Sampled-Data Audio Signal Compression with Huffman Coding

Shinjiro Ashida, Hironori Kakemizu, Masaaki Nagahara and Yutaka Yamamoto

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

Abstract: This paper proposes a new digital audio coding method. It combines subband-coding with an interpolator that recovers the high frequency spectrum according to an analog audio signal model. The interpolator can be designed via sampled-data control theory. Efficient compression can be achieved by the Huffman coding at low bit-rate transmission. The proposed method is seen to possess a better frequency characteristic and a simpler processing algorithm than MPEG-1 Audio.

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

1. Introduction

With the rapid development of digital devices, analog signals such as audio signals, pictures or videos are changed into digital signals. Analog signals are digitized by sampling with a fixed sampling period, and the total data become huge. Because of limits in transmission speed or storage capacity, these digital signals must be compressed efficiently. This paper discusses audio signal compression.

Many methods are proposed for digital audio signal compression. A well-known one is based on the so-called subband-coding[12)]. In subband-coding, digital signals are subdivided into multiple subbands or frequency ranges by multirate filterbanks, and then quantized according to the energy of each subband signal. In general, the frequency energy distribution of audio signals is nonuniform, so we decide how to allocate the total number of bits to each subband while minimizing of quantization noise arising from this process. MPEG-1 (Moving Picture Experts Group) Audio[3, 8, 9)] is such a subband-coding system, widely used over the Internet or with digital devices.

MPEG-1 Audio gives an efficient compression method, but it has the following problems:

- At a low bit-rate transmission, bits are often not allocated for high frequency subbands. This results in the lack of high frequency components in restored signals.
- The quantizer of MPEG-1 Audio employs the so-called psychoacoustic model or FFT (fast Fourier transform) of original signals to determine how much bits be allocated; this is computationally expensive, and not very suitable for real-time transmission.

We propose a new method of audio compression that improves upon these problems.

We first note that the original audio signals are almost always analog. They often have frequency components beyond the Nyquist frequency, that is, the sampling frequency is usually not high enough. Conventional methods, however, regard digital signals as the original, and do not take account of the original analog characteristic. In this paper, we propose a new method for audio signal compression by using sampled-data control theory, which can incorporate the analog signal characteristic into the filter design. This leads to a simple bit allocation algorithm that does not require a psychoacoustic model. This is a continuation of the work[16, 1)]. In addition to the basic framework developed there, we also adopt the Huffman coding[11)] for efficient compression at low bit-rate transmission. A design example shows that our method is superior to the conventional method.

2. MPEG-1 Audio

In this section we review the conventional MPEG-1 Audio. MPEG-1 Audio has three Layers: Layer I, Ⅱ and Ⅲ, and they are used depending on the situation. Layer Ⅲ, called MP3 for short, is superior to the two above, and is used more often over the Internet or in many digital audio devices, where the processing time is not a crucial issue. On the other hand, because of its heavy processing, lighter Layer I and Ⅱ are often preferred in real-time applications. We here discuss Layer Ⅱ in detail, because it has relatively simple procedures and reasonable reconstruction performance.

The encoding/decoding system of MPEG-1 Layer Ⅱ is shown in Figure 2. Observe that the inputs are assumed to be discrete-time signals. In the encoder, the input signal with sampling frequency 44.1 [kHz] is decomposed into 32 subbands by an analysis filter bank, and then each subband signal is quantized according to a suitable bit allocation rule, using the FFT of the input signal and psychoacoustic model. Quantized subband signals are then transmitted to the decoder. In

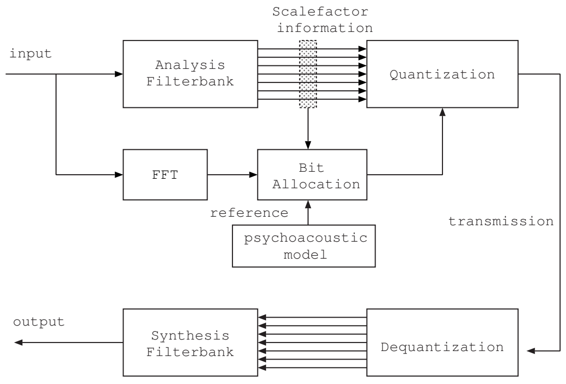

Figure 1: Signal compression system of MPEG-1 Audio

the decoder, transmitted signals pass through the dequantizer and an synthesis filterbank, and the signal is restored.

