

AN EFFICIENT TONAL COMPONENT CODING ALGORITHM FOR MPEG-2 AUDIO NBC

Yuichiro TAKAMIZAWA

Masahiro IWADARE

Akihiko SUGIYAMA

Information Technology Research Laboratories, NEC Corporation

E-mail: takami@dsp.CL.nec.co.jp

ABSTRACT

This paper proposes a tonal component coding algorithm for a codec that employs a transform followed by Huffman coding, such as MPEG-2 Audio NBC (Non-Backward Compatible)[1]. After the input audio signal is mapped onto a frequency domain, the proposed algorithm withdraws local maximum components that degrade coding efficiency. By this withdrawal, the flatness of the spectrum increases and the efficiency in Huffman coding is improved. The withdrawn components are encoded separately as side information. When the frequency resolution of the time/frequency mapping is high, this algorithm works more effectively since local maximum samples appear more frequently with such a mapping. Simulation results show that this algorithm achieves as much as 11% bit reduction per frame and improves the coding efficiency in 41% of all the audio frames.

1. INTRODUCTION

Recently, efficient audio signal coding algorithms have widely been studied for digital storage media and ISDN (Integrated Services Digital Network) applications. To achieve high coding efficiency, MDCT (Modified Discrete Cosine Transform), psychoacoustic analysis and entropy coding have been employed in recent coding algorithms. Among such algorithms are MPEG-1/Audio Layer III[2], ASPEC[3] and MPEG-2 Audio NBC. These algorithms consist of five major blocks, time/frequency mapping, psychoacoustic analysis, quantization, noiseless coding and frame packing. Figure 1 shows a block diagram of MPEG-2 Audio NBC Reference Model 2 (RM2)[4], which is the latest among these codecs. The input audio samples are transformed into a frequency domain by MDCT. The frequency-domain samples are quantized in the quantization block. The psychoacoustic analysis controls quantization to improve the subjective coding quality. The noiseless coding removes the redundancy from the quantized samples by utilizing Huffman coding. The frame packing block assembles the bitstream.

As a noiseless coding algorithm, similar algorithms are employed in MPEG-1/Audio Layer III, ASPEC and NBC RM2. However, the frequency resolution of the time/frequency mapping is different among these codecs. The frequency resolution of NBC RM2 is twice as higher compared to ASPEC and MPEG-1/Audio Layer III. When the frequency resolution is high, a few frequency-domain

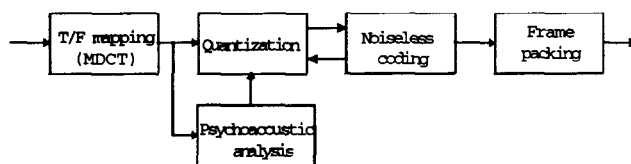


Figure 1. Block diagram of MPEG-2 Audio NBC RM2.

components tend to have relatively large amplitudes to adjacent components. Especially, this tendency becomes strong for an audio signal with tonal components. These components sometimes degrade the coding efficiency in the noiseless coding block.

This paper proposes a tonal component coding algorithm that improves the coding efficiency in noiseless coding. The tonal components which degrade the coding efficiency are withdrawn before the noiseless coding is performed. The withdrawn samples are separately encoded as side information. The performance of the proposed algorithm is evaluated by computer simulation results.

2. CONVENTIONAL NOISELESS CODING ALGORITHM

In the noiseless coding block where the redundancy of the quantized samples is removed, the samples are combined together in its frequency order and compose $\vec{X} = [X_1, X_2, \dots, X_N]$ where N is the block length of MDCT and $X_i (i = 1, 2, \dots, N)$ are quantized samples. These elements of $\vec{X}$ are partitioned into three regions, R1, R2 and R3, according to their amplitudes. Each region is encoded by different methods for better noiseless coding efficiency.

