INTERCHANNEL DEPENDENCY ANALYSIS OF BIOMEDICAL SIGNALS FOR EFFICIENT LOSSLESS COMPRESSION BY MPEG-4 ALS

Yutaka Kamamoto, Noboru Harada, Takehiro Moriya

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, JAPAN

ABSTRACT

This paper describes a new search algorithm that quickly finds interchannel relationships between a coding channel and a reference channel in the multichannel coding tool of the MPEG-4 Audio Lossless Coding (ALS) international standard. The algorithm has tree structure and can reduce data size with significantly smaller computation load than that of the conventional one. The devised method is based on a restricted greedy algorithm. It chooses the most efficient branch which does not make any loops in the existing path. The results of comprehensive evaluations show that this method maintains the compression performance (compression to around 1/3) and performs 1000 times as fast as the conventional method for the 512-channel magnetoencephalography signals. This algorithm enables practical lossless compression of biomedical data by the ALS, and at the same time, opens the way to a new multichannel analysis tool that may be used for purposes other than compression. The continual maintenance of this standard will make it possible to perfectly reconstruct encoded files even 100 years from now.

***Index Terms*—** Data compression, Audio coding, Magnetoencephalography, Electrocardiography

1. INTRODUCTION

A new international standard for lossless audio coding, ISO/IEC MPEG-4 Audio Lossless Coding (ALS), was published in March 2006 [1]. This standard can efficiently compress the time-series signals (such as speech, audio, and seismic signals) without any losses, and in order to improve compression performance not only for stereo signals but also for multichannel signals (such as biomedical and surround audio signals), it supports a Multi-Channel Coding (MCC) scheme that exploits interchannel correlation. In the MCC mode, adaptively weighted subtraction is carried out between the linear prediction residual signals of the coding channels and those of the reference channel. The combination of the reference channel and the coding channel affects the compression performance. A reasonable method for searching the combination has been introduced [2] and implemented in the reference software of the MPEG-4 ALS [3].

We have introduced the tradeoffs between compression performance and processing complexity with concern for stereo audio signals and enhanced processing speed [4]. For two-channel signals, it is easy to find the optimum combination of channel dependency (described in Sec.3). Additionally, we have pointed out that ALS will be used for not only audio signals but also biomedical signals. The conventional searching tool is practical for encoding 5.1-channel surround audio or 8-channel microphone array signals. However, it does not have sufficient performance when the number of channels of signals exceeds 100, such as 512-channel magnetoencephalography (MEG) signals, due to the rapid growth of complexity depending on the number of channels. So we need to use sub-optimal search within a restricted number of channels (i.e., dividing the large number of channels to search into groups of small number of channels) to reduce the computational cost in practical use by the sacrifice of compression performance.

In order to make interchannel dependency analysis for huge number of channels within a reasonable complexity, we have devised a new quick search algorithm. After an overview of the MPEG-4 ALS, the devised algorithm is explained, and then experimental results are shown.

2. OVERVIEW OF THE MPEG-4 ALS

2.1. Linear Predictive Coding

The MPEG-4 ALS is based on time-domain linear prediction and entropy coding. A P-th order linear prediction analysis provides estimates of the prediction parameters that minimize the prediction residual between the input value and the value predicted from a given number of past samples as shown in the following well-known equation

$$e(n) = x(n) - \hat{x}(n) = x(n) - \left(\sum_{k=1}^{P} a_k \cdot x(n-k) \right).$$

where $e(n)$ is the prediction residual signal and $x(n)$ is the input signal. The integer value of the prediction residual signal and the quantized PARCOR coefficients obtained from the prediction parameters are transmitted to the decoder. The decoder has a recursive filter that can reconstruct the original waveform losslessly from the transmitted bitstream.

Actually, the residual signal has smaller amplitude than the input signal, and the amplitude can be compressed by means of entropy coding using Golomb-Rice code or block Gilbert-Moore code. Obviously, the smaller the amplitude is, the shorter the code length becomes. Therefore, the minimum code length can approximately be obtained by minimizing the energy of the residual signals of every frame.

2.2. MCC

To improve compression performance for multichannel signals, adaptive subtraction from a reference channel with weighting factors is applied to the residual signal. There are three modes for each channel and each frame; an independent coding mode, a second mode with three taps, and a third one with six taps (i.e., three delayed taps added to the three in the second mode). At least one channel has to be encoded in the independent coding mode to decode all the channels losslessly, and here we call the second and the third mode "subtracted mode". For the three-tap mode, the following operation is performed to reduce the amplitude of the residual signal;

$$\tilde{e}_c(n) = e_c(n) - \left(\sum_{j=-1}^{1} \gamma_j \cdot e_r(n+j) \right),$$

