SICE Annual Conference in Fukui, August 4-6, 2003
Fukui University, Japan

Audio Signal Compression via Sampled-Data Control Theory

Shinjiro Ashida, Masaaki Nagahara, Yutaka Yamamoto

Graduate School of Informatics, Kyoto University
Kyoto 606-8501, JAPAN
{sashida, nagahara, yy}@acs.i.kyoto-u.ac.jp

Abstract: The paper proposes a new digital audio coding method. This method combines subband-coding with an interpolater which is designed by using sampled-data control theory. Moreover, a new bit-allocation method is proposed, which is simpler than the conventional method, MPEG-1 Audio. A design example is presented to show the advantages of the present method.

Keywords: Sampled-data control, subband-coding, bit allocation, MPEG-1 Audio

1. Introduction

In many fields, digital systems have taken the place of analog systems due to the rapid development of digital equipment. In digital communication, audio/speech signals are transmitted or stored digitally. Because of a limit on channel capacity or memory size, digital signals must be compressed economically and efficiently.

A well-known compression method is subband-coding[7]. In subband-coding, a digital signal is divided into multiple subbands of frequency by a multirate filterbank, and then quantized by allocating bits to each subband signal with respect to its energy. By using the fact that the frequency distribution of audio signals is generally uneven, we decide how to apportion the total number of code bits available for the quantization of the subband signals to minimize the audibility of the quantization noise. MPEG-1 (Moving Picture Experts Group) Audio[2] is a kind of subband-coding, and widely used in digital communication or digital audio.

However, MPEG-1 Audio has the following problems:

- At high compression ratio, the high-frequency component is cut terribly, and hence the restored sound may be blurred.
- Coding based on psychoacoustic model[4, 5] takes much time, which may make trouble in real-time transmission in the network.

We propose a new method of audio coding which improves these problems.

In general, the original audio signal is an analog one, which has frequency component beyond the Nyquist frequency. Conventional methods, however, do not take account of the original analog characteristic. Therefore, we propose a new method for audio signal compression by using the sampled-data control theory, which can take the analog characteristic into account, and natural reconstruction is possible by our method. Moreover, we adopt a simple bit allocation method without the psychoacoustic model. A design example shows that our method is superior to the conventional method.

2. MPEG-1 Audio

We here consider a conventional method, MPEG-1 Audio. MPEG-1 Audio has three layers, Layer I, II, and III, and there is a trade-off between complexity of computation and accuracy of reconstruction. Table. 1 summarizes this trade-off. Layer III, usually called MP3, can reconstruct audio signals better than Layer I or II, while the computation is involved, and hence in real-time computation, Layer I or II is often used.

We will now discuss Layer II in detail. The system of MPEG-1 Audio Layer II is shown in Fig. 1. First, the input audio signal with sampling frequency 44.1 [kHz] is decomposed into 32 subbands of frequency by the analysis filterbank (Discrete Cosine Transform (DCT) filter-

Layer	I	II	III
complexity	simple	simple	complicated
accuracy	poor	average	good

Table. 1: Comparison of three Layers: complexity of computation and accuracy of reconstruction

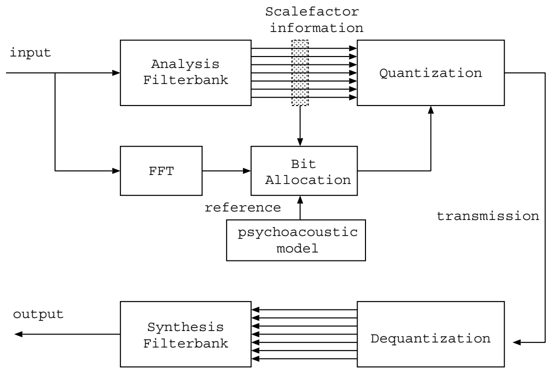

Fig. 1: Signal compression system of MPEG-1 Audio

bank). Secondly, the bit-allocation system decides the number of bits of each subband with respect to the FFT of the input signal, the scalefactor information, and psychoacoustic model. Each of the decomposed signals is then quantized and transmitted.