In general, when audio signal is mapped with MDCT, the higher the frequency is, the smaller the spectrum value is. Most of the quantized spectrum values are zero in the high frequency. These zero-valued elements compose R3. The quantized spectrum value gradually becomes larger as the frequency becomes lower. R2 contains the elements whose absolute values are less than two. The low frequency elements which have large amplitudes compose R1.

The partitioning is performed based on the amplitude of the elements. The elements are examined one after another from the highest-frequency element, X_N , until the first non-zero element, X_{R2end} , is encountered. These zero-valued elements, $X_{R2end+1}, X_{R2end+2}, \dots, X_N$, compose R3. Then,

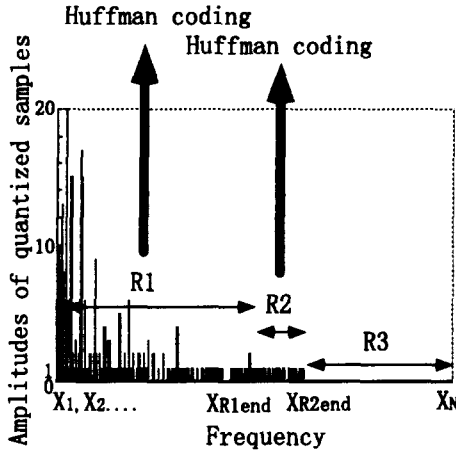


Figure 2. Conventional noiseless coding algorithm.

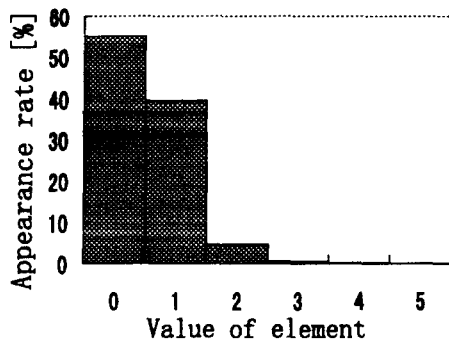


Figure 3. Distribution of element values in R1 which are neighboring R2.

the examination continues from X_{R2end} until the first element X_{R1end} , whose value is more than one, is detected. These elements, $X_{R1end+1}, X_{R1end+2}, \dots, X_{R2end}$, compose R2. All other elements compose R1. This partitioning is illustrated in Figure 2. The values of the elements in R1 and R2 are Huffman-encoded. An appropriate Huffman table is selected for R1 and R2 respectively so that the maximum value in the region can be encoded. Since any value up to the maximum value in the region must be encoded with the selected table, the average number of bits to encode an element increases in accordance with this maximum. Therefore, when a few large values exist in a region, the coding efficiency is degraded. This degradation becomes a serious problem especially in a low bitrate encoder with a high frequency resolution such as MPEG-2 Audio NBC. This is because a few components tend to have large value when the tonal signal is mapped onto the frequency domain with a high frequency resolution.

One example that degrades coding efficiency is the elements in R1 which are neighboring R2. The absolute values of these elements are generally less than two as well as el-

ements in R2. Figure 3 shows a distribution of element values in R1 when six different audio sources are encoded. The encoded sources are Harpsichord, Castanets, Speech, Bagpipes, Glockenspiel and Pitchpipe (48 kHz sampling, monaural). The distribution is examined for the elements in R1 which are neighboring R2. The number of examined elements is a quarter the number of elements in R2. The 95% of the elements in R1 which have been examined have the value less than two. The average code length for an element is shorter in R2 than in R1. Therefore, the coding efficiency is improved if these elements are included in R2. However, due to the remaining 5% elements whose values are more than two, these elements are included in R1.

3. PROPOSED TONAL COMPONENT CODING ALGORITHM

The proposed algorithm withdraws some large-amplitude samples from the quantized frequency-domain samples before Huffman coding. This withdrawal makes the maximum sample value in the region smaller and enables the use of a smaller Huffman code table. Generally, a smaller Huffman code table makes the average code length per symbol shorter. Therefore, the coding efficiency is improved. The withdrawn samples are encoded separately as side information. If this withdrawal does not make any gain, conventional noiseless coding algorithm with no withdrawals is performed.

