IEICE TRANS. INF. & SYST., VOL.E91–D, NO.12 DECEMBER 2008

LETTER

Realtime Joint Speech Coding and Transmission Algorithm for High Packet Loss Rate Wireless Channels*

Tan PENG[†a)], **Huijuan CUI**[†], **Kun TANG**[†], *Nonmembers*, *and* **Wei MIAO**[†], *Student Member*

SUMMARY In digital speech communication over noisy high packet loss rate wireless channels, improving the overall performance of the realtime speech coding and transmission system is of great importance. A novel joint speech coding and transmission algorithm is proposed by fully exploiting the correlation between speech coding, channel coding and the transmission process. The proposed algorithm requires no algorithm delay and less bandwidth expansion while greatly enhancing the error correcting performance and the reconstructed speech quality compared with conventional algorithms. Simulations show that the residual error rate is reduced by 84.36% and the MOS (Mean Opinion Score) is improved over 38.86%.

key words: *speech coding, joint source and channel coding, coded transmission*

1. Introduction

Reconstructed speech quality is seriously degraded when speech parameters are transmitted in noisy high packet loss rate wireless channels, especially in shortwave communications [1] where the packet loss rate (PLR, also called jamming rate) usually reaches 55%. While in the remaining packets, the channel bit error rate (BER) reaches 15%. Meanwhile inter-frame predictive coding, vector quantization, and super-frame techniques are widely used in speech coding algorithms, especially in low-bit-rate speech coding. Each speech parameter carries a great amount of information which results a lack of robustness. Once channel error or packet loss occurs, the reconstructed speech quality will be degraded not only in one frame but in consecutive frames.

A considerable amount of work [2]–[6] has been dedicated to protect speech parameters in harsh channel conditions. Multiple frames are first buffered then protected by RS (Reed Solomon) codes or RS, Turbo, LDPC concatenated codes with deep interleaving to counter the severe channel interference. However, these kinds of algorithms require a large amount of additional delay to buffer speech frames due to the relatively long code length of channel codes while the decoding complexity of Turbo or LDPC codes are not affordable in handset speech communication terminals. Neither of these can meet the requirements of real-time speech communication. Even without these restrictions, reconstructed speeches rendered by conventional algorithms are hardly intelligible under high PLR and BER wireless channels.

In this letter we proposed a novel speech coding and transmission algorithm based on joint source channel coding for real-time high quality speech communication over high PLR and BER wireless channels. It requires no additional algorithm delay and less bandwidth expansion while outperforming conventional algorithms. The proposed algorithm has already been implemented and used in real-time speech communication systems over shortwave channels.

2. Low-Bit-Rate Speech Coding and Channel Analysis

Speech signals are usually divided into frames to perform speech coding due to its short-term stability. Let's take the 600 bit/s SELP (Sinusoidal Excited Linear Prediction) [7] speech coder for instance. SELP is an excellent low-bit-rate speech coding algorithm with independent intellectual property rights. The reconstructed speech quality is better than the American federal standard MELPe [8] under both error free and error prone channels. Speech signal is divided into 25 ms each frame and three consecutive frames are jointly quantized as a super-frame (75 ms duration). In SELP, quantized speech parameters are LSF (Line Spectral Frequency), pitch, gain, U/V and fourier magnitude resulting totally 45 bits each super-frame. Note that different speech parameters bear different importance to the synthesized speech quality. For very low-bit-rate speech coding algorithms such as SELP and MELPe, the first and second stage quantization vector of multi stage vector quantized LSF parameters and U/V are the most important parameters under massive standard speech database listening evaluations. So providing more intensive protection for them will result better speech quality under the same bandwidth expansion.

The probabilistic Gilbert model [9] is considered throughout this letter to better characterize the packet switch wireless channel. Figure 1 illustrates the Gilbert model. The transition probability that a packet is lost, given the previous packet was not lost (Good→Bad) is denoted as α, and vice-versa for β if the previous packet is lost (Bad→Good). The conditional probability that the current packet is lost given the previous packet was lost is $1-\beta$. The steady state probability, referred to as the unconditional loss probability, for each state is:

Manuscript received June 16, 2008.
Manuscript revised August 17, 2008.

