

Golomb-Rice Coding Optimized via LPC for Frequency Domain Audio Coder

Ryosuke Sugiura*, Yutaka Kamamoto†, Noboru Harada†, Hirokazu Kameoka† and Takehiro Moriya†

*Graduate School of Information Science and Technology,

The University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo, 113-8656, Japan

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

Abstract—We have devised a method to optimize Golomb-Rice coding of frequency spectra, aiming at its use in frequency domain audio coder, using spectral envelopes extracted by linear predictive coding (LPC) from amplitude spectra instead of conventional power spectra according to theoretical investigations. This optimization improves the efficiency of the Golomb-Rice coding by allocating Rice parameter at each frequency bin based on the value of the envelopes, resulting in the enhancement of the objective and subjective quality of the state-of-the-art wide-band coder at 16 kbit/s. Therefore, the method introduced here is expected to be useful for coding audio signals at low-bit-rate and low-delay conditions, required in mobile communications.

Index Terms—Audio coding, Golomb-Rice coding, Linear prediction, TCX

I. INTRODUCTION

For years, speech coding for mobile communications have greatly developed by narrowing the target of coding to speech signals: limiting the frequency band of the inputs as possible and using models specialized in coding speech signals [1]. However, to make the communications more comfortable, higher quality is required in coding wide-band audio inputs such as music. To encode the audio signals other than speech, it is known that coding in frequency domain is effective, as is done in the coders, for example, ITU-T G.722.1, 3GPP Extended Adaptive Multi-Rate WideBand (AMR-WB+) and MPEG-D Unified Speech and Audio Coding (USAC) [2]–[6]. The goal of this work is to design a high-quality frequency-domain audio coder with low delay and low bit rate.

The state-of-the-art frequency-domain coder [7] represents input signals in Modified Discrete Cosine Transform (MDCT) coefficients domain. The input frequency spectra represented by MDCT are quantized and entropy coded after the analysis of Linear Predictive Coding (LPC). In a context of perfectly lossless coding, whitening the inputs by LPC filters and simply minimizing the power or L1 norm of prediction errors are required to optimize the entropy coding [8]–[10]. However, in lossy coding where bits cannot be sufficiently allocated to the spectra, it is not the best way to perfectly whiten the spectra by LPC filters since quantization noise gets amplified when decoded. So the coder first performs perceptual weighting to the spectra which are approximated by smoothing the spectral envelopes extracted by LPC. Then the weighted spectra, or residual spectra, are scalar quantized and entropy coded in that domain. The perceptual weights, which slightly whiten the spectra, shape the quantization noise to make it inaudible. In fact, the coefficients of LPC, which are also quantized and coded, are used only for this weighting. However, there is still

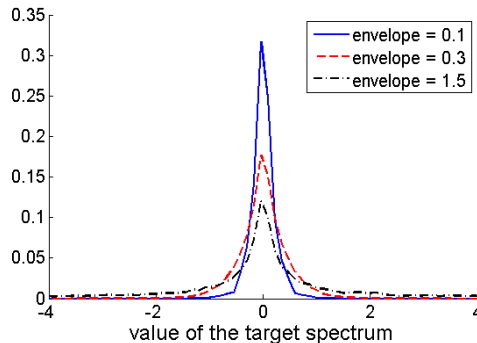


Fig. 1. Normalized histograms of the target spectra at the frequencies where the values of their envelopes are 0.1, 0.3 and 1.5, respectively.

redundancy between the quantized spectra and their envelopes since the perceptual weights do not perfectly whiten the input spectra. Fig. 1 shows the normalized histograms of the values of the weighted spectra at each frequency band where the values of their envelopes are 0.1, 0.3, and 1.5, respectively. The spectra are real values since they are MDCT coefficients. It can be seen that the spectra have relatively low variance where the values of their envelopes are low, and vice versa. In this paper, we choose for the entropy coding Golomb-Rice code [11], which requires low computational complexity to perform, and present a method to optimize the performance of Golomb-Rice coding using the envelopes.