Figure 4 illustrates the proposed algorithm. The elements in R1 are separated into two groups, local maxima that degrade the coding efficiency and the others. Local maxima are the elements whose values are more than one in R1 which are neighboring R2.

The actual encoding procedure consists of three steps. As the first step, local maximum samples to be withdrawn are detected. At the second step, $\tilde{Z} = [Z_1, Z_2, \dots, Z_N]$ that represents local maxima and $\tilde{Y} = [Y_1, Y_2, \dots, Y_N]$ that represents the others are generated. $\tilde{Y}$ is derived by replacing withdrawn elements of $\tilde{X}$ by zero. $\tilde{Z}$ is calculated by $\tilde{X} - \tilde{Y}$. In other words, $\tilde{Y}$ includes at least A non-zero elements and $\tilde{Z}$ includes at least $(N - A)$ zero elements where A is the number of withdrawn samples. In the last step, $\tilde{Y}$ and $\tilde{Z}$ are encoded. In order to realize these procedures, the MPEG-2 Audio NBC RM2 block diagram has been modified in noiseless coding as shown in Figure 5. More details of each step are described in the following subsections.

3.1. First step: detection of samples to be withdrawn

Local maximum samples to be withdrawn are detected by the following procedures.

1. Calculate the number of bits L_0 that are necessary to encode $\tilde{X}$ by conventional noiseless coding.
2. Let m and M be one and the maximum number of withdrawn samples.
3. Detect the frequency index P_m of withdrawn elements and their element values Q_m by the following equation.

$$P_m = \max\{i | i \leq X_{R1end+1}, |X_i| > 1\}, Q_m = X_{P_m}$$
4. Replace the value of X_{P_m} by zero and re-partition $\tilde{X}$ into R1, R2 and R3.

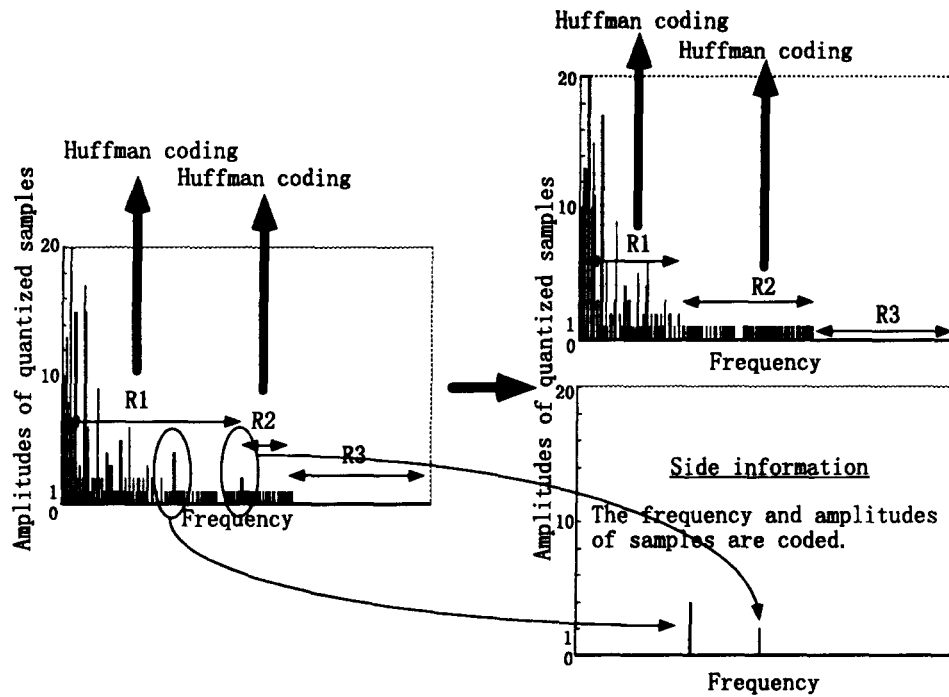


Figure 4. Proposed noiseless coding algorithm.