[†]The authors are with the State Key Laboratory on Microwave and Digital Communications, Tsinghua National Laboratory for Information Science and Technology, and Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China.

*This work is supported by the National Natural Science Foundation of China (NSFC), Grant No. 60572081.

a) E-mail: pengtanthu@gmail.com

DOI: 10.1093/ietisy/e91–d.12.2892

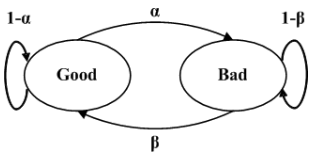

Fig. 1 The Gilbert model.

$$g(t) = \begin{cases} P_{good} = \dfrac{\beta}{\alpha+\beta}, \\ P_{loss} = \dfrac{\alpha}{\alpha+\beta}. \end{cases} \quad (1)$$

In the "Bad" state, the transmission packets which suffer significant signal degradations due to hostile jam or the very small channel gain, are considered to be lost. Therefore, they are zero-padded (corresponding to burst errors). While in the remaining packets (corresponding to "Good" state), the channel BER is set at an arbitrary level (corresponding to random errors).

3. The Proposed Realtime Joint Speech Coding and Transmission Algorithm

3.1 Coding and Transmission Protection at the Transmitter

The frame-based SELP encoder produces speech parameters which are quantized and mapped to bit combinations to form a frame and further equally divided into transmission groups. Each group is encoded by BCH code which enjoys good error correcting performance and strict algebraic structure while easy for constructing and encoding. Since realtime speech communication is highly sensitive to delay and complexity, the channel coding length for low-bit-rate speech communication is restricted within one frame (less than 60 bits). After intensive simulations and comparisons, the error correcting performance of BCH codes outperforms RS and RCPC codes. When Berlekamp iterative decoding algorithm [9] is introduced, additional indication can be provided if the number of errors exceeds the error correcting capability of BCH code. This property can be further utilized at the receiver side for majority based judgment recovery of speech parameter packets.

Taking into consideration of the harsh channel conditions, the information of speech parameters should be widely spread and multi-described as much as possible in different transmission groups while maintaining statistical correlation for better error correcting performance and synthesized speech quality. Multi superposition barrel shifting (MSBS) algorithm is designed to solve this problem which consists of the following steps:

1) Suppose the transmission block length is M, i and j are the block number and group number respectively. Set the block starting pointer to the first group.

2) Allocate M groups of data out of the BCH coded bit stream to form a transmission block. Relocate the starting pointer to the next group.

3) If the number of the current transmission block is larger than $N - M + 2$, the last $\{K_i \mid i > N\}$ groups are barrel-shifted to the $\{K_i \ mod \ (N) \mid i > N\}$ groups in the bit stream correspondingly in order to keep $M - 1$ consecutive groups super-posited with the previous block.

4) Check the starting pointer. If it points at the last group of data in the bit stream, then assemble all the transmission blocks from 1 to N for transmission. Otherwise return to Step 2).

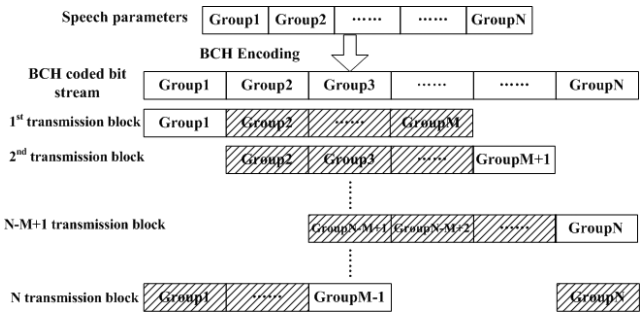

Fig. 2 Multi superposition barrel shifting algorithm procedure.

