm{LSP}_{10}$), which suggests intraframe dependency between the LSP's.
- There is a little change in the observed entropy rate of the process if it is computed assuming a higher-degree Markov process, e.g., a second-or third-order Markov chain[6].

Ultimately, of course, the most important judge of the "fit" of this model is the performance of a decoder designed to fit this model, which we shall see in section 5 is quite good.

In order to estimate the residual redundancy of the various parameters, a large training sequence of speech database was used; for every frame of speech, an EVRC algorithm is performed to get the parameters. The relative frequency of transitions between the three high-order bits values of each parameter was compiled to extract Markov transition probabilities. The entropy of the resulting Markov chains was computed to arrive at an estimate of the redundancy in each parameter in each frame. Let $H^* = \sum_{i=1}^{l} H(U_{i,j})$ be the process entropy rate (in bit/frame) if the parameters were independent (H^* is independent of j since (1) is independent of j). Note that we can write $\rho_{\mathrm{T}} = \rho_{\mathrm{D}} + \rho_{\mathrm{M}}$ where $\rho_{\mathrm{D}} = 36 - H^*$ denotes the frame redundancy due to the non-uniform distribution of the parameters and $\rho_{\mathrm{M}} = H^* - H(U)$ denotes the frame redundancy due to the memory between the parameters.

The results are compiled in Tab.2 in which we provide the values of ρ_D, ρ_M and ρ_T for each individual parameter as well as for the entire frame.

Tab.2 Residual redundancy bit/frame

Parameter	ρ_D	ρ_M	ρ_T
LSP	0.035	0.103	0.138
Pitch delay	0.564	0.698	1.262
Delta delay	0.528	0.629	1.157
ACB gain	0.184	0.289	0.473
FCB gain	1.356	1.483	2.839
Total	2.667	3.202	5.869

It is clear from Tab.2 that the amounts of 16% of the bits in each frame are redundant bits. Now, the residual redundancy of the EVRC parameters will be exploited with two channel coding schemes.

2 Exploiting Residual Redundancy with Convolutional Codes

The residual redundancy present in EVRC parameters can be exploited in a convolutional decoder by adjusting the Viterbi algorithm decoding metric to incorporate the Markov transitional probabilities. Maximum a posteriori decoding is optimal in the sense of minimizing the sequence error probability[5]. The MAP decision rule is to choose the code sequence, $\{\hat{\boldsymbol{c}}_k\}$ or $\{\hat{\boldsymbol{x}}_k\}$, which maximizes

$$P(\boldsymbol{y}^K|\boldsymbol{x}^K)P(\boldsymbol{x}^K)$$

where $\boldsymbol{y}^K = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_K)$ is the corrupted received sequence, $\boldsymbol{x}^K = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_K)$ is the sequence of the transmitted, modulated code symbols, and K is the sequence length.

For the AWGN and fully interleaved Rayleigh fading channels with noise variance $N_0/2$, the above metric reduces to choosing $\{\hat{\boldsymbol{x}}_k\}$ that minimizes

$$\sum_{k=1}^{K}\|\boldsymbol{y}_k - \boldsymbol{a}_k\hat{\boldsymbol{x}}_k\|^2 - N_0\ln P(\hat{\boldsymbol{x}}^K) = \sum_{k=1}^{K}[\|\boldsymbol{y}_k - \boldsymbol{a}_k\hat{\boldsymbol{x}}_k\|^2 - N_0\ln P(\hat{\boldsymbol{x}}_k|\hat{\boldsymbol{x}}_{k-1}, \hat{\boldsymbol{x}}_{k-2}, \cdots)] \quad (5)$$

where $\{\boldsymbol{a}_k\}$ is the sequence of Rayleigh fading coefficients (for AWGN, $\boldsymbol{a}_k$ is the all-one vector for all k), assumed to be available at the decoder. The prior distribution $P(\hat{\boldsymbol{x}}^K)$ is estimated using the Markov model in conjunction with a large training sequence. Hence (5) becomes

$$\sum_{k=1}^{K}\|\boldsymbol{y}_k - \boldsymbol{a}_k\hat{\boldsymbol{x}}_k\|^2 - N_0\ln P(\boldsymbol{u}_k|\boldsymbol{u}_{k-1}, \boldsymbol{u}_{k-2}, \cdots) \quad (6)$$

Note that (6) can easily be implemented via a modified version of the Viterbi algorithm. We will consider two soft decision-decoding schemes based on (6).