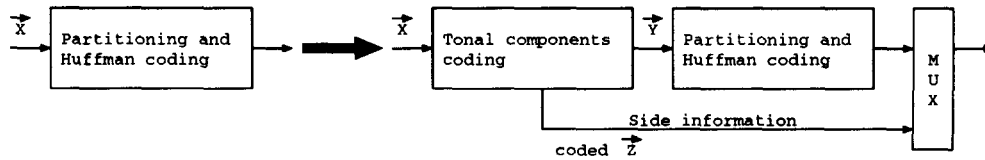


Figure 5. Differences between conventional and proposed algorithms .

5. Calculate the necessary number of bits L_m to encode the elements in R1 and R2, $P_1, P_2, \dots, P_m$ and $Q_1, Q_2, \dots, Q_m$.
6. Increase m by one.
7. If m is not greater than M then go back to Procedure 3.
8. Let A , giving $\min\{L_a | a = 0, 1, \dots, M\}$, be the number of withdrawn samples.

3.2. Second step: Withdrawal of local maxima

$\tilde{Z}$ and $\tilde{Y}$ that represent local maxima and the others are generated. Let $\tilde{Y}$ be equal to $\tilde{X}$, $\tilde{Z}$ be zero-vector whose dimension is the same as that of $\tilde{X}$. If A is greater than one, let Z_m be equal to Q_m and X_{P_m} be zero for each $m = 1, 2, \dots, a$.

3.3. Third step: Encoding of main and side information

The elements of $\tilde{Y}$ and $\tilde{Z}$ are encoded. The conventional noiseless coding algorithm is employed for $\tilde{Y}$. For $\tilde{Z}$, the frequency indexes and amplitudes of non-zero elements are encoded. Better coding efficiency is achieved by encoding $(P_1 - R1_{end} - 1), (P_2 - P_1), \dots, (P_a - P_{a-1})$ instead

of $P_1, P_2, \dots, P_a$ for the frequency indexes. As the amplitudes, $(|Q_m| - 2)$ and the sign of Q_m are encoded instead of Q_m . This is because Q_m is a withdrawn local maximum sample and $|Q_m|$ can not be less than two. The codes for local maxima ($\tilde{Z}$) and the others ($\tilde{Y}$) are multiplexed and fed into the frame packing block.

3.4. Decoding operation

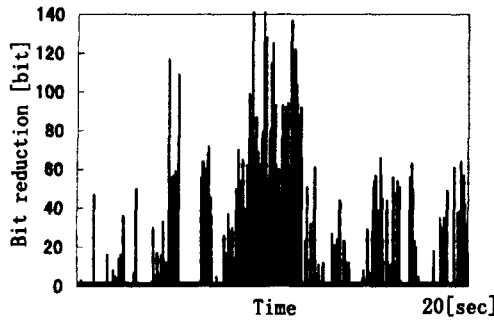
The additional decoding operation to the conventional algorithm is to reproduce the local maxima and merge them with the Huffman decoding outputs. Therefore, the complexity increase is negligible.

4. EXPERIMENTAL RESULTS

The proposed algorithm was evaluated by computer simulations for the same six sound sources as in Section 2. The bitrate was 64 kbps and M was set to four. The coding efficiency was compared with MPEG-2 Audio NBC RM2. Table 1 shows the rate of "effective-frame" where this algorithm needs less bits than MPEG-2 Audio NBC RM2, the maximum improvement of coding efficiency in a frame and average improvement in the effective-frame. As much as 11% improvement of the coding efficiency is achieved and

Table 1. Coding efficiency.