The paper is organized as follows: We first explain, in section II, how to optimize Golomb-Rice coding by showing the relation between the optimization and LPC in a specific situation. Then in section III, the integration of the optimization with the real coder is described. Finally, the results of objective and subjective evaluations related to the method are presented in section IV.

II. OPTIMIZATION OF GOLOMB-RICE CODING

A. Basic idea

Golomb-Rice code is a variable-length code which is optimal when the targets of the coding are exponentially distributed, and it is used, for example, in SHORTEN [12], [13], LOW COMPLEXITY LOSSLESS COMPRESSION for Images (LOCO-I) [14], MPEG-4 Audio Lossless Coding (ALS) [15]–[20], and ITU-T G.711.0 [21]–[23]. In this paper, the targets of the coding are spectra which are quantized after perceptually

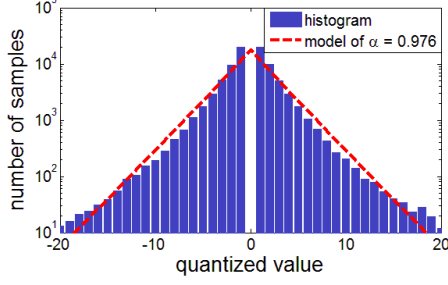


Fig. 2. Histogram of quantized spectra in 16 kbps. Red dashed line indicates generalized Gaussian distribution fitted to the histogram. 10 seconds each from 50 items in RWC music database are used in 16 kHz sampling rate.

weighted. Since zero values, especially in a low-bit-rate situation, appear in the targets with exceptionally high frequency, we think about coding the zeros by an exclusive method such as zero run-length coding. Fig. 2 shows the histogram of the targets excluding the zeros and the model of generalized Gaussian distribution fitted by the method reviewed in [24]. The generalized Gaussian distribution corresponds to Laplacian and Gaussian distribution when the shape parameter α is set to 1 and 2 respectively. The shape parameter of the histogram was approximated as $\alpha = 0.976$, which is near to 1, thus we can expect the targets excluding zeros to be exponentially distributed in this case. So here, as mentioned above, we use the Golomb-Rice code for entropy coding the weighted spectra and use run-length for zeros. Therefore, the following discussions focus on non-zero elements of the spectra, and for simplicity, we consider only the absolute values of the targets by separately coding their signs.

Golomb-Rice code has a parameter r called Rice parameter which stands for the length of fixed-length part. The length for coding an integer $Z (> 0)$ is written as

$$\mathcal{L}(Z|r) = 1 + r + \lfloor Z/2^r \rfloor \quad (1)$$

where $\lfloor \cdot \rfloor$ is a flooring operation. The Rice parameter r for coding Z should be a small value if Z is small, and vice versa.

As stated above, there is a relationship between the values of the target spectra and the values of their envelopes. Therefore, by choosing the proper Rice parameters $\{r_k\}_{k=0}^{N-1}$ for coding quantized spectra $\{y_k (> 0)\}$ based on the values of their envelopes at each frequency bin k , the performance of the coding can be enhanced. However, the conventional LPC is not optimal for this way of using since it assumes the signal to be Gaussian distributed, and the envelopes are expected only for the approximation of the perceptual weights. So here, we optimize the Golomb-Rice coding by considering both the optimal way to calculate the Rice parameters from the envelopes and to extract the envelopes from the spectra. For the approximation of the weights, these optimized envelopes can be used.

B. Requirements for the optimization

Two requirements must be considered for this optimization. First, the Rice parameters for coding have to be calculated

from the envelopes so that there is no need for sending additional information. Second, to save computational complexity, the conventional algorithm for LPC should be used for the extraction of the envelopes.

