Low-complexity PARCOR coder designed for entropy coding of prediction residuals

Yutaka Kamamoto*, Takehiro Moriya† and Noboru Harada‡

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation,
3-1 Morinosato-Wakamiya, Atsugi, 243-0198 Japan

(Received 7 October 2011, Accepted for publication 27 December 2011)

Keywords: Speech coding, Lossless compression, Linear predictive coding, Standardization, ITU-T G.711.0
PACS number: 43.72.Gy [doi:10.1250/ast.33.262]

1. Introduction

The ITU-T recommendation, G.711, is the prevailing speech coding technology for narrow-band telephony applications [1]. Recently, it is also used for packet-based networks such as voice over IP (VoIP). In order to reduce the bit rate of the G.711 payload without any degradation, a new ITU-T recommendation, G.711.0, was established [2]. In other words, G.711.0 enables the compression of the A-law/ μ -law logarithmic pulse-code modulation (PCM) signals losslessly.

This new codec is based on linear prediction and entropy coding; therefore, the quantization method of the partial autocorrelation (PARCOR) coefficients, which will be quantized and transmitted, affects the compression performance. For lossy coding, several quantization algorithms for PARCOR coefficients have already been investigated [3–6]. However, with G.711.0, there is a need for such a quantization algorithm for lossless coding. In this paper, we describe an algorithm for lossless coding based on a criterion that minimizes the entropy of the prediction residual signals [7–9]. Since this tool shows efficient performance in terms of bit reduction and complexity, this method is included in the G.711.0 standard (lower complexity is required for this standard [10]).

After the theoretical description, the experimental results will be shown.

2. Basic structure of mapped domain linear predictive coding in G.711.0

G.711.0 uses several coding tools, but more than half of the input speech signals are processed by mapped-domain linear predictive coding. The coding tool takes a sequence of N , number of samples per frame, G.711 A-law symbols $I_A(n)$ ($n = 1, 2, \dots, N$) or G.711 μ -law symbols $I_\mu(n)$. These G.711 symbols are converted into $x_{\text{pcm}}(n)$ in the linear PCM domain using the A-law/ μ -law for linear PCM conversion (i.e., decoding into the linear PCM domain). Once the input signals are mapped into the linear domain, they can be handled as ordinary linear PCM signals.

By transmitting the quantized PARCOR coefficients and the residual signals that are encoded by the Golomb-Rice code [11,12], the adaptive recursive Golomb-Rice code, or the

Huffman code with escape symbol (E-Huffman) [13], a recursive filter at the decoder can reproduce the original waveform losslessly. Consequently, the original G.711 data $I_{A/\mu}(n)$ are also reconstructed perfectly. This mapping strategy has also been discussed in other lossless compression schemes [14,15].

3. Low-complexity PARCOR coder for entropy coding of prediction residuals

3.1. Estimation of approximate code length

Here, we assume that the mean input speech signal is zero and that the Golomb-Rice code, the adaptive recursive Golomb-Rice code, and the E-Huffman code asymptotically approach the entropy of residual signals. As used in the intermediate process of the Levinson-Durbin algorithm, the energy of residual signals after P -th-order prediction, $E(P)$, can be estimated by PARCOR coefficients, so the variance of residual signals becomes

$$\sigma_p^2 = E(P)/N = (E(0)/N) \cdot \prod_{i=1}^P (1 - k_i^2), \quad (1)$$

where k_i ($i = 1, 2, \dots, P$) denotes the PARCOR coefficients ($\|k_i\| < 1$), and $E(0) = \sum_{n=1}^N \{x_{\text{pcm}}(n)\}^2$. When these signals follow a Gaussian distribution, their entropy can be denoted as

$$H_G(\sigma_p^2) = \log_2(\sqrt{2\pi e \sigma_p^2}). \quad (2)$$

When they follow Laplacian distribution, their entropy is described as

$$H_L(\sigma_p^2) = \log_2(\sqrt{2e^2 \sigma_p^2}). \quad (3)$$

In any case, these entropies depend on the variance, σ_p^2 ; therefore, Eqs. (2) and (3) can be transcribed as

$$H(\sigma_p^2) = \beta + \frac{1}{2} \log_2(\sigma_p^2), \quad (4)$$

where the constant β is about 2 for Gaussian distribution and around 1.9 for Laplacian distribution. From Eqs. (1) and (4), the estimated code length per frame is

$$\begin{aligned} C(P) &= H(\sigma_p^2) \cdot N + \sum_{i=1}^P \gamma_i \\ &= \left\{ \beta + \frac{1}{2} \log_2(\sigma_p^2) \right\} \cdot N + \sum_{i=1}^P \gamma_i \end{aligned}$$