Figure 2 shows the procedure of the algorithm. The superposition groups between the current and previous block are indicated by dashed lines. Since each transmission group is covered once in M transmission blocks, there will be exactly M different multi-description copies in a barrel shifting way at the receiver for countering the adverse effect of packet losses and bit errors in the wireless channel after transmission.

3.2 Decoding and Error Correcting at the Receiver

After receiving all the parameters transmission packets at the receiver side, N transmission blocks are extracted and rearranged by performing the inverse MSBS algorithm. Each transmission group in the transmission block is decoded using the corresponding BCH code with Berlekamp iterative decoding algorithm. If the current transmission group is within its correcting capability then the decoded bits are stored into the buffer matrix $\{D_{i,j} \mid 1 \le i \le M, 1 \le j \le N\}$ and the corresponding group status $\{F_{i,j} \mid 1 \le i \le M, 1 \le j \le N\}$ is set. Otherwise the group status $F_{i,j}$ is cleared which indicates directly discarding the corrupted transmission group.

$$F_{i,j} = \begin{cases} 1, within\ capability. \\ 0, exceed\ capability. \end{cases} 1 \le i \le M, 1 \le j \le N \quad (2)$$

After transmitted through the high PLR and BER wireless channels, some of the transmission groups may be corrupted. In order to counter such adverse effect, all the transmission copies of each speech parameter is performed with majority judgment recovery (MJR) algorithm described as

$$T_j = D_{i,j} \quad (3)$$

st

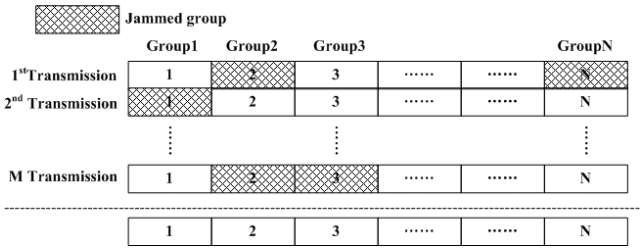

Fig. 3 Receive buffer arrangement.

$$\exists 1 \le n, l \le N, \quad s.t. \quad D_{n,j} = D_{i,j} \tag{4}$$

$$F_{i,j} \neq 0 \tag{5}$$

where T_j is the final recovered transmission group. Figure 3 shows the receive buffer arrangement. The upper bound of the probability $P_{recover}$ that all data groups can be correctly recovered after majority judgment recovery is given by:

$$P_{Max} = 1 - (P_{loss})^M - C_M^l(1 - P_{loss})(P_{loss})^{M-1} \tag{6}$$

taking into Eq. (1) we get:

$$\begin{aligned} P_{recover} &\le P_{Max} \\ &= 1 - \left(\frac{\alpha}{\alpha+\beta}\right)^{M-1} \left\{\left(\frac{\beta}{\alpha+\beta}\right) \cdot M + \frac{\alpha}{\alpha+\beta}\right\} \end{aligned} \tag{7}$$

It has been proved that the proposed algorithm can correctly recover all speech groups with relatively small bandwidth expansion. As the transmission block number M increases, the probability of correctly recovery approaches 1 exponentially. Restricted by the actual bandwidth expansion requirements in practical wireless speech communications, M should not be increased without limitation. So there is a tradeoff between speech quality and bandwidth consumption.

3.3 Unequal Protection of Speech Parameters and Error Concealment

In order to further enhance the reconstructed speech quality, unequal protection and error concealment of speech parameters are both introduced in the proposed algorithm. As mentioned before, the first and second stage LSF quantization vectors and the U/V speech parameter were found to be the most important speech parameters in terms of reconstructed speech quality. Therefore, protection of these parameters is appropriate since the contribution of other parameters is relatively trivial. Note that this algorithm is also applicable for other speech coders such as the American federal standard MELPe which has LSF and U/V parameters.