LPC can be written as a minimization problem of Itakura-Saito (IS) divergence between an all-pole filter and power spectra $\{|x_k|^2\}$ [25]:

$$\arg \min_{\sigma^2, \{a_n\}} \sum_k D_{IS}(\sigma^2 h_k || x_k|^2) \quad (2)$$

where

$$h_k = \left| \sum_{n=0}^p a_n e^{-j \frac{\pi n k}{N}} \right|^{-2}, \quad D_{IS}(X|Y) = Y/X - \ln(Y/X) - 1$$

with prediction gain σ^2 and LPC coefficients $\{a_n\}_{n=0}^p$. Therefore, the Rice parameter $\{r_k\}$ for Golomb-Rice coding the target spectra $\{y_k\}$ can be optimally parameterized by the method of LPC if the code length is represented in the form of IS divergence from the all pole model.

C. Minimization of the code length

Neglecting rounding effects, the code length of Golomb-Rice coding $\{y_k\}$ by Rice parameters $\{r_k\}$ can be written as

$$\begin{aligned} & \mathcal{L}(\{y_k\}|\{r_k\}) \\ & \simeq \sum_{k=0}^{N-1} (1 + r_k + \frac{y_k}{2^{r_k}}) = \sum_k (1 + \log_2 2^{r_k} + \frac{y_k}{2^{r_k}}) \\ & = (\log_2 e) \sum_k \left(\frac{y_k}{(\log_2 e) 2^{r_k}} - \ln \frac{y_k}{(\log_2 e) 2^{r_k}} - 1 \right) \\ & \quad + N(1 + \log_2 \ln 2 + \log_2 e) + \sum_k \log_2 y_k \\ & = (\log_2 e) \sum_k D_{IS}((\log_2 e) 2^{r_k} | y_k) + C(\{y_k\}) \end{aligned} \quad (3)$$

where C is a constant for $\{r_k\}$. By modeling $\{r_k\}$ with parameters $\tilde{\sigma}^2, \{\tilde{a}_n\}$ and rounding operation $\lfloor \cdot \rfloor$ as

$$\begin{aligned} r_k & \equiv \max(\lfloor \log_2((\ln 2) \tilde{\sigma}^2 \left| \sum_n \tilde{a}_n e^{-j \frac{\pi n k}{N}} \right|^{-2}) \rfloor, 0) \\ & \equiv \max(\lfloor \log_2((\ln 2) \tilde{\sigma}^2 \tilde{h}_k) \rfloor, 0), \end{aligned} \quad (4)$$

the code length approximately becomes

$$\mathcal{L}(\{y_k\}|\{r_k\}) \simeq (\log_2 e) \sum_k D_{IS}(\tilde{\sigma}^2 \tilde{h}_k | y_k) + C(\{y_k\}), \quad (5)$$

thus leading to

$$\arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \mathcal{L}(\{y_k\}|\{r_k\}) = \arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \sum_k D_{IS}(\tilde{\sigma}^2 \tilde{h}_k | y_k). \quad (6)$$

Just as the case in equation (2), the minimization of the code length can be solved by the way of LPC with $\{y_k\}$ regarded as power spectra. Moreover, $\{\tilde{\sigma}^2 \tilde{h}_k\}$ represents an envelope since $\{\tilde{\sigma}^2 \tilde{h}_k\}$ is fitted to the spectra $\{y_k\}$.

The solution of this minimization problem results in the same process which has been applied to TwinVQ [26]. However, it was solely intended for complexity reduction.

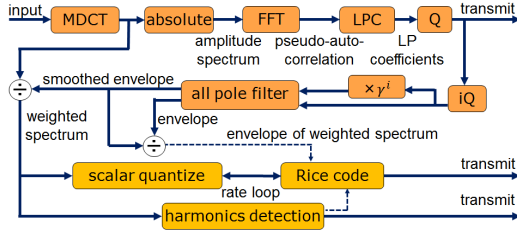


Fig. 3. Proposed coder based on [7]. "Q" and "iQ" stand for quantization and inverse quantization respectively.