*e-mail: kamamoto.yutaka@lab.ntt.co.jp

†e-mail: moriya.takehiro@lab.ntt.co.jp

‡e-mail: harada.noboru@lab.ntt.co.jp

$$= \frac{N}{2} \left\{ 2\beta + \log_2 \left(\frac{E(0)}{N} \right) + \sum_{i=1}^P \log_2(1 - k_i^2) \right\} + \sum_{i=1}^P \gamma_i. \quad (5)$$

Here, γ_i ($i = 1, 2, \dots, P$) are the bits needed to represent the PARCOR coefficients corresponding to k_i . We can see from Eq. (5) that the code length depends on the characteristic of the input signal and calculated PARCOR coefficients. Therefore, we can only consider variable terms of Eq. (5) as follows:

$$\begin{aligned} \tilde{C}(P) &= \frac{N}{2} \sum_{i=1}^P \log_2(1 - k_i^2) + \sum_{i=1}^P \gamma_i \\ &= \sum_{i=1}^P \left(\frac{N}{2} \log_2(1 - k_i^2) + \gamma_i \right) \\ &= \sum_{i=1}^P c(i), \end{aligned} \quad (6)$$

where $c(i)$ ($i = 1, 2, \dots, P$) denotes an estimated bit reduction for each prediction order. When N is large, γ_i becomes relatively small. However, when N is small, such as 40 and 80, γ_i should not be ignored, even though it will be encoded by the entropy coder. We should carefully select the precision for PARCOR coefficients and the number of coefficients. In this work, we focus on the case of 40 samples per frame ($N = 40$) because it is the most challenging condition.

3.2. Non-linear bit shift (NLBS) quantizer

There is a trade-off between the bit reduction term $(N/2) \log_2(1 - k_i^2)$ and bit representation term γ_i , depending on the quantizing precision. The quantizing error of PARCOR coefficients affects the compression performance. Here, the expected bit reduction λ_i in each sample is (the constant value is avoided)

$$\lambda_i = \log_2(1 - k_i^2) \leq 0. \quad (7)$$

Consequently, the inverse function can be assumed as

$$\|k_i\| = \sqrt{1 - 2^{\lambda_i}} \approx 1 - 2^{\lambda_i - 1} = \tilde{k}_i. \quad (8)$$

Its partial differential becomes

$$\frac{\partial \tilde{k}_i}{\partial \lambda_i} = -(\ln 2) \cdot 2^{\lambda_i - 1}. \quad (9)$$

Obviously, the PARCOR coefficients should be quantized as the criterion of the power-of-two space if each coefficient has the same responsibility for bit reduction. A quantization algorithm based on this criterion can be constructed by bit shift operation as well as by inverted A-law/ μ -law compression.

In G.711.0, the PARCOR coefficients are calculated using the 16-bit precision fixed-point arithmetic formula. Actually, Q15 format, which has 15 fractional bits and the real numbers between -1 and $+1$ mapped into the integer values between $-32,768$ and $+32,767$, is used and the maximum absolute value of integer PARCOR coefficients is limited to 32,750, i.e., $\|k_i\| \leq (32,750/32,768) = 0.99945$. For a simple explanation, we here demonstrate the Q8 format case and define K_i

as an integer notation. The developed non-linear bit shift (NLBS) quantization algorithm performs only the following several basic operations.

- (1) The absolute integer values of PARCOR coefficients (Q8 format ‘Sabc defg,’ where ‘S’ means sign, and each letter corresponds to zero or one) are defined as $K_i^+ = \|K_i\|$. In other words, ‘Sabc defg’ becomes ‘0abc defg.’
- (2) Integer values of K_i^+ are used for quantization and are arithmetically bit shifted to the right by $(8 - U_i)$, where U_i denotes the precision bits (let it be three here) for quantizing and depends on i ,

$$W_i = \lfloor 2^{-(8-U_i)} K_i^+ \rfloor$$

(i.e., ‘0abc defg’ $\rightarrow$ ‘0000 00ab’).