The transmitted signal passes through the dequantization block and the synthesis filterbank (Inverse Discrete Cosine Transform (IDCT) filterbank), and finally the signal is restored.

In this system, the complicated procedures are frequency decomposition and bit allocation. For decomposing the frequency, various efficient filterbanks are proposed[1, 6, 7], and DCT filterbank is one of them. On the other hand, the total number of code bits are allocated to minimize the audibility of the quantization noise according to psychoacoustic model. It is true that this procedure improves the quality of the restored signal, but it costs much time, which results in transmission delay. Layer I and III also have this problem.

In addition, Layer III has a lot of complicated procedures such as adaptive filterbank, nonlinear quantization, or Haffman coding. Therefore Layer III requires much more time for encoding than Layer I and II.

We next consider the reconstruction characteristic of MPEG-1 Audio Layer II. The FFT (fast Fourier transformation) of the restored signal is shown in the middle of Fig. 2, where the bit rate for transmission is 64 [kbps]. We can see that this FFT is similar to that of the original signal (shown above in Fig. 2) up to around 10 [kHz]. But when the bit rate is 48 [kbps] (shown below in Fig. 2), the FFT of the restored signal is sharply cut at the frequency around 6 [kHz]. In other words, the high frequency spectrum is not reconstructed adequately, which often makes the restored sound blurred.

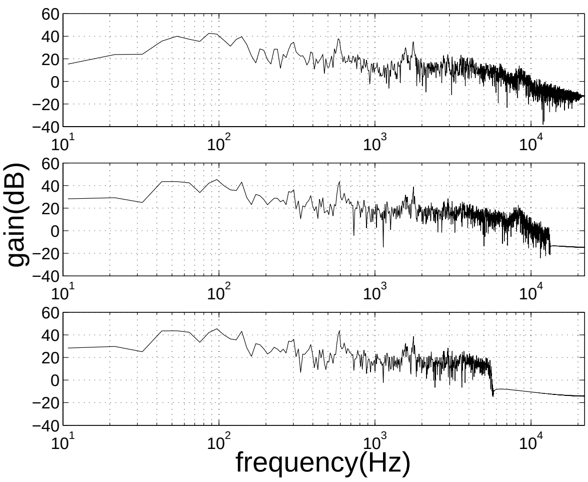

Fig. 2: FFT of signal (above: original, middle: 64 [kbps], below: 48 [kbps])

In the next section, we propose a new method which improves these problems.

3. Proposed Method

3.1 Subband-Coding

The subband-coding system we propose is shown in Fig. 3. In this figure, u is the input audio signal with sampling frequency 44.1 [kHz]. First, we downsample u by the downsampler $\downarrow 2$, and the data size is reduced by half. Then, the filterbank in the encoder (analysis filterbank) decomposes the signal into 16 subbands of frequency[1], and each of the decomposed signals is quantized by the quantizer Q. The quantized signal is then transmitted to the decoder.

In the decoder, the quantized signal passes through the synthesis filterbank to produce a restored signal $\hat{v}$. Finally, the signal $\hat{v}$ is upsampled by the upsampler $\uparrow 2$ and filtered by the digital filter $Y(z)$, and we have the restored audio signal $\hat{u}$ with sampling frequency 44.1 [kHz].

The filter $Y(z)$ is designed by sampled-data control theory [3]. Fig. 4 illustrates the error system for designing the filter $Y(z)$. We design $Y(z)$ which minimizes the H^∞ norm of the sampled-data system T_{ew} from the analog input w_c to the analog error output e_c. Our problem is formulated as follows:

[1]Each bandwidth is around 690 [Hz], which is the same as MPEG-1 Audio.