III. INTEGRATION WITH CODER

A. Frequency domain audio coder

In this section, we apply the optimization of Golomb-Rice coding explained above to a frequency domain audio coder based on the idea in [7]. As explained in the introduction, the conventional coder performs perceptual weighting by approximating the weights from the smoothed envelopes as

$$w_k = \left| \sum_n (\gamma^n a_n) e^{-j \frac{\pi n k}{N}} \right|, \quad (k = 0, \dots, N-1) \quad (7)$$

where $0 < \gamma < 1$, and $\{a_n\}$ are the coefficients of LPC which are extracted from the auto-correlation function or the Fourier-transform of power spectra, or squared MDCT coefficients, of the signal in each frame. The perceptually optimal weights make the distortion in quantized spectra smaller in the peaks than the valleys of the spectra. It is experimentally known that the weights can be approximated by using $\gamma = 0.92$, and this γ is actually used in the state-of-the-art coders [3], [7].

B. LPC of amplitude spectra

Assuming that the perceptually optimal weights $\{w_k^*\}$ are given, the coding targets, or the quantized weighted spectra, can be written with the amplitude spectra $\{x_k\}$, or the absolute of the MDCT coefficients, as $y_k \equiv [w_k^* x_k / s]$ where s is the given step size of the scalar quantization. By modifying the model in section II as

$$r_k \equiv \max([\log_2((\ln 2) \tilde{\sigma}^2 w_k^* \tilde{h}_k / s)], 0), \quad (8)$$

the minimization of the length for Golomb-Rice coding the weighted spectra can be represented as

$$\begin{aligned} & \arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \mathcal{L}(\{y_k\} | \{r_k\}) \\ &= \arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \sum_k D_{IS} \left(\frac{\tilde{\sigma}^2}{s} w_k^* \tilde{h}_k | y_k \right) \\ &= \arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \sum_k D_{IS} \left(\frac{\tilde{\sigma}^2}{s} w_k^* \tilde{h}_k | \frac{1}{s} w_k^* x_k \right) \\ &= \arg \min_{\tilde{\sigma}^2, \{\tilde{a}_n\}} \sum_k D_{IS} (\tilde{\sigma}^2 \tilde{h}_k | x_k). \end{aligned} \quad (9)$$

Thus, the coding can be optimized by LPC of the amplitude spectra. Moreover, the spectral envelope $\{h_k\}$ has a similar property with the conventional envelope so that we approximate the weights $\{w_k\}$ using $\{\tilde{a}_n\}$ as

$$w_k \equiv \left| \sum_n (\gamma^n \tilde{a}_n) e^{-j \frac{\pi n k}{N}} \right|^2. \quad (10)$$

Considering the discussions above, we propose a coder outlined in Fig. 3. This coder performs LPC regarding Fourier-transformed amplitude spectra as pseudo-auto-correlation functions. In the Golomb-Rice coding, the Rice parameter of each frequency bin is calculated from the quantized LPC coefficients as

$$r_k = \max([\log_2(w_k h_k) + \bar{r}], 0) \quad (11)$$

where $\bar{r} = \log_2 \frac{\tilde{\sigma}^2}{s}$ stands for the average Rice parameter in the frame. Step size s of the quantization for the frame is chosen by a bisection search to meet the bit rate, and both the step size and the average Rice parameter are quantized.

Additionally, to enhance the performance of the coder, the harmonics of the inputs are detected and transmitted, which roughly indicates the interval of frequencies in which the non-zeros are likely to be. This harmonics information is used for modifying the zero run-length coding, and the encoder decides whether to use the information or not. Since the proposed method indicates that the code length can always be shortened by decreasing the IS divergence between amplitude spectra and their envelopes, we also used 3 bits in the proposed coder for compensating the envelopes of the harmonic components. The compensation was calculated by second order LPC of the harmonic components of the targets.