- (3) The absolute quantized PARCOR coefficient $\|\hat{K}_i\|$ is obtained by the bit shift operations $(8 - U_i - V_i - W_i)$ -right shifting and $(8 - U_i - V_i - W_i)$ -left shifting, where V_i is determined in advance (here, it is one):

$$\|\hat{K}_i\| = \begin{cases} 011c \text{ def0} & \text{if } ab = 11, \text{ i.e., } W = 3 \\ 010c \text{ de00} & \text{if } ab = 10, \text{ i.e., } W = 2 \\ 001c \text{ d000} & \text{if } ab = 01, \text{ i.e., } W = 1 \\ 000c \text{ 0000} & \text{if } ab = 00, \text{ i.e., } W = 0 \end{cases}. \quad (10)$$

- (4) Finally, an appropriate value ϵ is added for bias and a quantized PARCOR coefficient with sign is provided as

$$\hat{K}_i = \text{sign}(K_i) \cdot \|\hat{K}_i\| + \epsilon. \quad (11)$$

In summary, sparse quantization is applied to values around zero, whereas absolute values close to one are quantized densely. To reduce complexity, the table look-up method is adopted in the G.711.0 implementation.

3.3. Comparison of estimated bit reduction

The quantization of PARCOR coefficients affects compression performance. Theoretically, the energy reduced by the P -th coefficient is $E(P) = E(P-1) \cdot (1 - k_p^2)$. However, it changes to $\hat{E}(P) = E(P-1) \cdot (1 - k_p^2 + \Delta_p^2)$ when the quantized error of the coefficient is $\Delta_p = k_p - \hat{k}_p$ (where $\hat{k}_p$ denotes the quantized PARCOR coefficient). According to Eq. (1), the energy of predicted residual signals obtained by the first quantized PARCOR coefficient is

$$\begin{aligned} \hat{E}(1) &= \sum_{n=1}^N \{e(n)\}^2 \\ &= \sum_{n=1}^N \{x(n) + (k_1 + \Delta_1) \cdot x(n-1)\}^2 \\ &\approx R(0) + 2(k_1 + \Delta_1) \cdot R(1) + (k_1^2 + 2k_1 \cdot \Delta_1 + \Delta_1^2) \cdot R(0) \\ &= E(0) \cdot (1 - k_1^2 + \Delta_1^2), \end{aligned} \quad (12)$$

because $R(0) = \sum_{n=1}^N \{x(n)\}^2 \approx \sum_{n=2}^N \{x(n-1)\}^2$, $R(1) = \sum_{n=2}^N \{x(n) \cdot x(n-1)\}$, and $k_1 = -R(1)/R(0)$, as assumed in the Levinson-Durbin algorithm. Consequently, we must transcribe Eq. (6) as

$$\hat{c}(1) = \frac{N}{2} \log_2(1 - k_1^2 + \Delta_1^2) + \gamma_1. \quad (13)$$

Fig. 1 Comparison of estimated bit reductions (average of $\hat{c}(1)$, the bottom is the better).

Fig. 2 Comparison of estimated bit reductions (average of $\hat{c}(2)$, the bottom is the better).

Generally, when the quantization errors up to the $(i-1)$ -th order are ignored, we can assume the estimated bit reduction as described below:

$$\hat{c}(i) = \frac{N}{2} \log_2(1 - k_i^2 + \Delta_i^2) + \gamma_i. \quad (14)$$

Figures 1 and 2 show the difference in the estimated bit reduction of the first and second PARCOR coefficients, respectively, among our NLBS and well-known conventional quantization methods (bit-shift, arcsin, and arctanh) [6]. Since the first and second PARCOR coefficients have unique distributions and they are sensitive to the quantization in general, we show the results of $\hat{c}(1)$ and $\hat{c}(2)$, respectively. Here, frame size N is 40, and both γ_1 and γ_2 correspond to the X-axis. Input speech signals were obtained from ITU-T P.501 and these signals were resampled to narrow-band data. As can be seen from this graph, the developed method achieves bit reduction similar to the arcsin quantizer. Actually, the conversion of arcsin/sin and arctanh/tanh was performed with floating-point precision in this experiment, so the results of these two methods are the best case. The arcsin quantizers

Table 1 Comparison of estimated bit reduction for approximated arcsin quantizer and NLBS.