Three additional bits are added each super-frame to parity check the first and second stage LSF quantization vectors and the U/V speech parameter. The frame-based speech encoder produces speech parameters which are quantized and mapped to bit combinations to form a frame. Suppose the frame index is k. The parameter with index r is assigned with a bit combination consisting of N_b bits as

$$x_k^r = (x_k^r(0), x_k^r(1), \ldots, x_k^r(N_b - 1)), x_k^r \in (1, 0) \tag{8}$$

When transmitted through the channel, possible error will be introduced to the speech parameter bit combination. Assume the received bit combination at the receiver side is

$$\hat{x}_k^r = (\hat{x}_k^r(0), \hat{x}_k^r(1), \ldots, \hat{x}_k^r(N_b - 1)), \hat{x}_k^r \in (1, 0) \tag{9}$$

Odd numbers of transmission bit errors within the protected speech parameters will be detected. Since more than three-bits error is a very rare case, one bit error is considered throughout the error concealment process. Each corrupted speech parameter will be recovered using one of the two kinds of error concealment criterions depending on its unique characteristic and quantization algorithm at the transmitter side. The arbitrary parameter error concealment criterion should reflect the impact of parameter errors on the subjective speech quality. For vector quantized LSF parameters, error concealment based on minimum mean square (MMSE) criterion is appropriate which is described as

$$\begin{aligned} v_k^{MMSE} &= \sum_{r=0}^{N_b-1} v_k^{(r)} \cdot P(x_k^r \mid \hat{x}_k, \overline{X}_{k-1}) \\ &= \sum_{r=0}^{N_b-1} v_k^{(r)} \cdot \frac{P(\hat{x}_k \mid x_k^r) \cdot P(x_k^r \mid x_{k-1}^{r'})}{\sum_{r=0}^{N_b-1} P(\hat{x}_k \mid x_k^r) \cdot P(x_k^r \mid x_{k-1}^{r'})} \end{aligned} \tag{10}$$

where $\overline{X}_{k-1}$is the decoded bit combinations from time index 0 to $k-1$, $P(x_k^r \mid x_{k-1}^{r'})$is the inter-frame transition probability calculated offline by using the first markov chain which is trained by intensive standard speech database. $v_k^{(r)}$ is the corresponding code vector.

In contrast, the maximum a posteriori (MAP) criterion should be applied when the speech parameter is uniformly quantized at the encoder side such as the U/V parameter which is formulated as

$$\begin{aligned} v_k^{MAP} = \{&v_k^m \mid P(x_k^m \mid \hat{x}_k, x_{k-1}^{r'}) \\ &\ge P(x_k^t \mid \hat{x}_k, x_{k-1}^{r'}), t \neq m\} \end{aligned} \tag{11}$$

It should be noted that although the error correcting performance will not be enhanced since the transmission errors will only be detected but not corrected, better reconstructed speech quality can be achieved when combined with error concealment.

4. Simulation Results

Probabilistic Gilbert model is used throughout the simulations to characterize the noisy wireless packet switch channels where the packet loss rate (corresponding to a steady state probability of $\frac{\alpha}{\alpha+\beta}$) ranges from 5% to 55%. While in the remaining packets (corresponding to a steady probability of $\frac{\beta}{\alpha+\beta}$), the bit error rate ranges from 5% to 15% respectively. The 600 b/s SELP is used as the speech coder without losing generality.

4.1 Error Correcting Performance Comparison

Extensive simulations were carried out to evaluate the error correcting performance of the proposed algorithm compared with conventional algorithms which are widely used in outdoor, shortwave radio and long distance wireless communications.

Scheme (1) "RS codes and deep interleaving". RS codes with long code length and deep interleaving methods are applied. 10 super-frames are buffered and encoded using $(31, 1, 31)$ RS code with $(93, 150)$ deep interleaving. The algorithm delay is 750 ms and the output bandwidth is 18.6 kb/s.

Scheme (2) "RS and Turbo concatenated codes". Totally 10 super-frames are buffered and encoded using $(31, 3, 29)$ RS code (the outer code) and rate $1/3$ Turbo code (the inner code) with $(36, 39)$ deep interleaving. The algorithm delay is 750 ms and the output bandwidth is 18.6 kb/s.

