LETTER

An Iterative Joint Source-Channel (De-)Coding and (De-)Modulation Algorithm for G.729EV in Ultrashort Wave Communication*

Tan PENG†, Xiangming XU†, Huijuan CUI†, Kun TANG†, Nonmembers, and Wei MIAO†, Student Member

SUMMARY Improving the overall performance of reliable speech communication in ultrashort wave radios over very noisy channels is of great importance and practical use. An iterative joint source-channel (de-)coding and (de-)modulation (JSCCM) algorithm is proposed for ITU-T Rec.G.729EV by both exploiting the residual redundancy and passing soft information throughout the receiver while introducing a systematic global iteration process. Being fully compatible with existing transmitter structure, the proposed algorithm does not introduce additional bandwidth expansion and transmission delay. Simulations show substantial error correcting performance and synthesized speech quality improvement over conventional separate designed systems in delay and bandwidth constraint channels by using the ISCCM algorithm.

key words: joint source-channel coding and modulation, G.729EV, speech coding, soft decoding

1. Introduction

Ultrashort wave radio communication is one of the main modes of communication for navigation, long distance communication and civil use. It serves as one of the very few available means of communication, especially in emergency conditions such as the communication support in the Sichuan earthquake recently in China and the great Hanshin earthquake in Japan when the infrastructure of other communication methods such as PSTN, GSM and WCDMA were severely damaged. Its unique properties of strong survivability, rapid deployment and mobility make it indispensable. With rapid development of modern ultrashort wave communication systems, the available bit rate can allow the transmission of high-quality wideband (50Hz-7kHz) speech and audio signals such as the new ITU-T standard scalable wideband coder G.729EV [1].

Conventional ultrashort wave speech communication systems are designed based on Shannon’s classical separation theorem [3] with the premise of no constraint on code length, delay or complexity. However limited bandwidth, delay and the restriction of resources such as complexity and power in speech communication terminals, reveal the limitation of the separate design of source and channel. Moreover the separation theorem also requires the source encoder to completely remove all the redundancy and the channel code to give zero error probability above a certain channel condition. In practice, the residual redundancy is inevitably inherent in the output of the source encoder because of imperfection of practical encoders. And some redundancy is left intentionally to improve robustness. Meanwhile perfect protection can not always be provided by the channel encoder because of the limited bandwidth and the effect of noise in wireless channels. These residual redundancy, mutual information and structural correlation can be fully exploited by jointly designing the source coder, channel coder and modulation in a systematic way which seeks for global optimality comparing with the conventional separate design which only aims at local optimality. For a fixed degree of delay and/or bandwidth expansion, the joint source-channel (de-)coding and (de-)modulation (JSCCM) system can achieve better reconstructed speech quality and overall error correcting performance in ultrashort wave communications over very noisy channels. It also means that for a fixed degree of performance gain, the JSCCM system will require less bandwidth (bitrate reduction) and/or algorithm delay than the conventional separate designed systems.

In [4]–[6] speech coding and channel coding are jointly designed with unequal error protection or error concealment. In [7], [8] joint trellis coded quantization and continuous phase modulation (CPM) is proposed. However, the source coder, channel coder and modulation have not been jointly designed. In this letter, an iterative joint source-channel (de-) coding and (de-) modulation algorithm is proposed for speech communication using G.729EV which fully exploits the mutual information and correlation between the demodulator, channel decoder and source decoder. We also provide experimental results for the conventional separate designed system and the proposed JSCCM scheme under various channel conditions.

2. Redundancy within ITU-T Rec.G.729EV

Using the conventional ITU-T 8 kb/s speech coding standard G.729 CS-ACELP [2] as a core layer, G.729EV gener-
ates enhancement layers on top of it where each additional layer successively improves the audio quality. The full rate frame-based G.729EV consists of twelve codec layers: the core layer, the narrowband enhancement layer, the TDBWE (Time Domain Bandwidth Extension) layer and eight TDAC (Time Domain Aliasing Cancellation) layers. Starting from the TDBWE layer, the G.729EV can produce wideband signals. Within the constraint of complexity, we focus only on the most important layers, from the core Layer to the TDBWE layer. Each speech parameter between successive frames is modeled by a first order Markov chain to characterize redundancies in G.729EV with good approximation [10]. A large training speech sequence (over 100,000 frames) from the TIMIT [9] speech database is applied to estimate the Markov probabilities of each parameter set and compute its entropy rate \( H_r \). Suppose \( M \) is the number of bits assigned to the quantized parameter, the parameter with index \( i \) in the \( k^{th} \) frame is \( x_i^k \). So the redundancy rate is