IV. EVALUATION

A. Performance of Golomb-Rice coding

The effects of optimizing the Golomb-Rice coding were evaluated. We focused only on encoding audio signals in the evaluations since the speech and audio coders such as AMR-WB+ and USAC are expected to encode, in most cases, speech signals in time domain by adaptively changing their modes and enhancing time-domain coding is beyond the scope of this paper.

We first prepared quantized spectra by using the proposed coder at 16 kbps, 32 kbps, 64 kbps, 128 kbps, 320 kbps, respectively. 10 seconds of signals, each from 50 items in RWC music database [27], were down-sampled into 16 kHz and coded at 20 ms per frame. Then, the quantized spectra were coded in Golomb-Rice code by using

- 1) One optimal Rice parameter for each frame,
- 2) Rice parameters calculated from the envelopes of the conventional LPC, or LPC of power spectra,
- 3) Rice parameters calculated from the envelopes of the proposed LPC, or LPC of amplitude spectra,

and the code lengths were compared with the ideal description length where the optimal Rice parameters were used for each bin. Fig. 4 shows the result. The proposed method of choosing the Rice parameters actually enhanced the performance of the coding in all cases. Meanwhile, the parameters calculated from the conventional LPC did not always enhance the performance since they were not optimized for the Golomb-Rice coding.

B. Objective quality of the coder

Objective experiments were performed to evaluate the effects of the proposed coder. We first, to prove the quality of the proposed coder, compared the coder with AMR-WB+, a reference method. The same signals used in the last section were coded in 16 kbps by both coders. The proposed

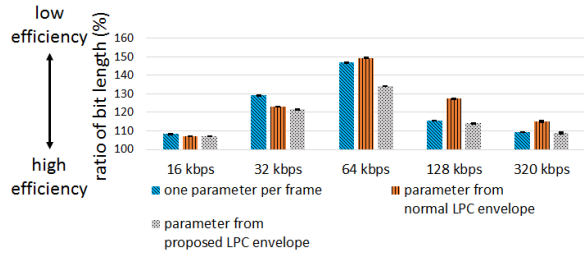


Fig. 4. The ratio to the ideal description length of Golomb-Rice code at each bit-rate using the Rice parameters calculated by each method. Average and 95% confidential interval. The 16th order LPC was used without quantizing the linear prediction coefficients.

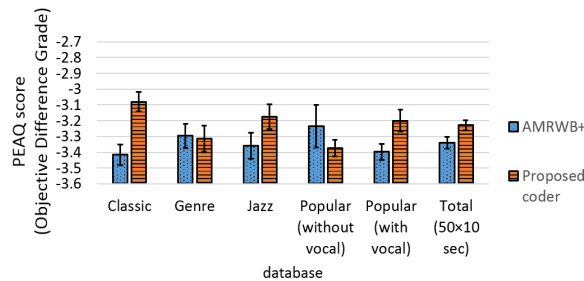


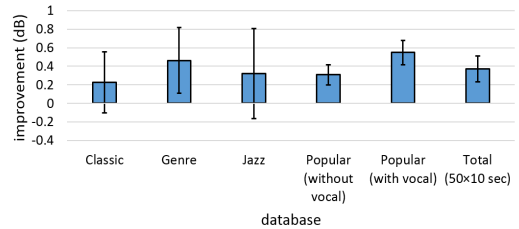
Fig. 5. PEAQ scores of AMR-WB+ and the proposed coder. Average and 95% confidential interval. Each category contains 10 items.

coder produces 40 ms of algorithmic delay while AMR-WB+ produces 144 ms at 16 kHz of internal sampling rate [28]. Fig. 5 shows the improvements from AMR-WB+ in objective quality. The objective quality was calculated by the method of Perceptual Evaluation of Audio Quality (PEAQ) in AFsp [29]. The proposed coder scored higher than AMR-WB+ in average.