Quantization method	Minimum value of average of $\hat{c}(1)$
arcsin	
(w/o approximation)	-42.46
arcsin (4th polynomial Taylor series)	
$x + (1/6) \cdot x^3 + (3/40) \cdot x^5 + (5/112) \cdot x^7$	-42.22
arcsin (3rd polynomial Taylor series)	
$x + (1/6) \cdot x^3 + (3/40) \cdot x^5$	-42.00
arcsin (2nd polynomial Taylor series)	
$x + (1/6) \cdot x^3$	-41.95
NLBS	
(only a few operations)	-42.23

achieve maximum bit reduction in the case of the first prediction order. However, it was degraded when the function was approximated using Taylor-series for complexity reduction, as shown in Table 1. Finally, the developed NLBS quantizer yields better performance in terms of the trade-off between bit reduction and complexity.

4. Conclusion

We described the low-complexity PARCOR coder designed for entropy coding of prediction residual signals. The quantization of PARCOR coefficients is based on a criterion that minimizes the entropy of the prediction residual signals. The devised NLBS algorithm is included in the G.711.0 standard because this tool shows efficient performance in terms of bit reduction and complexity. It will be widely used in the near future because G.711.0 with the described low-complexity tool can compress the data rate of G.711, the prevailing speech coding technology.

References

- [1] ITU-T Recommendation G.711, *Pulse Code Modulation (PCM) of Voice Frequencies* (1988).
- [2] ITU-T Recommendation G.711.0, *Lossless Compression of G.711 Pulse Code Modulation* (2009).
- [3] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust. Speech Signal Process.*, **23**, 309–321 (1975).
- [4] A. H. Gray, Jr. and J. D. Markel, "Quantization and bit allocation in speech processing," *IEEE Trans. Acoust. Speech Signal Process.*, **24**, 459–473 (1976).
- [5] A. H. Gray, Jr., R. M. Gray and J. D. Markel, "Comparison of optimal quantizations of speech reflection coefficients," *IEEE Trans. Acoust. Speech Signal Process.*, **25**, 9–23 (1977).
- [6] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis* (Elsevier Science B. V., Amsterdam, 1995).
- [7] Y. Kamamoto, T. Moriya and N. Harada, "Low-complexity PARCOR coefficient quantizer for G.711.0 (lossless speech coding)," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 375–376 (2010).
- [8] Y. Kamamoto, T. Moriya and N. Harada, "Low-complexity PARCOR coefficient quantizer and prediction order estimator for lossless speech coding," *Proc. ICASSP 2010* (2010).
- [9] Y. Kamamoto, T. Moriya and N. Harada, "Low-complexity PARCOR coefficient quantizer and prediction order estimator

- for G.711.0 (lossless speech coding),” *Proc. DCC 2010* (2010).
- [10] ITU-T SG16 TD-33R1/WP3 Annex Q10.E, *Terms of Reference (ToR) and Time Schedule for G.711 Lossless Compression (G.711-LLC)* (2009).
- [11] S. W. Golomb, “Run-length encodings,” *IEEE Trans. Inf. Theory*, **12**, 399–401 (1966).
- [12] R. F. Rice, “Some practical universal noiseless coding techniques — part I–III,” *Jet Propulsion Lab. Tech. Rep.*, Vol. JPL-79-22, JPL-83-17, JPL-91-3 (1979, 1983, 1991).
- [13] N. Harada, Y. Kamamoto and T. Moriya, “Lossless compression of mapped domain linear prediction residual for ITU-T Recommendation G.711.0,” *Proc. DCC 2010* (2010).
- [14] F. Ghido and I. Tabus, “Accounting for companding nonlinearities in lossless audio compression,” *Proc. ICASSP 2007* (2007).
- [15] D. Huang, “Lossless compression for *mu*-law (A-law) and IMA ADPCM on the basis of a fast RLS algorithm,” *Proc. ICME 2000* (2000).