A Lossless and Lossy Audio Compression using Prediction Model and Wavelet Transform

Se Yil Park, Se Hyoung Park, Dae Sik Lim and Jaeho Shin
Department of Electronic Engineering,
Dongguk University,
Seoul,100-715, Korea.
Tel. +82-2-2260-3336, Fax.: +82-2-2272-5667
e-mail : wildred@dongguk.edu

Abstract: In this paper, we propose a structure for lossless audio coding method. Prediction model is used in the wavelet transform domain. After DWT,wavelet coefficients is quantized and decorrelated by prediction modeling. The DWT can be constructed to critical bands. We can get a lower data rate representation of audio signal which has a good quality like the result of perceptual coding. Then the prediction errors are efficiently coded by the Golomb-coding method. The prediction coefficients are fixed for reducing the computational burden when we find prediction coefficients.

1. Introduction

In recent years, lossless audio coding is highly concerned. There are two parts in audio coding. One is lossy coding method, and the other is lossless one. In lossy part, there are MPEG I layer 1/2/3, ACC, and AC-3 etc. Especially, the Layer-3 of MPEG I is used popular in audio service through Internet since it has a high compression ratio(more than 1:12) and provides CD-quality sound. Thus, it is frequently used to store and signals in disc, Meanwhile, lossless audio signal compression methods haves been rarely studied compared to lossy compression. However, the demand of high quality and lossless compression of the original audio signals is gradually increased. Typical applications are It is useful in digital versatile disc(DVD) and audio signal database systems which require the nature of the original audio signal or the characteristics. [1].

Lossy audio coding schemes are mostly using perceptual property of human hearing, but lossless coding is using prediction[1] or transform[2] method which reduce the redundancy of signal. Also, it must be decoded easier then lossy compression. General lossless data compression method like zip, pkzip are not adequate for audio signal because the properties of audio signal is different from text data properties. Therefore, audio signal is required its own lossless compression method [5].

There are several methods for lossless audio coding such as Shorten, LTAC, MusiCompress, and AudioPaK [1][2][5]& [11]. Shorten[1] proposed linear prediction method in time domain. OggSquish and DVD-audio are using IIR prediction filters[10]&[11]. But LTAC[2] is a transform coding using discrete cosine transform(DCT). It has the flexibility of audio quality and bit rate.

•In this paper, we propose a new method which execute prediction analysis in Discrete Wavelet Transform (DWT) domain. Since DWT has the properties of perfect reconstruction and energy compaction, it is possible to apply the transform coding based on lossy coding. Furthermore, the coefficients of DWT are very correlated like original signal in time domain, and we can reduce the redundancy by decorrelating the DWT coefficients.

The basic operations in lossless compression algorithms are well defined in [1]&[5]. The first step of the operation is the framing operation that provides an important and necessary property for most applications with digital audio. In this operation, digital audio signal is divided into independent frames of equal time duration. If this time scale is too short or long, it incurs overhead in the computation and transmission. And the framing gives editing ability. Thus, it is important to determine the appropriate frame length.

After framing, input signals go through intrachannel decorrelation block followed by Entropy coding. This intra-channel decorrelation is inevitable to achieve large compression ratio. There are two types to reduce data redundancy. One is prediction (analysis), and the other is transform coding. We will examine these two methods in the following subsections, and then we describe the entropy coding for residual signals and error signals that are generated from intrachannel decorrelation operation.

The rest of the paper is organized as follows. In the next sections(2, 3, 4) we describe the operations involved in the proposed compression method and they are prediction process, transform coding process, and entropy coding process. In Section 5 the proposed coding scheme is presented. It also explains the transform and the prediction procedure for the proposed lossless coding and evaluates the performance. Finally, we make a conclusion in Section 7.

2. Transform Coding

Figure. 1-(a) shows a lossless coding scheme as a combination of lossy coding and additional error transmission. LTAC[2] is an example of lossless transform coding system exploiting discrete orthonormal transform as shown in Figure. 1-(b). The T and T^{-1} represent an arbitrary orthonormal transform with block length N and its inverse, respectively, and the transform is discrete cosine transform (DCT) in [2]. The reason why DCT is used in lossless transform coding is energy compaction property of DCT, which permits a good reconstruction from a few quantized values of the coeffici-ents. Many coefficients are very small or even zero, and moreover, they constitute an uncorrelated source when using a suitable transform.