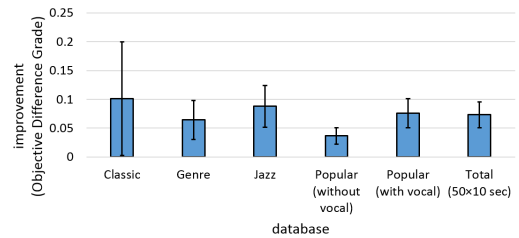
Next, we prepared the same coder as the proposed one except of the LPC and the compensation part: the conventional LPC was used instead of the proposed LPC without the compensation of envelopes. Here, we call this coder the conventional coder. The effects of changing the way of LPC were compared by the same items and the same conditions as stated above. Fig. 6 describes the database-wise difference in SNR of the weighted spectra and PEAQ scores. The SNR in the domain of perceptually weighted spectra increased by the proposed LPC since the performance of the Golomb-Rice coding was enhanced by the proper Rice parameters. Moreover, the improvements in PEAQ scores prove that the envelopes extracted by the proposed LPC can still be used for the approximation of the perceptual weights.

C. Subjective quality of the coder

Finally, an informal AB test was conducted to compare the subjective quality of the proposed and the conventional coder. Five items, 10 seconds each from RWC music database, were coded by the two coders: item 1 (a violin piece in classical music database), item 2 (a trumpet piece in music genre database), item 3 (a piano piece in jazz music database), item 4 (a guitar piece in popular music database) and item 5 (a



(a) SNR of the weighted spectra.



(b) PEAQ scores.

Fig. 6. Improvements by comparing LPC of amplitude spectra (proposed) over the conventional LPC. Average and 95% confidential interval.

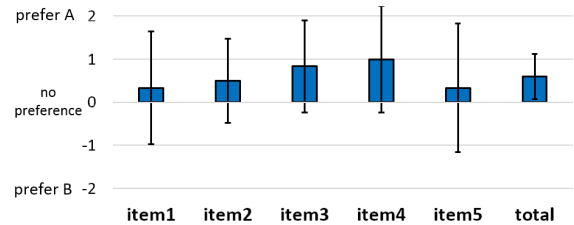


Fig. 7. Result of the subjective AB test. "A" for the proposed coder and "B" for the conventional coder. Score from -2 (prefer conventional) to 2 (prefer proposed). Average and 95% confidential interval.

male vocal piece in popular music database). Six participants evaluated the preference by scoring -2 to 2 points. The result is shown in Fig. 7. Although there was no significant preference in each item, total score improved on average at the significance level of 5%. The proposed method had a positive effect on the subjective quality of the coder.

V. CONCLUSION

In this paper, we introduced a method for optimizing Golomb-Rice coding by showing a theoretical consideration about the relation between the code length and IS divergence. This optimization enables us to calculate an efficient Rice parameter for each frequency bin from the value of the spectral envelope and enhances the performance of the coder. The proposed method of extracting envelopes can be combined with other techniques related to the representation of the envelopes like [30] or the conversion of LPC coefficients as in [31], which is expected to make further enhancement.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 26730100, 26280060.