where $e_r(n)$ is the residual of the reference channel, $e_c(n)$ is that of the coding channel, and $\tilde{e}_c(n)$ is the subtracted signal of the coding channel. The reference channel is found by searching among available channels. The subtracted residual signal and the reference channel index are coded together with the multi-tap gain parameters $\gamma_j (j = -1, 0, 1)$, which can be calculated by minimizing the energy of the subtracted residual signal. The decoder reconstructs the original residual signal simply by applying the reverse operation, $e_c(n) = \tilde{e}_c(n) + \left(\sum_{j=-1}^{1} \gamma_j \cdot e_r(n+j) \right)$. For the six-tap mode, a similar operation is carried out using $\tilde{e}_c(n) = e_c(n) - \left(\sum_{j=-1}^{1} \left(\gamma_j \cdot e_r(n+j) + \gamma_{\tau+j} \cdot e_r(n+\tau+j) \right) \right)$, where the time difference τ can be estimated by cross-correlation between the coding channel and the reference channel. In the decoder, the original residual signal can be reconstructed by the inverse operation. Anyway, the reconstructed residual signal is used for LPC synthesis.

3. INTERCHANNEL DEPENDENCY ANALYSIS

3.1. Principle

In the conventional method that is implemented in the reference software of the ALS, the coding channel is determined if the candidate of the reference channel has already connected to the channel that is encoded in the independent coding mode. In other words, the root-to-leaf approach is applied. In contrast, the new algorithm is based on a tree structure and the leaf-to-root approach. For the following reason, we approximate this problem by using a minimum spanning tree with a directed graph.

To simplify the explanation without losing generality, we assume one-tap subtraction

$$\tilde{e}_c(n) = e_c(n) - \gamma \cdot e_r(n) \tag{1}$$

where, the code length is approximated by the energy as mentioned in the previous section, so we should minimize the subtracted energy $\tilde{E}_c = \sum_{n=1}^{N} (\tilde{e}_c(n))^2$, where N denotes samples per frame. Therefore, γ in eq.(1) is calculated as

$$\gamma = \frac{\sum_{n=1}^{N} e_c(n) \cdot e_r(n)}{\sum_{n=1}^{N} e_r(n) \cdot e_r(n)}. \tag{2}$$

From eqs.(1) and (2), $\tilde{E}_c$ can be transcribed as

$$\begin{aligned} \tilde{E}_c &= \sum_{n=1}^{N} (\tilde{e}_c(n))^2 = \sum_{n=1}^{N} (e_c(n) - \gamma \cdot e_r(n))^2 \\ &= E_c - \frac{(C_{c,r})^2}{E_r}. \end{aligned}$$

where $E_c = \sum_{n=1}^{N} (e_c(n))^2$, $E_r = \sum_{n=1}^{N} (e_r(n))^2$, and $C_{c,r} = \sum_{n=1}^{N} e_c(n) \cdot e_r(n) = C_{r,c}$.

For stereo signals, for example, there are two cases of the total energy with the MCC process. When the left (1) channel is the reference channel, a total energy E_{1ref}^{tot} is calculated as $E_{1ref}^{tot} = E_1 + \tilde{E}_2 = E_1 + \left(E_2 - \frac{(C_{2,1})^2}{E_1} \right) = E_1 + E_2 - \frac{(C_{1,2})^2}{E_1}$, and another energy (when the right (2) channel is the reference) becomes $E_{2ref}^{tot} = E_2 + \tilde{E}_1 = E_1 + E_2 - \frac{(C_{1,2})^2}{E_2}$. Obviously, we should choose the channel that has smaller energy as the reference because we want to minimize the total energy. For stereo signals, it is easy to find the optimum combination, but doing so becomes much more complex as the number of channels increase.

In general, let C be the number of channels, let I be the channel that is coded in the independent coding mode, and let $\rho(j)$ the reference channel for the coding channel j, then the total energy is denoted as

$$\begin{aligned} E_{Iref}^{tot} &= \sum_{k=1}^{C} E_k - \left(\sum_{j=1}^{I-1} \frac{(C_{j,\rho(j)})^2}{E_{\rho(j)}} + \sum_{j=I+1}^{C} \frac{(C_{j,\rho(j)})^2}{E_{\rho(j)}} \right) \\ &= \sum_{k=1}^{C} E_k - G, \end{aligned}$$

so we should maximize G to improve compression ratio. The optimum relationships of the channel dependency could be solved by Kruskal's algorithm if it were undirected graph [5]; however, unfortunately, it is directed one. Therefore our new algorithm is based on restricted greedy algorithm.

3.2. Devised algorithm

To speed up the search for reasonable interchannel dependency without degrading compression performance, the following algorithm (see Fig.1) has been implemented in the codec software that is developed by NTT and is compliant with the ALS. The calculations of the energies and cross-correlations are same as in the conventional one (i.e., the criterion is same); there is just difference in the search module. The devised algorithm chooses sequentially from the effective branch without any loops. Calculating $\frac{(C_{j,\rho(j)})^2}{E_{\rho(j)}}$ in all cases and sort them into decrease order, then checking the i-th element ($i = 1, 2, \cdots$), and the channel $\rho(j)$ becomes the reference channel of channel j if the channel j has not encoded or the connection from j to $\rho(j)$ dose not make any loops.