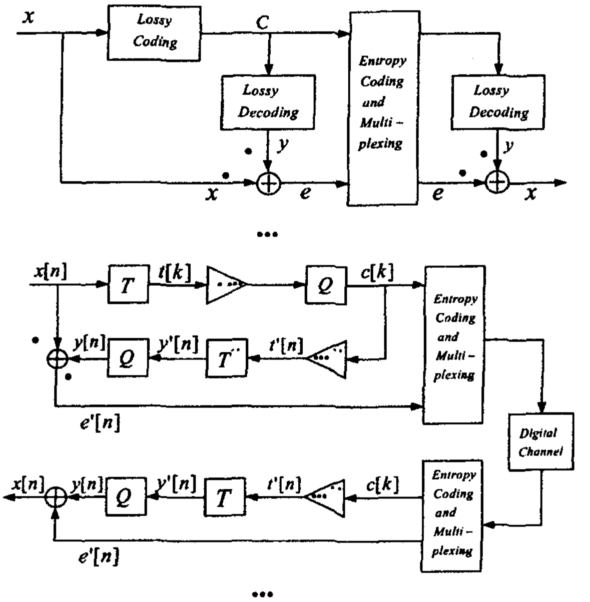

Figure 1. Transform coding based on lossy coding a) losseless coding scheme as a combination of lossy coding as a combination of lossy coding and additional error transmiss-ion, b) a lossless transform coding system proposed by M. Purat [2].

The next stage of Figure 1-(b) is scaling α and quantization Q . The coefficients are scaled by α =1 and then quantized with a unitary quantization step size $\Delta = 1$. The result of the quantization is equivalent to integer truncation, so we cannot reconstruct the signal perfectly.

In this paper, we also propose a transform coding scheme using discrete wavelet transform(DWT). The DWT also supports orthogonal transform and perfect reconstruction and has good energy compaction property.

An advantage of transform coding based on lossy coding is that the lossless representation contains both lossy representa tion and error of the original audio signal. This is useful when the data rate is variable, or perfect reconstruction is not necessary.

3. Prediction Modeling

One of the most popular speech coding techniques is linear predictive coding. Although audio signals are different from speech, it is possible to apply to audio signals. The basic idea behind the linear predictive coding is that a speech/audio sample can be approximated by a linear combination of the past speech/audio samples. A unique set of predictor coefficients can be determined by minimizing the sum of the squared differences between the actual speech samples and the linearly predicted samples[9].

Compression is achieved by building a predictive model of the waveform, also known as linear predictive coding (LPC). The purpose of the prediction in the lossless audio coder is to remove redundancy by decorrelating the samples within a frame. Figure. 2-(a) and (b) show predictive modeling scheme and general structure of the prediction block, respectively [5][10]& [11].

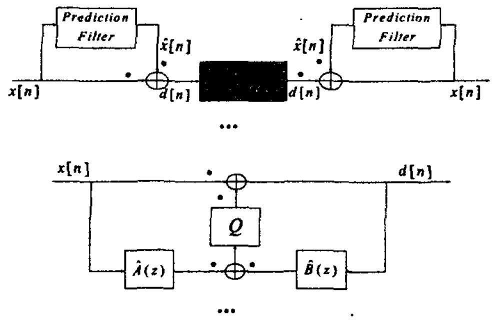

Figure 2. a) Linear prediction lossless coding system. b) general structure for prediction

Figure 2(a) depicts the equation ,

$$d[n] = x[n] - \hat{x}[n] , \qquad (1)$$

and $\hat{x}[n]$ is given by

$$\hat{x}[n] = Q\left\{ \sum_{i=1}^{M} \hat{a}_i x[n-i] - \sum_{i=1}^{N} \hat{b}_i d[n-i] \right\} \qquad (2)$$

where $Q\{\}$ denotes quantization. In Figure. 2-(b) we find two elements $\hat{A}(z)$ and $\hat{B}(z)$. If the feedback element is $\hat{B}(z)=0$, the prediction filter has finite impulse response (FIR). If the feedback element $\hat{B}(z)\neq 0$, the prediction filter has infinite impulse response(IIR).

When all-pole model is used, the coefficients of the prediction filter are computed by the standard Durbin's algorithm from autocorrelation coefficients[1]. Meanwhile, the characteristic of IIR predictor are explained in [10]. The IIR prediction has not been widely used in other lossless coding and speech coding because of the computation complexity of IIR prediction.