Scheme (3) "RS and LDPC concatenated codes". Totally 10 super-frames are buffered and encoded using $(31, 2, 30)$ RS code (the outer code) and rate $\frac{1}{2}$ $(1000, 2000)$ LDPC code (the inner code) with $(100, 140)$ deep interleaving. The algorithm delay is 750 ms and the output bandwidth is 18.6 kb/s.

Scheme (4) "The proposed realtime joint speech coding and transmission algorithm without error concealment". Three zero bits are added to make 48 bits each super-frame. No parity check or error concealment is applied. Each super-frame is divided into $N = 8$ groups and encoded by $(31, 6)$ BCH code. The transmission block length M is 5. There is no algorithm delay and the output bandwidth is just 16.5 kb/s.

Scheme (5) "The proposed realtime joint speech coding and transmission algorithm with error concealment". Speech parameters are classified by their importance and unequally protected. Three bits are added each super-frame to parity check the first and second stage LSF quantization vectors and the U/V speech parameter, resulting 48 bits each super-frame. Each super-frame is divided into $N = 8$ groups and encoded by $(31, 6)$ BCH code. The transmission block length M is 5. At the receiver side, error concealment is applied to the first and second stage LSF quantization vectors and the the U/V speech parameter if any error is detected by parity check bits. There is no algorithm delay and the output bandwidth is also 16.5 kb/s.

Table 1-4 show the error correcting performance of five schemes under different PLR and BER. Although RS codes enjoy good burst error correcting performance while Turbo and LDPC codes have near optimum performance and deep interleaving can transform burst errors into random errors, correcting performances of conventional shcemes are not satisfactory under harsh wireless channel conditions. When the PLR is higher than 1%, residual BER of conventional algorithms is above 10% which is unacceptable in speech communication. However the proposed algorithm actively utilizes joint source channel coding to counter the severe interference in channels. Even under the most severe condition (55% PLR and 15% BER), the residual BER is reduced by 84.36% compared with the conventional algorithms. The proposed algorithm achieves better error correcting performance than conventional algorithms with less bandwidth and no additional delay.

4.2 Reconstructed Speech Quality Comparison

The reconstructed speech quality is evaluated by ITU-T Rec PESQ [10] measurement which provides overall MOS scores identifying the reconstructed speech quality. Figures 4-5 present the MOS scores of five different schemes under different channel PLR and BER respectively. We can see that the proposed algorithm can effectively protect the important speech parameters hence improving the reconstructed speech quality compared with conventional algorithms. After unequal protection and error concealment are introduced, even better reconstructed speech quality is achieved, especially under harsh channel conditions where PLR is above 45%. Even when PLR and BER reach 55%, 15% respectively, the speech quality is still acceptable for wireless speech communication.

Table 1 Residual BER of scheme (1).

BER \ PLR	15%	30%	45%	50%
5%	5.15%	18.94%	25.34%	29.67%
10%	13.94%	21.96%	28.05%	31.94%
15%	20.04%	25.47%	30.75%	34.23%

Table 2 Residual BER of scheme (2).

BER \ PLR	15%	30%	45%	50%
5%	0.00%	20.00%	21.11%	34.00%
10%	0.00%	20.88%	30.66%	38.44%
15%	25.33%	28.00%	34.00%	34.44%

Table 3 Residual BER of scheme (3).

BER \ PLR	15%	30%	45%	50%
5%	13.11%	22.61%	28.80%	32.43%
10%	19.85%	25.78%	31.04%	34.53%
15%	24.23%	28.84%	33.50%	36.41%

Table 4 Residual BER of scheme (4) and scheme (5).