REFERENCES

- [1] ITU-T G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)," 2012.
- [2] ITU-T G.722.1, "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 2005.
- [3] 3GPP TS 26.290 version 11.0.0 Release 11, 3GPP, 2012.
- [4] ISO/IEC 23003-3:2012, Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding
- [5] S. Quackenbush, "MPEG Unified Speech and Audio Coding," *MultiMedia*, IEEE Computer Society, vol. 20, issue 2, pp. 72-78, 2013.
- [6] M. Neuendorf, *et al.*, "MPEG Unified Speech and Audio Coding - The ISO/MPEG standard for high-efficiency audio coding of all content types", in Proc. AES 132nd Convention Paper, #8654, Apr., 2012.
- [7] G. Fuchs, *et al.*, "MDCT-based coder for highly adaptive speech and audio coding," in Proc. EUSIPCO, pp. 1264-1268, 2009.
- [8] Y. Kamamoto, *et al.*, "Low-complexity PARCOR coefficient quantizer and prediction order estimator for lossless speech coding," *Acoustical Science and Technology*, vol. 34, no. 2, pp. 105-112, 2013.
- [9] Y. Kamamoto, *et al.*, "Low-complexity PARCOR coefficient quantizer and prediction order estimator for G.711.0 (Lossless Speech Coding)," in Proc. Data Compression Conference, IEEE, pp. 475-483, 2010.
- [10] H. Kameoka, *et al.*, "A linear predictive coding algorithm minimizing the Golomb-Rice code length of the residual signal," *IEICE Transactions on Fundamentals of Electronics*, vol. J91-A, no. 11, pp. 1017-1025, Nov. 2008 (in Japanese).
- [11] R. F. Rice, "Some practical universal noiseless coding techniques - part I-III," *Jet Propulsion Laboratory Technical Report*, vol. JPL-79-22, JPL-83-17, JPL-91-3, 1979, 1983, 1991.
- [12] T. Robinson, "SHORTEN: Simple lossless and near-lossless waveform compression," *Cambridge Univ. Eng. Dept.*, Cambridge, UK, Tech. Rep. 156, 1994.
- [13] M. Hans and R. W. Schafer, "Lossless compression of digital audio," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 21-32, Jul., 2001.
- [14] M. J. Weinberger, *et al.*, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," in Proc. Data Compression Conference 1996, pp. 140-149, 1996.
- [15] ISO/IEC 14496-3:2009, Information technology – Coding of audiovisual objects – Part 3: Audio
- [16] T. Liebchen, *et al.*, "MPEG-4 Audio Lossless Coding," in Proc. AES 116th Convention, #6047, May, 2004.
- [17] T. Liebchen and Y. Reznik, "MPEG-4ALS: an emerging standard for lossless audio coding," in Proc. Data Compression Conference 2004, pp. 439-448, Mar., 2004.
- [18] Y. Reznik, "Coding of prediction residual in MPEG-4 standard for lossless audio coding (MPEG-4 ALS)," in Proc. ICASSP 2004, pp. III-1024-1027, 2004.
- [19] T. Liebechen, *et al.*, "The MPEG-4 Audio Lossless Coding (ALS) standard - technology and applications," in Proc. AES 119th Convention, Paper #6589, Oct., 2005.
- [20] S. Salomon and G. Motta, "Handbook of data compression," Springer, 2010.
- [21] ITU-T G.711.0, "Lossless compression of G.711 pulse code modulation," 2009.
- [22] N. Harada, *et al.*, "Emerging ITU-T standard G.711.0 - lossless compression of G.711 pulse code modulation," in Proc. ICASSP 2010, pp. 4658-4661, 2010.
- [23] N. Harada, *et al.*, "Lossless compression of mapped domain linear prediction residual for ITU-T recommendation G.711.0," in Proc. Data Compression Conference 2010, p. 532, Mar., 2010.
- [24] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, Jul., 1989.
- [25] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36-43, 1970.
- [26] T. Moriya, *et al.*, "Extension and complexity reduction of TwinVQ audio coder," in Proc. ICASSP 1996, IEEE, vol. 2, pp. 1029-1032, 1996.
- [27] [Online]. Available: <https://staff.aist.go.jp/m.goto/RWC-MDB/> (as of June '14).
- [28] [Online]. Available: <http://www.voiceage.com/AMR-WBplus.html> (as of Oct. '14)
- [29] [Online]. Available: <http://www-mmsp.ece.mcgill.ca/Documents/Software/Packages/AFsp/AFsp.html> (as of July '14)
- [30] R. Sugiura, *et al.*, "Representation of spectral envelope with warped frequency resolution for audio coder," in Proc. EUSIPCO, vol. TU-L03-1, 2014.
- [31] R. Sugiura, *et al.*, "Direct linear conversion of LSP parameters for perceptual control in speech and audio coding," in Proc. EUSIPCO, vol. TU-L03-2, 2014.