Residual signal $d[n]$ after prediction is quantized for effective coding. When residual signals and predictors are quantized, the effect of quantization error has to be investigated. Shorten[1] reported that there is no change in SNR if FIR predictors have been coded by more than 7bit word length. In IIR predictors' case, the word length is required more than 8bits [10].

4. Entropy Coding

The residual signal $d[n]$ from prediction and the error signal $e[n]$ from transform coding must be converted into bit streams for storage or transmission. Residual signal and error signal are assumed to be uncorrelated. Entropy coding removes redundancy from $d[n]$ and $e[n]$ themselves. In this process, no information will be lost, error signal and residual signal may be coded, respectively.

The Golomb-Rice coding which is designed to encode integer is a simple and efficient way. Golomb code is optimal for exponentially decaying(geometric) probability distributions of the nonnegative integers [12]. The Golomb code is actually a family of codes parameterized by an integer $m > 0$. In the Golomb code with parameter m , we represent an integer $I > 0$ using two numbers q and r , where

$$q = \left\lfloor \frac{I}{m} \right\rfloor \quad (3)$$

and $r = n - qm$ (4)

where $\lfloor x \rfloor$ is the integer part of x, q is the quotient and r is the remainder when I is divided by m. The remainder r can take on the values $0,1,2,\ldots,m-1$. If m is a power of two, we use the $\log_2 m$ -bit binary representation of r. If m is not a power of two, we could still use $\lceil \log_2 m \rceil$ bits, where $\lceil x \rceil$ is the smallest integer greater than or equal to x. We can reduce the number of bits required if we use the $\lfloor \log_2 m \rfloor$-bit binary representation of r for the first $2^{\lceil \log_2 m \rceil} - m$ values, the $\lceil \log_2 m \rceil$-bit binary representation of $r + 2^{\lceil \log_2 m \rceil} - m$ for the rest of the values [7].

Golomb-Rice coding is the special case of Golomb code with $m = 2^k$. Choosing m to be a power of 2 leads to very simple encoding/decoding procedure: the code for I consists of the k least significant bits of I, followed by the number formed by the remaining higher order bits of I, in unary representation. The length of the encoding is $k + 1 + \lfloor I/2^k \rfloor$ [12]. Golomb-Rice coding is the algorithm for lossless data compression recommended by the Consultative Committee on Space Data Standards(**CCSDS**).

In equation (1), $d[n]$ (or error signal $e[n]$) will have a small magnitude when our prediction is good and a large value when it is not. Let input signal x_{max} and x_{min} be the largest and smallest values that the sequence $\{x[n]\}$ takes on.

Define

$$T_i = \min\{x_{max} - \hat{x}, \hat{x} - y_{min}\}. \quad (5)$$

The sequence $\{d[n]\}$ can be converted into a sequence of nonnegative integers $\{g[n]\}$ using the following mapping:

$$g[n] = \begin{cases} 2d[n] & 0 \le d[n] \le T \\ 2|d[n]| - 1 & -T \le d[n] < 0 \\ T + |d[n]| & otherwise. \end{cases} \quad (6)$$

The value of $g[n]$ will be small whenever the magnitude of $d[n]$ is small. Therefore, the value of $g[n]$ will be small with higher probability. And then, the coded block is transmitted along with an identifier that indicates which particular option was used [5][7]&[12].

5. Prediction in wavelet transform domain

In this section, we propose a prediction in DWT domain and then execute entropy coding for residual signals from the prediction in DWT and transform error signals. Figure. 3 is the entire block diagram of proposed lossless audio coding scheme.

In the first, the input signal that is coded by PCM sample of 16bit or 20bit goes to framing block. In framing block, input signal is divided into a frame which has the length power of 2, 512~4096. In simulation, the lengths of 2048 or 4096 have shown good performance for a sampling rate of 44.1kHz. The time•interval for the lengths are 46 and 92ms, respectively. The block length can be varied, but we fixed to 4096 for convenience.

After framing, each frame block $x[n]$ is transformed by discrete wavelet transform(DWT). The DWT is constructed to critical bands. The bandwidths corresponding to one critical band are shown in Table 1.