ML: Maximum likelihood decoding which chooses the code sequence $\{\hat{\boldsymbol{x}}_k\}$ that minimizes

$$\sum_{k=0}^{K}\|\boldsymbol{y}_k - \boldsymbol{a}_k\hat{\boldsymbol{x}}_k\|^2 \quad (7)$$

Here, no redundancy in the EVRC parameters is exploited in the decoding metric.

MAP: Maximum a posteriori decoding exploiting the residual redundancy, by choosing $\{\hat{\boldsymbol{x}}_k\}$ to minimize

$$\sum_{k=0}^{K}\|\boldsymbol{y}_k - \boldsymbol{a}_k\hat{\boldsymbol{x}}_k\|^2 - N_0\ln P_A^{([k\bmod l])}(\boldsymbol{u}_k|\boldsymbol{u}_{k-1}) \quad (8)$$

where $[k\bmod l]$ refers to the unique integer between 1 and l.

3 Exploiting Residual Redundancy with Turbo Codes

The turbo coder that we used is parallel concatenated convolutional coder (PCCC). The iterative turbo decoding process was used at the receiver five times.

We conducted simulations using plain turbo coding and turbo coding enhanced with EVRC's residual redundancy information. A conventional turbo decoder produces for each data bit b a "log-likelihood" ratio $L(b)$:

$$L(b) = \log(P(b=1|\boldsymbol{r})/P(b=0|\boldsymbol{r})) \quad (10)$$

where $\boldsymbol{r}$ is the received sequence. Given a random binary 3-tuple $\boldsymbol{b} = [b_0 \ \ b_1 \ \ b_2]$, we associate the following reliability metric with $\boldsymbol{b}$:

$$L(\boldsymbol{b}) = \sum_{i=0}^{2}\log[P(b_i=1|\boldsymbol{r})/P(b_i=0|\boldsymbol{r})] \quad (11)$$

Now, we shall assume a first order Markov relationship among a sequence of binary triples, i.e., assume the statistical model we have been using for the high-order bits in the EVRC gains, pitch, and LSP. Let $\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \cdots\}$ be the sequence of binary triples, then the metric that we employ in the simplified approach is

$$M(\boldsymbol{b}_i) = L(\boldsymbol{b}_i) + \log\left[\frac{P(\boldsymbol{b}_i|\boldsymbol{b}_{i-1})}{P(000|\boldsymbol{b}_{i-1})}\right] \quad (12)$$

where $L(\boldsymbol{b}_i)$ is the contribution computed from the turbo decoder's log-likelihood values and the second term accounts for the *a priori* transition probabilities computed from the training sequence.

4 Channel Models

The channel models considered in this work are the additive white Gaussian noise (AWGN) channel and the fully interleaved Rayleigh fading channel, both are used with BPSK modulation. More specifically, we assume the jth received signal x_j according to

$$y_j = a_j x_j + n_j \tag{13}$$

where $x_j \in \{+\sqrt{E_s}, -\sqrt{E_s}\}$ and n_j is a zero-mean Gaussian random variable with variance $N_0/2$. The distribution of the fading coefficient a_i depends on the channel assumption:

- For a purely AWGN channel, we assume $a_i = 1$;
- For the fully interleaved Rayleigh fading channel, we assume that a_i has a Rayleigh distribution with $E[a_j^2] = 1$. Moreover, to accommodate the "fully interleaved" description, we assume that a_i and a_j are independent for $i \neq j$.

5 Simulation Results

For convolutional codes, the simulations of decoding schemes are based on the Viterbi algorithm with the 32-state rate 3/4[7], with free distance of 5 and a generator matrix

$$G(D) = \begin{pmatrix} 6 & 2 & 2 & 6 \\ 1 & 6 & 0 & 7 \\ 0 & 2 & 5 & 5 \end{pmatrix}$$

The turbo encoder is implemented with the 16-state rate 3/4[8], with free distance of 9 and a generator matrix

$$G(D) = \begin{pmatrix} 1 & 0 & 0 & \dfrac{f(D)}{g(D)} \\ 0 & 1 & 0 & \dfrac{f(D)}{g(D)} \\ 0 & 0 & 1 & \dfrac{f(D)}{g(D)} \end{pmatrix}$$

where $f(D) = 1 + D^2 + D^3 + D^4$, and $g(D) = 1 + D + D^4$, and the maximum number of soft decoding the residual redundancy is exploited with the non-interleaved data. It can be seen from Fig.1 and Fig.2 that for the BPSK-modulated system and at a decoding bit error rate (BER) of 0.01, MAP decoding gains (over ML decoding) are as high as 0.9 dB for the Gaussian channel and 1.1 dB for the Rayleigh fading channel. While the enhanced turbo decoder gains (over the conventional one) are as high as 0.5 dB for the Gaussian channel and 0.4 dB for the Rayleigh channel.