\[
\rho_r = \frac{(M - H_r)}{M}
\]

\[
= \left( M + \sum_{x_i^k \rightarrow x_j^{k-1}} p(x_i^k | x_j^{k-1}) \log p(x_i^k | x_j^{k-1}) \right) / M \quad (1)
\]

The results are listed in Table 1. We can see that of the considered 109 parameter bits, more than 7.05% are redundant which can be further utilized.

### Table 1: G.729EV parameters and their redundancy.

<table>
<thead>
<tr>
<th>G.729EV parameter</th>
<th>( M )</th>
<th>( H_r )</th>
<th>( \rho_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSP MSVQ 1st stage</td>
<td>7</td>
<td>6.01</td>
<td>14.14%</td>
</tr>
<tr>
<td>LSP MSVQ 2nd stage</td>
<td>10</td>
<td>9.15</td>
<td>8.50%</td>
</tr>
<tr>
<td>Pitch Delay</td>
<td>13</td>
<td>12.58</td>
<td>3.23%</td>
</tr>
<tr>
<td>Gain codebook</td>
<td>4</td>
<td>3.97</td>
<td>2.32%</td>
</tr>
<tr>
<td>CELP 12k extra codebook</td>
<td>13</td>
<td>12.68</td>
<td>2.46%</td>
</tr>
<tr>
<td>CELP 12k extra codebook sign</td>
<td>4</td>
<td>3.98</td>
<td>2.46%</td>
</tr>
<tr>
<td>CELP 12k extra gain index</td>
<td>5</td>
<td>4.26</td>
<td>14.80%</td>
</tr>
<tr>
<td>TDBWE time envelope mean</td>
<td>5</td>
<td>4.03</td>
<td>19.40%</td>
</tr>
<tr>
<td>TDBWE time envelope</td>
<td>14</td>
<td>13.46</td>
<td>3.86%</td>
</tr>
<tr>
<td>TDBWE freq envelope</td>
<td>14</td>
<td>12.66</td>
<td>9.57%</td>
</tr>
</tbody>
</table>

3. **System Model Overview and Analysis**

In this section, we give a comprehensive overview of the JSCCM system for G.729EV. The system model is described in Fig. 1.

#### 3.1 Transmitter Structure

The full rate G.729EV encoder extracts speech parameters each frame (of 20 ms duration) which are then quantized and mapped to bit combinations consisting of 640 bits at the rate of 32 kb/s. The parameter with index \( i \) is assigned with a bit combination as

\[
x_i^k = (x_i^k(0), x_i^k(1), \ldots, x_i^k(M - 1)), x_i^k \in (1, 0)
\]

Speech parameter bit combinations are then interleaved and passed to the channel encoder. Since speech communication is highly sensitive to delay and complexity, irregular (640, 1280) LDPC code (the outer code) is used to encode the bit combinations of each G.729EV frame which enjoys good error correcting performance without introducing additional frame delay. The output sequence of LPDC encoder is then coded by (2, 1, 3) convolutional code (the inner code) and modulated by CPM which is the natural choice of modulation in ultrashort wave communication when constant envelope of the transmitted signal is required [10]. The phase transition trellis of CPM and the state transition trellis of convolutional code are further incorporated into one joint trellis graph which forms the trellis-coded CPM (TCCPM).

The transmitted signal can be described as

\[
s(t) = \sqrt{\frac{2E}{T}} \cos \left[ 2\pi f_c t + \phi(t; I) \right]
\]

where \( T \) is the pulse duration, \( f_c \) is the carrier frequency and \( \phi(t; I) \) is the carrier phase.