There are many different scales proposed, and one of the most popular method is given by [13],

$$CB = 13\tan^{-1}\left(\frac{0.76f}{1000}\right) + 3.5\tan^{-1}\left(\left(\frac{f}{7500}\right)^2\right) \; Barks \quad (7)$$

or

$$\Delta f / Hz = 25 + 75[1 + 1.4(f/kHz)^2]^{0.69} \quad (8)$$

We divide input audio signal(sampling rate is 44.1kHz) into 25 critical bands using wavelet packet tree as shown in Figure. 4. Some mismatch is inevitable in DWT that divides audio band into critical band by dyadic, but this approximation is acceptable in this algorithm. New critical band by DWT is generated. DWT is well suited to construct cochlear filter of Human Auditory System(HAS) which is non-linear filter bank[6].

The basic idea of DWT transform coding based on lossy coding scheme is that the lossless representation contains a lossy lower data rate representation of the audio signal in each critical band. It is easily obtained by discrete wavelet transform which constructs cochlear filter bank(Critical band). Every critical subband is coded respectively. If we take a different scaling factor α and quantization step size $\Delta \ge 1$, we can get a lower data rate representation of audio signal which has a good quality like the result of perceptual coding. But error signal $e[n]$ will be increased. It is useful where transmission bit rate is varied through the network.

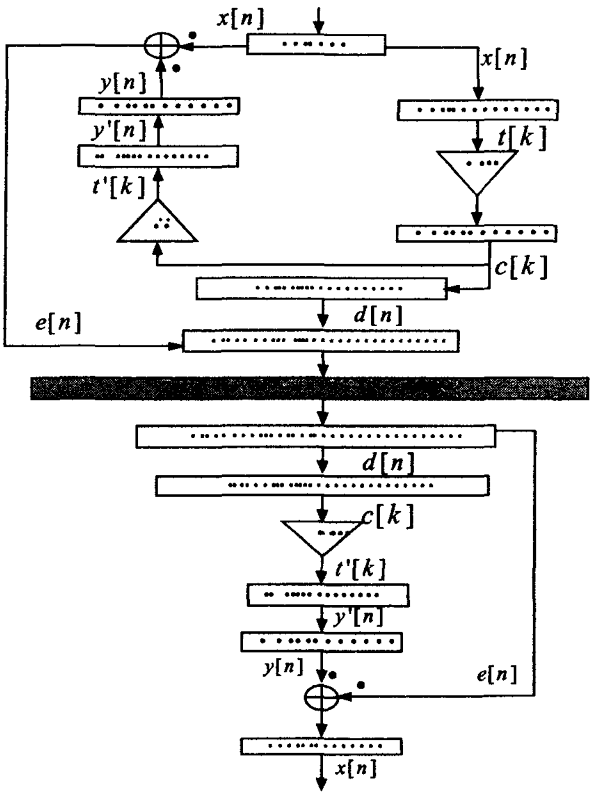

Figure. 3 Entire block diagram proposed lossless coding algorithm

After DWT, the coefficient $c[k]$ must be efficiently coded. But the coefficient is a kind of filtered downsample version of input signal, so it still has correlation itself. So we apply

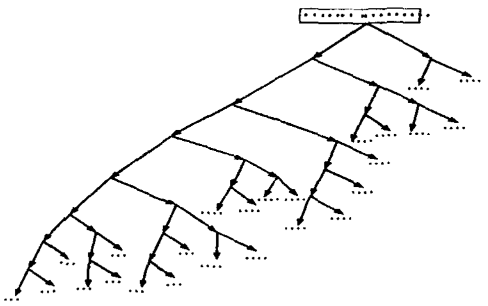

Figure. 4 Wavelet decomposition tree for critical bands.

prediction filtering to the coefficient $c[k]$ and get prediction residual signal $d[k]$.

As described above in section 3, we can use two types of predictor, FIR and IIR. The predictor of FIR type which is computed by Durbin's method is not work well in DWT domain. The amplitude of DWT coefficients is varied faster than in time domain, because DWT executes filtering with wavelet filters and downsampling. The predictor of IIR type which is computed by prony method is work well in DWT domain. We use 2nd order IIR filter that the coefficients of the filter is quantized to 16bits. We don•• t consider about the quantization effect of predictor, because word length to quantized the coefficients is enough to neglect the effcts. And then the residual signal $d[k]$ goes through inverse prediction filter and inverse discrete wavelet transform (IDWT). So we generate error signal $e[n]$ is calculated by

$$e[n] = x[n] - y[n] \qquad (9)$$

where $x[n]$ is blocked signal from framing and $y[n]$ is reconstred signal with lossy coding scheme. Lastly residual signal $d[k]$ and error signal $e[n]$ is coded by Golomb-Rice coding which is designed for integers.