Sources	effective frame rate[%]	maximum improvement [%]	average improvement [%]
Harpsichord	55.8	5.78	1.18
Castanets	40.6	6.04	1.18
Speech	39.2	7.59	1.16
Bagpipes	30.5	5.96	0.84
Glockenspiel	30.4	11.01	2.99
Pitchpipe	50.4	3.38	0.90
Total	41.2	11.01	1.28

**Figure 6. Bit reduction per frame (Glockenspiel).**

the algorithm reduced coding bits in 41% of all the frames.

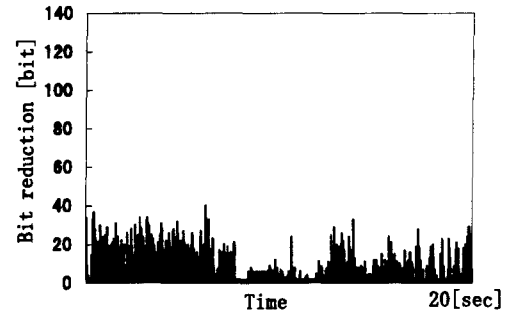
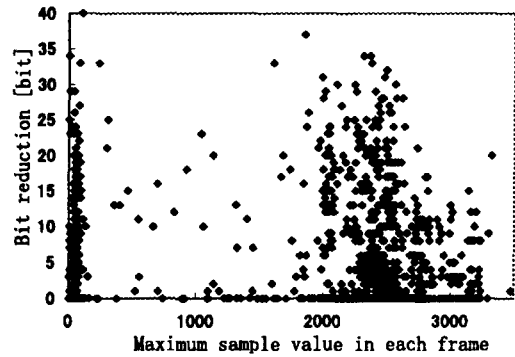
The number of reduced bits strongly depends on the encoded sound source. Figures 6 and 7 show two typical examples of the number of reduced bits per frame. For Glockenspiel in Figure 6, the large amount of bit reduction is derived from some limited areas along the X-axis. On the other hand, more averaged amount of bit reduction is derived from a larger areas along the X-axis for Pitchpipe in Figure 7.

Figure 8 shows the relation between a maximum sample value and bit reduction per frame. The maximum sample value per frame approximately represents the dynamic range of the input audio signal. As the dynamic range becomes wider, the encoding efficiency degrades. The bit reductions above 2,000 along the X-axis are derived from the frames which have a wide dynamic-range signal and are difficult to encode. The proposed algorithm improves the decoded sound quality by using these reduced bits for encoding.

5. CONCLUSION

An efficient tonal component coding algorithm has been proposed and its performance has been shown by computer simulation results. The proposed algorithm improves the coding efficiency by withdrawing local maximum samples in the frequency domain with few increase of decoding complexity. Simulation results show that this algorithm achieves as much as 11% bit reduction per frame and improves the coding efficiency in 41% of all the audio frames.

This algorithm can reduce the quantization bits in other coding algorithms by narrowing the dynamic range of the

**Figure 7. Bit reduction per frame (Pitchpipe).****Figure 8. Maximum sample value and bit reduction per frame**

input signal. It can also be applied to runlength coding for making the runlength longer. Therefore, the proposed algorithm can be incorporated into various kinds of coding algorithms as well as MPEG-2 Audio NBC.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Takao Nishitani, deputy general manager of Information Technology Research Laboratories, NEC Corporation, and Dr. Kazunori Ozawa, senior manager of Signal Processing Research Laboratory, NEC Corporation, for their continuous encouragement and valuable comments on this paper.

REFERENCES

- [1] MPEG Audio subgroup, "Call for contributions for MPEG-2 Audio NBC coding algorithms," ISO/IEC JTC1/SC29/WG11/N0866, Nov. 1994.
- [2] ISO/IEC 11172-3, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3: Audio," Aug. 1993.
- [3] ISO/IEC JTC1/SC2/WG8 MPEG-Audio, "ASPEC," No.89/205, 1989.
- [4] MPEG Audio subgroup, "Definitions and interface descriptions for NBC RM2," ISO/IEC JTC1/SC29/WG11/N1095, Jul. 1995.