#### 3.2 Receiver Structure

Different from conventional separate designed systems using hard-bit decoding, the JSCCM algorithm passes soft information (bit log-likelihood ratio, LLR) of each speech parameter through the demodulator, channel decoder and source decoder while introducing a systematic global iteration process between them to further enhance the reliability of each decoded bit. The utmost useful information of the received signal is preserved which makes it apparently superior. There are two levels at which iterations are carried out. At the first level, local iterations are performed within LDPC decoder by exchanging reliability information between check and variable nodes while giving soft outputs. At the second level, the global iteration is formed by exchanging soft information of each bit among the demodulator, channel decoder and source decoder. The global iteration process is indicated in Fig. 1 by the dotted lines showing two-way soft information flow throughout the receiver. After transmitted through the AWGN channel, the filtered received signal is
\[ r(t) = s(t) + n(t) \]  

where 
\[ n(t) = n_s(t) \cos 2\pi f_s t - n_s(t) \sin 2\pi f_s t \]

Since the phase transition trellis of CPM and state transition trellis of convolutional code are incorporated into one joint trellis graph, coherent demodulation of CPM is performed with trellis based BCJR algorithm [11] which forms a soft-in soft-out (SISO) structure. After TCCPM demodulation, the demodulated soft information is passed to the LDPC decoder with sum-product algorithm (SPA) which is an efficient iterative decoding algorithm based on belief propagation. The decoded speech parameter bit combination can be described as
\[ \hat{x}_k^i = (\hat{x}_k^i(0), \hat{x}_k^i(1), \ldots, \hat{x}_k^i(M - 1)) \]  

with the corresponding LLR of every bit as
\[ L_{\text{ldpc}}^{i} = (L_{\text{ldpc}}^{i}(0), L_{\text{ldpc}}^{i}(1), \ldots, L_{\text{ldpc}}^{i}(M - 1)) \]

Suppose \( L_{\text{prior}}^{i} \), initialized to be zero, is the priori information of trellis based demodulation and channel decoding. The extrinsic information
\[ L_{\text{source-prior}}^{i} = L_{\text{ldpc}}^{i} - L_{\text{prior}}^{i} \]

is passed to the source decoder and further converted to probability values.

\[
\begin{cases}
L_{\text{source-prior}}^{i}(x_k^i(m) = 0) = \frac{1}{1 + e^{L_{\text{source-prior}}^{i}(x_k^i(m) = 1)}} \\
L_{\text{source-prior}}^{i}(x_k^i(m) = 1) = \frac{e^{L_{\text{source-prior}}^{i}(x_k^i(m) = 1)}}{1 + e^{L_{\text{source-prior}}^{i}(x_k^i(m) = 1)}}
\end{cases}
\]

The a posteriori probability of each bit in the received speech parameter is calculated by jointly utilizing:
1. Received bit combinations in all the previous frames.
2. Decoded bit combinations of each global iteration in the current frame.
3. Inter-frame transition probabilities of parameters.

Since each quantized parameter is modeled by a first-order markov chain, the a posteriori probability in the \( k^{th} \) frame is calculated as
\[
P(x_k^i | \hat{x}_k^i, \overline{X}_{k-1}) = P(\hat{x}_k^i | x_k^i) \cdot P(x_k^i | \overline{X}_{k-1})
\]

\[
P(\hat{x}_k^i | x_k^i) = \prod_{t=0}^{M-1} P_{\text{source-prior}}(\hat{x}_k^i | \overline{X}_{k-1})
\]

where \( P_{\text{source-prior}}(\hat{x}_k^i | \overline{X}_{k-1}) \) is the decoded bit combinations from time index 0 to \( k - 1 \). The a posteriori LLR value of each received bit is
\[
L_{\text{source}}(x_k^i(m) = 0) = \ln \frac{P(x_k^i(m) = 0 | \hat{x}_k^i, \overline{X}_{k-1})}{P(x_k^i(m) = 1 | \hat{x}_k^i, \overline{X}_{k-1})}
\]

\[
= \ln \left\{ \left[ P(\hat{x}_k^i | x_k^i(m) = 0) \cdot \sum_{j=0}^{2^m-1} P(x_k^i(m) = 0 | x_k^j \cdot P(x_k^j | \overline{X}_{k-1}) \right] \right\}
\]

\[
\left\{ P(\hat{x}_k^i | x_k^i(m) = 1) \cdot \sum_{j=0}^{2^m-1} P(x_k^i(m) = 1 | x_k^j \cdot P(x_k^j | \overline{X}_{k-1}) \right\}
\]

Therefore a better estimation of the probability of each bit is derived. The extrinsic information from the source
\[
L_{\text{TCCPM-prior}} = L_{\text{source}} - L_{\text{source-prior}}
\]

is passed back to the demodulator, thereby forming a global iteration among the demodulator, channel decoder and source decoder which substantially enhances the overall error correcting performance and the synthesized speech quality without extra bandwidth expansion and transmission delay. The iterative process can be repeated until convergence, or a maximum number of global iterations has been reached.