In this system, the crucial step is the decomposition to subbands combined with a bit allocation algorithm. For this purpose many filterbanks are proposed Many filterbanks are proposed for frequency[2, 10, 12]. A convenient tool for this is DCT (discrete cosine transform). The quantization bits number for each subband are determined to minimize quantization noise according to the FFT of the input signal and psychoacoustic model. On the other hand, computing FFT is computationally heavy, and results in a transmission delay. Layer I and Ⅲ also share this problem. In addition, Layer Ⅲ employs an adaptive filterbank, nonlinear quantization and the Huffman coding, so it requires more time than Layer Ⅱ.

We next consider the characteristic of restored signals by MPEG-1 Layer Ⅱ. The frequency characteristic of the original digital signal is shown in Figure 2. There are a fair amount of frequency components up to the Nyquist frequency (22.05 [kHz]). The middle figure of Figure 2 shows the result of compression with bitrate 64 [kbps/channel] by MPEG-1 Layer Ⅱ. The signal is well reconstructed up to 15 [kHz], but anything above this is truncated sharply. No bits are allocated beyond this frequency, to allow more bits to low and middle frequency. In the case of low bit-rate transmissions, this problem can be worse. When the bit-rate is 48 [kbps/channel] (shown at the bottom of Figure 2), the FFT of the restored signal is sharply truncated at around 6 [kHz]. Practically, the sound is dull due to the lack of the high frequency range.

3. Proposed Method

3.1 Subband-Coding

The subband-coding system we propose is shown in Figure 3. In this figure, u is the input audio signal with sampling frequency 44.1 [kHz]. In the encoder, we downsample u by the downsampler $\downarrow 2$, and obtain the signal v with sampling frequency 22.05 [kHz] (the data

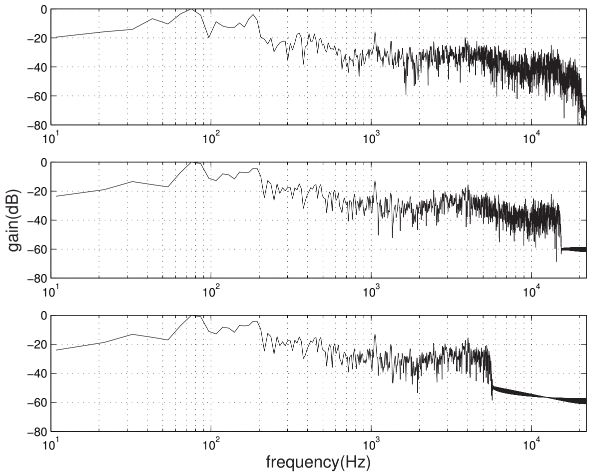

Figure 2: FFT of signal (above: original, middle: 64 [kbps/channel], below: 48 [kbps/channel]

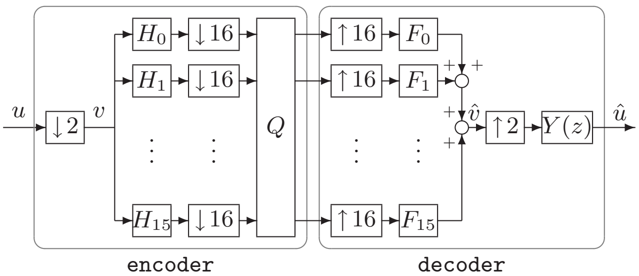

Figure 3: Proposed audio signal compression system

size is reduced to half). Then the analysis filterbank decomposes the signal into 16 subbands of frequency 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 signal is dequantized and passes through the synthesis filterbank to produce a restored signal $\hat{v}$ with sampling frequency 22.05 [kHz]. This signal $\hat{v}$ is then 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].

We also use DCT for implementing the filterbank. The number of subbands is however only 16—half of that in MPEG 1, as a result of the downsampler $\downarrow 2$. The crucial step is to recover the lost high frequency via the interpolator $Y(z)(\uparrow 2)$, and this results in saving the allocation bits. This works as follows: The downsampler $\downarrow 2$ discards the high frequency components beyond the half of Nyquist frequency (11.025 [kHz]). This is the signal to be reconstructed after the synthesis filterbank. The upsampler $\uparrow 2$ creates the imaging components beyond 11.025 [kHz] as the mirror image of the lower frequency components. The role of the digital filter $Y(z)$ is to recover the original signal including the high frequency components, using the imaging components. To this end, sampled-data control theory gives an ideal platform in that it optimizes the intersample behavior, which is nothing but the high frequency compo-

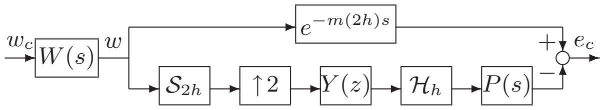

Figure 4: Error system

nents beyond the Nyquist frequency[5, 6, 7]. The crucial ingredient here is the analog frequency characteristic of the original signal, which is best utilized by sampled-data control theory, but not yet much appreciated in the signal processing literature.