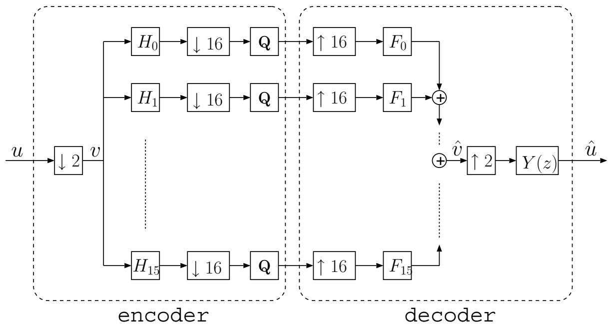

Fig. 3: Proposed audio signal compression system

Problem 1 Given a stable and strictly proper $W(s)$, stable and proper $P(s)$, delay step m and sampling time h, find $Y(z)$ which minimizes

$$\|T_{ew}\|_\infty := \sup_{w_c \in L^2[0,\infty)} \frac{\|T_{ew} w_c\|_{L^2}}{\|w_c\|_{L^2}}.$$

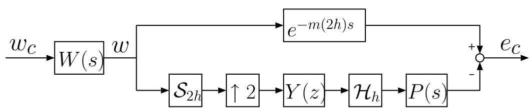

Fig. 4: Error system

3.2 Bit Allocation Algorithm

It is necessary for efficient compression to take account of the frequency characteristic when we allocate bits to each subband signal. The fundamental way is to allocate bits in proportion to the logarithm of the amplitude of each subband signal. MPEG introduces adaptive bit allocation with psychoacoustic model[4)]. However, this method requires a lot of time for computation (e.g., computation of FFT), and the delay is not desirable for real-time audio signal transmission.

Instead, we propose a simple algorithm for bit allocation. To begin with, let us consider a conventional algorithm of the logarithm bit allocation.

step1: Store M samples

$$v_i[0], v_i[1], \ldots, v_i[M-1],$$

of the i-th subband (i=1,..., 16).

step2: Define scale factor S_i as the maximum of the absolute value of the stored samples:

$$S_i := \max\{|v_i[0]|, |v_i[1]|, \ldots, |v_i[M-1]|\}. \quad (1)$$

step3: Allocate b_i bits to the i-th subband in proportion to $\log_2 S_i$ under the constraint that $\sum_{i=1}^{16} b_i = B$, where B is the total number of bits.

step4: Quantize the i-th subband data according to the quantization bit b_i.

Although this method is very simple, a problem arises; no bit can be allocated to the low or middle frequency signals when the scale factor of a high-frequency subband is as large as that of a low or middle frequency one. An example of this situation is shown in Fig. 5. The left graph in Fig. 5 shows scale factors of the subbands, and the right graph shows the number of bits allocated to each subband. Because the human auditory system is sensitive to low or middle frequency signals, the emptiness at the middle frequency results in an uncomfortable noise. In order to solve this problem, we preferentially allocate 2 or 3 bits to low or middle frequency subbands. Then we reallocate the remains of the bits in proportion to the logarithm of the amplitude. Step 3 in the algorithm mentioned above is modified as follows:

step3-1: Allocate 2 bits to the i-th subbands ($i = 1, \ldots, 9$). Then divide the scale factor S_i by 2^2, $i = 1, \ldots, 9$, and subtract 2×9 from total bit number B;

$$\begin{aligned} b_i &:= 2, \\ S_i &:= S_i / 2^{b_i}, \quad i = 1, \ldots, 9, \\ B &:= B - 2 \times 9. \end{aligned}$$

step3-2: Reallocate b_i bits to i-th subband with

$$b_i := \begin{cases} b_i + \lfloor c \log_2 S_i \rfloor, & i = 1, \ldots, 9, \\ \lfloor c \log_2 S_i \rfloor, & i = 10, \ldots, 16, \end{cases}$$

where c is a factor decided such that $\sum_{i=1}^{16} b_i = B$, and $\lfloor a \rfloor$ denotes the greatest integer equal to or smaller than a.

The effectiveness of this algorithm appears in Fig. 6, from which we can see that to the low or middle frequency bands, 2 or 3 bits are allocated adequately.