4. Simulation Results

4.1 Objective Comparisons of Error Correcting Performance

Extensive simulations were carried out to evaluate the performance of the proposed JSCCM algorithm compared with the conventional separate designed system. We apply the irregular (640, 1280) LDPC code (the inner code) on the full rate G.729EV. The generator of convolutional encoder (the inner code) is \([1001; 1101]\). The modulation index of CPM is \( h = 1/4 \) with pulse shape of length \( L = 1 \) raised cosine (1RC). The TCCPM structure results a total of 64 transition states. All the results are based on AWGN channels. The objective performance measurement used for speech parameters is the signal-to-noise ratio (SNR), defined as

\[
\text{SNR} = 10 \log_{10} \frac{\sum_{k=1}^{N} E((x_k^i)^2)}{\sum_{k=1}^{N} E((\hat{x}_k^i - x_k^i)^2)} \] (dB)
while forming systematic global iterations to gradually improve the error correcting performance. In order to fully preserve the useful information in the received signal, the JSCCM converts the soft information into hard-decoded bits only at the final stage where the synthesized speech has to be rendered. There is $0.1 \sim 1$ dB improvement in terms of parameter SNR with only one global iteration. Performance is gradually improved as the global iteration number increases. There is $0.1 \sim 0.81$ dB further improvement in the 5th iteration. Especially when the channel $E_b/N_0$ is between $-2 \sim -1$ dB which is the usual working range of ultrashort wave communication systems in harsh wireless channel conditions, there is $1.32$ dB improvement in terms of parameter SNR comparing with the system without global iteration. Note that the performance gain is negligible after 5th iteration. It is due to the fact that as the number of iteration increases, the correlation between the prior probability and the calculated a posterior probability of each decoded speech parameter bit increases accordingly which results less improvement gain to the performance.

4.2 Subjective Comparisons of Reconstructed Speech Quality

The quality of reconstructed speech is evaluated by both the ITU-T Rec PESQ [12] measurement which provides an overall MOS (Mean Opinion Score) identifying the subjective synthesized speech quality and additional standard subjective listening tests. Sample speeches, consisting of over 30,000 speech frames, are all standard English utterances from the TIMIT [9] speech database. There is no speaker occurring both in the training and the test database.

Table 2 present the comparison of average speech PESQ scores for the proposed JSCCM system with different number of global iterations respectively. There are $0.01 \sim 0.12$ improvement in terms of PESQ scores after only one global iteration is introduced. When the number of global iteration reaches five, $0.02 \sim 0.24$ further improvement is achieved. Under the usual working range (channel $E_b/N_0$ between $-2 \sim -1$ dB) of ultrashort wave communication systems, there is $0.092 \sim 0.317$ improvement of PESQ scores which corresponds to apparently audible improvement between sample speeches compared with the separate designed system without global iteration. Additional subjective listening tests aiming at further evaluating the synthesized speech quality using standard headphones also certify and support the results of PESQ scores.

5. Conclusion

We have studied both the separate designed scheme and JSCCM schemes with different iteration numbers used to protect speech parameters of G.729EV. The JSCCM algorithm does not strictly emphasize bit exactness recovery, which we also can not always achieve. Instead, it focuses on gradually enhancing the reliability of each parameter bit in a systematically iterative way. Both objective SNR measures and subjective listening tests show that the JSCCM algorithm can achieve better overall system performance under bandwidth and transmission delay constraint channels. The JSCCM algorithm is applicable to any SISO structured receiver.

Acknowledgment

The authors would like to thank the editor and the anonymous reviewers for providing valuable comments which greatly improved the quality of this paper.
References