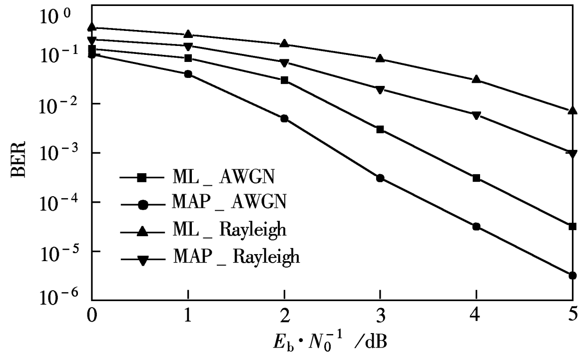

Fig.1 Performance of convolutional codes in AWGN and Rayleigh fading channels

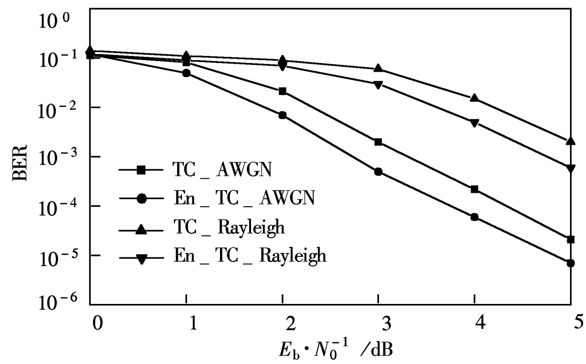

Fig.2 Performance of turbo codes in AWGN and Rayleigh fading channels

6 Conclusion

This paper quantifies the residual redundancy in EVRC encoder bitstream, more specifically; it shows that substantial residual redundancy is present in the quantized pitch and gain parameters and less (but still noticeable) redundancy is present in the output of the LSP vector quantizer. Techniques for exploiting this residual redundancy assuming various channel-coding strategies were developed and assessed; the channel codes included convolutional codes and turbo codes. Noticeable improvements were obtained with each of these methods, especially at low signal-to-noise ratio and especially when the performance criterion is closely linked to a redundancy-rich parameter.

References

[1] Shannon C E. A mathematical theory of communication [J]. *Bell System Technical Journal*, 1948, **27**: 379 – 423.

[2] Phamdo N, Farvardin N. Optimal detection of discrete markov sources over discrete memoryless channels-applications to combined source-channel coding[J]. *IEEE Transactions on Information Theory*, 1994, **40**(1): 186 – 193.

[3] Hagenauer J. Source-controlled channel decoding[J]. *IEEE Transactions on Communications*, 1995, **43**(9): 2449 – 2457.

[4] Enhanced Variable Rate Codec. *Speech Service Option* 3 *for Wideband Spread Spectrum Digital Systems* [S]. IS-127, July 19, 1996.

[5] Alajaji F I, Phamdo N, Fuja T E. Channel codes that exploit the residual redundancy in CELP-Encoded speech[J]. *IEEE Transactions on Speech and Audio Processing*, 1996, **4**(5): 325 – 336.

[6] Ahmed J. Jameel. *Joint source-channel decoding for mobile communications*[D]. Nanjing: Department of Radio Engineering, Southeast University: in preparation.

[7] Shu Lin, Costello Daniel J. *Error control coding*: *fundamentals and applications* [M]. Prentice-Hall, 1983.

[8] Heegard C, Wicker S B. *Turbo coding*[M]. Kluwer Academic Publisher, 1998.

利用残留冗余量的联合信源信道解码在 EVRC 语音解码器中的应用

阿哈麦德　尤肖虎　高西奇

（东南大学移动通信国家重点实验室，南京 210096）

摘　要　宽带扩频系统的语音服务采用增强型的可变比率编码器，该标准已在 IS－95 和 J－STC－008 个人移动通信系统得到应用．本文致力于利用增强可变比率编码器中内在残留冗余信息的信道解码器．由于残留冗余信息可以用一阶马尔夫模型表示，同时相关频率的变化可以用熵率来表示，从而，信道解码器可利用这种残留冗余量．仿真结果表明，和没有利用这种信息量的解码器相比，由于编码增益，系统性能有明显改善，其中，信道编码采用了卷积码、Turbo 码两种方式．

关键词　残留冗余，卷积码，Turbo 码

中图分类号　TN929.533