This process may be explained as a psychoacoustic model, which is simpler because it only requires comparison of the amplitude.

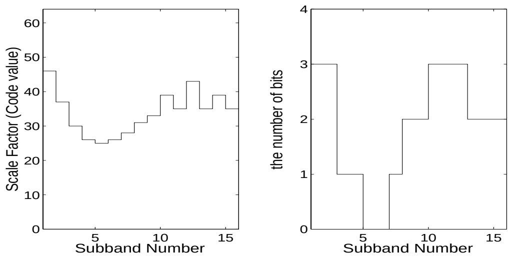

Fig. 5: Basic bit allocation

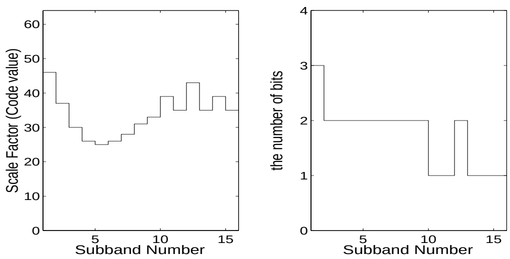

Fig. 6: Proposed bit allocation

4. Design Example

We here present a design example. We compare the proposed method mentioned above with the conventional method (MPEG-1 Audio Layer Ⅱ). The bit rate for transmission is 48 [kbps]. The FFT (fast Fourier transformation) of the original signal is shown above in Fig. 7. Then we show the FFT of restored signal by the proposed method (in the middle of Fig. 7) and by MPEG-1 Audio (below in Fig. 7). We can see that the FFT of the restored signal by MPEG-1 Audio is sharply cut at the frequency about 6 [kHz], while the FFT by our method has some frequency component to the Nyquist frequency (22.05 [kHz]). The reason is that the interpolation filter $Y(z)$ is designed with analog performance. Practically, the sound by MPEG-1 Audio is blurred, while our sound is clearer.

5. Conclusion

We present a new method for audio signal compression by using the sampled-data H^∞ control theory. The proposed method is not only simpler to compute than the conventional one, but also takes account of the analog performance.

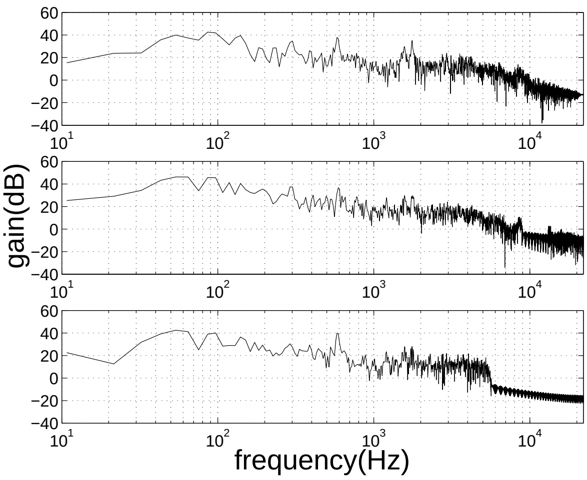

Fig. 7: FFT of signal (above: original, middle: proposed, below: MPEG-1 Audio)

References

1) R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice Hall, 1983.

2) ISO/IEC JTCI/SC29, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – IS 11172 (Part 3: Audio)," 1992.

3) M. Nagahara and Y. Yamamoto, "Sampled-data H^∞ design of interpolators," *Systems, Information and Control*, **vol. 14, No. 10**, pp. 483–489, 2001.

4) P. Noll, "MPEG Digital Audio Coding," *IEEE Signal Processing Magazine*, **vol. 14, No. 5**, pp. 59–81, September 1997.

5) D. Pan, "A Tutorial on MPEG/Audio Compression," *IEEE Trans. on Multimedia*, **vol. 2, No. 2**, pp. 60–74, 1995.

6) J. Princen, J. Johnston and A. Bradley, "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Alias Cancellation," Proc. ICASSP '87, pp. 2161–2164, 1987.

7) P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.