BER \ PLR	15%	30%	45%	50%
5%	0.050%	0.456%	2.150%	4.626%
10%	0.047%	0.477%	2.132%	4.605%
15%	0.121%	0.851%	2.987%	5.754%

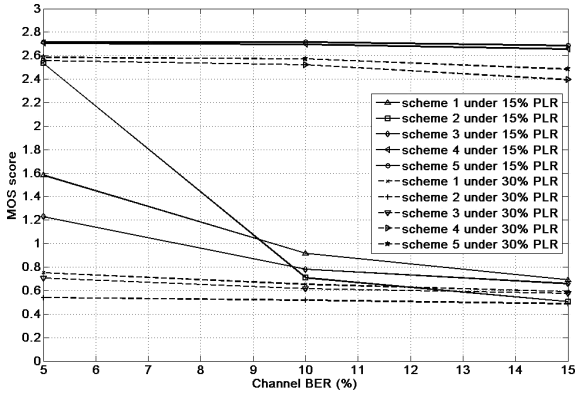

Fig. 4 MOS scores comparison under 15% and 30% PLR.

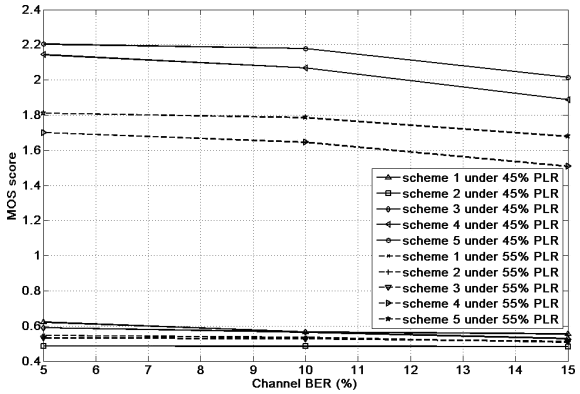

Fig. 5 MOS scores comparison under 45% and 55% PLR.

5. Conclusion

We have studied many classical coding and transmission algorithms and the proposed joint source channel coding algorithm for speech communication over noisy high packet loss rate wireless channels. Simulations show that the proposed algorithm can achieve better error correcting performance and reconstructed speech quality with less bandwidth expansion and no additional delay. The improvement is due to the inherent ability of MSBS at the transmitter, the MJR mechanism and error concealment at the receiver. Reasonable speech intelligibility can be achieved even under wireless channels with 55% PLR and 15% BER by using the proposed algorithm.

Acknowledgment

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

References

[1] Rahikka, T.E. Fujia, and T. Fazel, "Enhanced error correction of the US federal standard MELP vocoder employing residual redundancy of harsh tactical applications," NATO Tactical Mobile Communications Symposium TMC-99, Lillehammer, Norway, June 1999.

[2] Q. Wang and S.N. Koh, "Joint MELP turbo code with unequal error protection," IET Electronics Letters, vol.37, no.10, pp.637–639, May 2001.

[3] F. Lahouti and A.K. Khandani, "Soft reconstruction of speech in the presence of noise and packet loss," IEEE Trans. Audio, Speech, and Language Processing, vol.15, issue 1, pp.44–56, Jan. 2007.

[4] L. Yin, J. Lu, and Y. Wu, "LDPC-based joint source-channel coding scheme for multimedia communications," IEEE 8th International Conference on Communication Systems, vol.1, pp.337–341, Nov. 2002.

[5] W. Jiang and A. Ortega, "Multiple description speech coding for robust communication over lossy packet networks," IEEE International Conference on Multimedia and Expo, vol.1, pp.444–447, 2000.

[6] J.D. Edward and A.T. Keith, "Performance of FNBDT and low rate voice (MELP) over packet networks," Thirty-Fifth Asilomar Conference on Signals, Systems and Computers, vol.2, pp.1568–1572, Nov. 2001.

[7] J. Zhang, T. He, and J. Li, "High quality 0.6 kb/s speech coding algorithm," Journal of Tsinghua University, vol.43, no.4, pp.49–452, 2003.

[8] MIL-STD-3005, "Analog-to-digital conversion of voice by 2.400 bit/second mixed excitation linear prediction (MELP)," Department of Defense Telecommunications Systems Standard.

[9] S. Lin and D.J. Costeilo, Error Control Coding, 2nd ed., Prentice Hall Press, NJ, 2001.

[10] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," 2001.