Figure 4 illustrates the error system for designing the digital filter $Y(z)$. The analog input signal w_c is bandlimited by $W(s)$ and sampled by the sampler $\mathcal{S}_{2h}$. The signal is then processed by the interpolator $Y(z)(\uparrow 2)$. It passes through the zero-order hold $\mathcal{H}_h$ and the analog filter $P(s)$, and then we obtain the restored signal. The important point here is to take the analog error signal e_c as the difference between the restored signal and the delayed input signal with the m step delay, and then minimize these errors over all possible analog inputs w_c.

To this end, $Y(z)$ is designed via sampled-data control theory[7]. We design $Y(z)$ that 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 design problem is thus formulated as follows:

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*

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

A difficulty in Problem 1 is that it contains a continuous time-delay, which makes the problem infinite-dimensional. Another is that it is time-varying due to the upsampler $\uparrow 2$. By using the discrete-time lifting[13], however, we can reduce this problem to a single-rate problem. We can further reduce this to a finite-dimensional H^∞ problem[5], but for numerically tractability, we rather resort to the fast sample/hold approximation[4]. (For its convergence, see [14, 15].) By this method, Problem 1 is converted into a discrete-time problem approximately. We can then design the digital filter $Y(z)$ by using Matlab.

3.2 Bit Allocation Algorithm

The role of the quantizer Q is to assign a suitable number of bits to each subband according to a certain predetermined rule. The simplest is to allocate bits in proportion to the logarithm of the amplitude of each subband signal. This proceeds as follows:

step 1: Store M samples

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

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

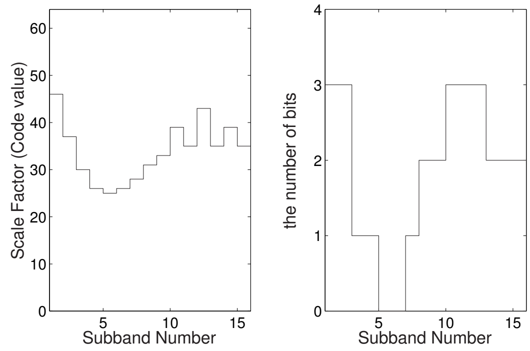

Figure 5: Basic bit allocation

step 2: Define the *scale factor* S_i of the i-th subband as the maximum of the absolute value of the stored samples:

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

step 3: Determine the quantizing bit number b_i of 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 which is defined by the bit-rate.

step 4: Quantize the i-th subband data linearly according to the quantizing bit number b_i.

In many cases, this simple method works quite well. However, a problem arises when scale factors of high frequency subbands are larger than those of low or middle frequency ones. Figure 5 shows this problem: no bit is allocated to middle frequency subbands. This raises more audible and intolerable noise due to the human auditory characteristic which is more sensitive in the low or middle frequency ranges. To avoid this, MPEG-1 employs a psychoacoustic model and the FFT of the input signal to calculate quantizing bits. On the other hand, this process is computationally burdensome.

We here propose a simple algorithm without the FFT or a psychoacoustic model. We start by allocating 2 bits to low and middle frequency subbands (from the lowest subband to the prespecified n-th subband). Then we allocate the remaining bits in proportion to the logarithm of the amplitude. Thus **Step 3** in the algorithm mentioned above is modified as follows:

step 3-1. Allocate 2 bits to the i-th subbands ($i = 1, \ldots, n$). Then divide the scale factor S_i by 2^2, $i = 1, \ldots, n$, and subtract $2n$ from total bit number B;

$$S_i' := \begin{cases} S_i/2^2, & i = 1, \ldots, n, \\ S_i, & i = n+1, \ldots, 16, \end{cases}$$