6. Results

Compreesion ratio is given by

$$r = \frac{\textit{original file size}}{\textit{compressed file size}} \qquad (10)$$

As you can see in Table 1, the result of compression ratio of proposed lossless coding scheme has a little bit lower than other scheme. However it is available both for lossless and lossy coding. For lossy coding, we removed a cb25 (dropping one band) of the critical and three bands (cb25,cb24 and cb23) in Figure 4.

7. Conclusion

We have presented a lossless audio coding algorithm, which is based on orthonormal signal transforms and predictive coding. The use of a simple linear predictor and DWT followed by Rice coding has been exploited. We have shown that perfect reconstruction of the 16-bit input signal can be achieved by integer transform coefficients and additional transmission of the error correction information. The proposed algorithm can also support lossy compression for a specified bit rate and a specified signal to noise ratio.

Table 1. Compression ratio LTAC, Shorten,and Proposed Scheme.

Max CD (20 Second)	LTAC	Shorten	Proposed Scheme		
			Lossless	Lossy	
				drop band	drop 3band
Track1	1.298	1.227	1.229	1.707	2.384
Track2	1.511	1.489	1.389	1.918	2.606
Track3	1.301	1.289	1.236	1.716	2.377
Track4	1.394	1.325	1.313	1.808	2.397
Track5	1.754	1.642	1.555	2.119	2.766
Track6	1.415	1.355	1.314	1.839	2.568
Track7	1.557	1.467	1.410	1.953	2.614
Track8	1.398	1.362	1.286	1.773	2.389
Track9	1.552	1.468	1.436	1.987	2.589
Track10	1.473	1.351	1.384	1.933	2.534
Average	**1.465**	**1.397**	**1.356**	**1.876**	**2.523**

Acknowledgments

This research was supported by Brain Korea 21 Program of Korea Research Foundation.

References

[1] T.Robinson, "SHORTEN: Simple lossless and near-lossless waveform compression", Cambridge Univ.Eng. Dept., Cambridge, UK, Tech.Rep.156, 1994.

[2] M.Purat, T.Liebchen, and P.Noll, "Losselss transform coding of audio signals," in Proc. 101nd AES Conv., Munich, Germany, 1997, preprint 4414.

[3] M.Hans, "Optimization of digital audio for Internet transmission," Ph.D. dissertation, School Elect. Comp. Eng., Georgia Inst. Technol., Atlanta, 1998.

[4] N.Jayant and P.Noll, *Digital Coding of Waveforms, Principles and Applications to Speech and Video*, Eng ewood Cliffs, NJ: Prentice-Hall, 1984.

[5] M.Hans, R,W. Schafer, "Lossless Compression of Digital Audio," IEEE Signal Precessing Magazine, July, 2001.

[6] Ali N. Akansu, Mark J.T. Smith, *Subband And Wavelet Transforms Design and Applications*, Kluwer Academic Publishers, 1996.

[7] Khalid Sayood, *Introduction to Data Compression*, Morgan Kaufmann, 2000.

[8] Peter Graven, M.Law, and J.Stuart, "Lossless compression using IIR prediction filters,"in *Proc. 102nd AES Conv.*, Munich, Germany, 1997, preprint 4415.

[9] L.R. Rabiner and R.W. Schafer, *Digital Preocessing of Speech Signals*,Englewood Cliffs, NJ: Prentice-Hall, 1978.

[10] P.Craven, M.Law, and J.stuart, "Lossless compression using IIR prediction filters," in Proc.102nd AES Conv., Munich, Germany, 1997, preprint 4415.

[11] A.Bruekers, A.Oomen, and R.van der Vlenuten, "Lossless coding for DVD audio, " in Proc. 101st AES Conv., Los Angeles, Ca, Nov.1996, preprint 4358.

[12] M.weinberger, G. Seroussi, and G.Sapiro, "LOCO-I: A low complexity,context-based,lossless image compression algorithm," in Proc. Data Compression conf., 1996.

[13] E. Zwicker and H. Fastl, *Psychoacoustic: Facts and Models*, 2nd edition , Springer-Verlag, Berlin Heidelberg, 1999.