$$B' := B - 2n.$$

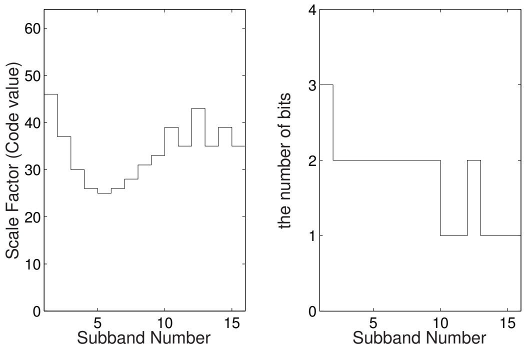

Figure 6: Proposed bit allocation

Table 1: The probability of each symbol generated in the quantizer (subbands allocated 2 bits)

symbol	0	1	2	3
probability	0.0873	0.4122	0.4132	0.0873

step 3-2. Reallocate b_i bits to i-th subband with

$$b_i := \begin{cases} 2 + \lfloor c \log_2 S'_i \rfloor, & i = 1, \ldots, n, \\ \lfloor c \log_2 S'_i \rfloor, & i = n+1, \ldots, 16, \end{cases}$$

where c is a factor determined 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 Figure 6, from which we can see that bits for quantization are allocated adequately to the low and middle frequency bands. The process may be viewed as a simple psychoacoustic model, because it requires only a simple comparison of the scale factors.

3.3 Huffman Coding

The probability distribution of each symbol generated in the quantizer is generally uneven. An example is shown in Table 1 (subbands allocated 2 bits). Since the gap between each probability will be wide if the number of allocated bits is small, efficient compression is possible by using the Huffman coding[11)] particularly at a low bit-rate transmission (each subband is expected to have few bits to be quantized).

In the quantizer of the proposed method, low or middle (or high) frequency subbands are allocated 2 bits preferentially, and we have many subbands allocated with 2 bits (see Figure 6). Because of this, and for simplicity, we use the Huffman coding only for subbands with 2 bits in series. The Huffman code of each symbol is shown in Table 2.

4. Design Example

We here compare the reconstruction performance of our proposed method and that of MPEG-1 Layer Ⅱ to show

Table 2: Huffman code of each symbol (2bit-allocated band)

symbol	0	1	2	3
Huffman code	001	01	1	000

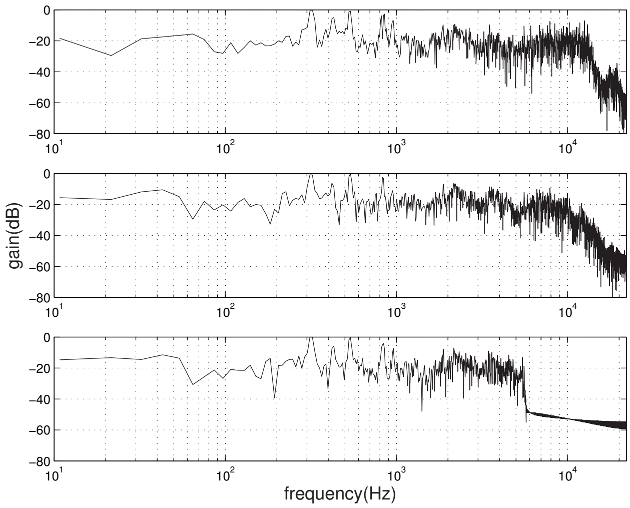

Figure 7: Comparison of performance with MP2 (above: original, middle: proposed, below: MPEG-1 Audio Layer Ⅱ)

the effectiveness of our method. The bit-rate of compression is 48 [kbps/channel]. Design parameters of the digital filter $Y(z)$ in our method are as follows:

$$W(s) = \frac{1}{(Ts+1)(10Ts+1)}, \quad T = 0.70187, \qquad (2)$$

$$P(s) = 1, \quad h = 1, \quad m = 2.$$

Here $W(s)$ simulates the upper envelope of the frequency energy distribution of a wide range musical source (for example, orchestral musics): flat response up to 1 [kHz] and decays by 20 [dB/dec] beyond there and then decays again beyond 10 [kHz].

The FFT plots of the original audio signal, the restored signal by the proposed method and the result by MPEG-1 Layer Ⅱ are shown in Figure 7. The result by MPEG-1 Audio Layer Ⅱ shows no higher frequency components beyond 6 [kHz], while the present method shows a natural frequency decay up to the Nyquist frequency (22.05 [kHz]). In actual comparison with listening, the sound by MPEG-1 Audio Layer Ⅱ lacks sharpness while the present method gives much crisp sounds.

We next compare the result of the proposed method with that by MPEG-1 Layer Ⅲ (using joint stereo encoding). Design parameters of $Y(z)$ are the same as (2), and the bit-rate for transmission is also 48 [kbps/channel]. The FFT of the reconstructed audio signal by our method and MPEG-1 Layer Ⅲare shown in Figure 8. Two results both exhibit similar high fre-

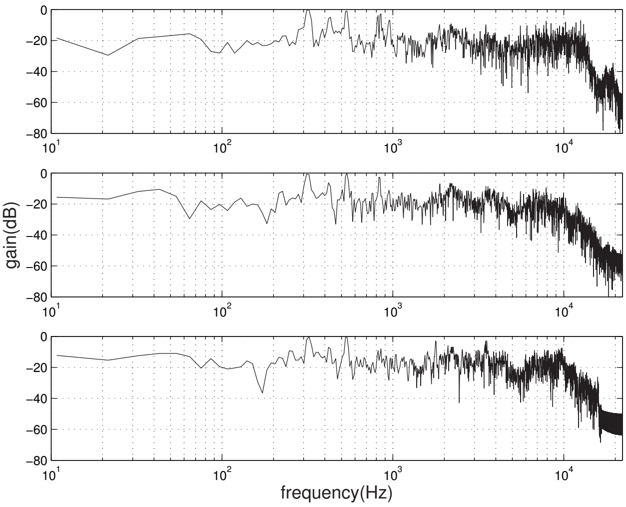

Figure 8: Comparison of performance with MP3 (above: original, middle: proposed, below: MPEG-1 Audio Layer Ⅲ using joint stereo encoding)

Table 3: The effect of Huffman coding on the bit-rate

method	bit-rate [kbps/channel]
with Huffman coding	47.3
without Huffman coding	56.0

quency response. An advantage of the present method is its simplicity in processing.

Table 3 shows the effect of Huffman coding. In this case, the bit-rate is reduced about 8 [kbps/channel] by using Huffman coding only to subbands allocated 2 bits. This is because many subbands are allocated 2 bits in case of low bit-rate transmission. More efficient compression can be done if we adopt Huffman coding to every subbands.

5. Conclusion

We have proposed a new method for audio signal compression using sampled-data H^∞ control theory. We have also employed a simple bit allocation algorithm and the Huffman coding, and this led to an efficient compression algorithm. The proposed method is not only simpler than conventional ones, but also successfully restores high frequency components—advantage of sampled-data control theory.

References

[1] S. Ashida, M. Nagahara, Y. Yamamoto, "Audio signal compression via sampled-data control theory," *Proc. SICE Annual Conference*, pp. 1182–1185, 2003.

[2] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice Hall, 1983.

[3] 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.

[4] J. P. Keller and B. D. O. Anderson, "A new approach to the discretization of continuous-time controllers," *IEEE Trans. on Automatic Control*, **vol. 37**, **No. 2**, pp. 214–223, 1992.

[5] P. P. Khargonekar and Y. Yamamoto, "Delayed signal reconstruction using sampled-data control," *Proc. of 35th Conf. on Decision and Control*, pp. 1259–1263, 1996.

[6] M. Nagahara and Y. Yamamoto, "A new design for sample-rate converters," *Proc. of 39th Conf. on Decision and Control*, pp. 4296–4301, 2000.

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

[8] P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Magazine*, **vol. 14**, **No. 5**, pp. 59–81, 1997.

[9] D. Pan, "A tutorial on MPEG/Audio compression," *IEEE Trans. on Multimedia*, **vol. 2**, **No. 2**, pp. 60–74, 1995.

[10] 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.

[11] J. G. Proakis, *Digital Communications*, McGraw-Hill, 1995.

[12] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[13] Y. Yamamoto, H. Fujioka and P. P. Khargonekar, "Signal reconstruction via sampled-data control with multirate filter banks," *Proc. 36th Conf. on Decision and Control*, pp. 3395–3400, 1997.

[14] Y. Yamamoto, A. G. Madievski and B. D. O. Anderson, "Approximation of frequency response for sampled-data control systems," *Automatica*, **35**, pp. 729–734, 1999.

[15] Y. Yamamoto, B. D. O. Anderson and M. Nagahara, "Approximating sampled-data systems with applications to digital redesign," *Proc. of 41st Conf. on Decision and Control*, pp. 3724–3729, 2002.

[16] Y. Yamamoto et al., *Japanese patent application number H11-302231, 2000-174799, and 2003-285330.*