For example, in the three-channel case, there are three energies ($E_1 = 60$, $E_2 = 48$, $E_3 = 20$) and three cross-correlations ($(C_{1,2})^2 = 4$, $(C_{1,3})^2 = 1$, $(C_{2,3})^2 = 1/10$). Therefore, we obtain a vector $\mathbf{G}$=($\frac{(C_{2,1})^2}{E_1}$, $\frac{(C_{1,2})^2}{E_2}$, $\frac{(C_{3,1})^2}{E_1}$, $\frac{(C_{1,3})^2}{E_3}$, $\frac{(C_{3,2})^2}{E_2}$, $\frac{(C_{2,3})^2}{E_3}$) = ($\frac{1}{15}$, $\frac{1}{12}$, $\frac{1}{60}$, $\frac{1}{20}$, $\frac{1}{480}$, $\frac{1}{200}$), and the vector sorted into decreasing order becomes $\mathbf{G}$=($\frac{(C_{1,2})^2}{E_2}$, $\frac{(C_{1,2})^2}{E_1}$, $\frac{(C_{1,3})^2}{E_3}$, $\frac{(C_{1,3})^2}{E_1}$, $\frac{(C_{2,3})^2}{E_3}$, $\frac{(C_{2,3})^2}{E_2}$). Then, the first element is checked. Channel 1 has not been encoded yet, and the connection from 1 to 2 does not make a loop, so the channel 1 is encoded in the subtracted mode and the reference channel of channel 1 is the channel 2. At the second element, channel 2 has not been encoded yet, but the connection from 2 to 1 makes a loop because the connection in the opposite direction has been encoded. Next, at the third element, the channel 1 has already been encoded. And, at the fourth element, channel 3 has not been encoded, and the connection from 3 to 1 does not make a loop; therefore, channel 3, whose reference channel is 1, is encoded. Finally, the leftover channel 2 is encoded in the independent mode.

This greedy algorithm is not optimum, however the channel dependency obtained by the new method keeps decodable channel relationships in the international standard. In addition, it has same approach as the conventional one to perform a partial search within a subset of channels instead of searching for all channels.

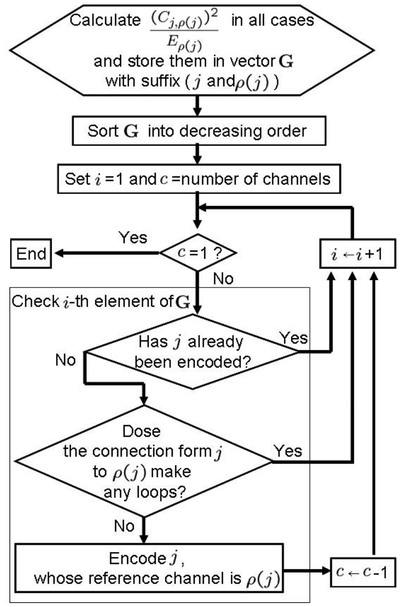

Fig. 1. Flowchart of the devised algorithm.

4. EXPERIMENTAL EVALUATION

4.1. Condition

To evaluate our methods, encoding and decoding experiments were carried out for MEG and electrocardiography (ECG) signals; these are donated by Prof. Takeda from the Univ. of Tokyo and downloaded from PhysioNet [6], respectively. The parameters of these input materials are shown in Table 1. The obtained encoding and decoding times are the total processing times for all input signals measured by "timeit.exe" on a Microsoft Windows Server 2003 with an AMD Opteron processor operating at 2.4 GHz with 2 MB of memory. We defined the compression ratio (the smaller the better) in percent and the processing-time ratio (the smaller the faster; 50% means double speed) in percent as follow

$$\text{Compression ratio (\%)} = \frac{\text{Compressed file size}}{\text{Original file size}} \times 100$$

$$\text{Processing-time ratio (\%)} = \frac{\text{Total processing time}}{\text{Total duration time}} \times 100.$$

The maximum prediction order was set to 31, and the residual signals were coded by the Golomb-Rice code.

4.2. Results

The results of encoding for MEG signals are shown in Figs.2—3 and for ECG signals in Table 2. At the decoder, there were no significant differences between the conventional and the new method in processing time because the devised algorithm does not harm the ALS decoder, and obviously we confirmed that all bitstreams were reconstructed perfectly and compliant with the standard. For reference, the well-known ZIP can

Table 1. Input-signal parameters

Data	MEG	ECG
Number of channels	512	12
Sampling rate	2 kHz	1 kHz
Word length	16 bit	16 bit
Total file size	163.8 Mbyte	151.8 Mbyte
Total duration time	80.0 sec.	6323.2 sec.
Number of files	4	57

compress the MEG and ECG files to only 54.9% and 75.2% of compression ratios, respectively.

Figure 2 shows the tradeoffs between encoding time and compression ratio, where the five symbols on each curve correspond to systems with different searching subset channels of 4, 8, 16, 32, and 64, from top to bottom. Compressed file size becomes smaller as processing time increases in both methods. At the same encoding time, e.g. 35% of processing-time ratio, the developed method (solid line and circle dots) yields a 0.35% better compression ratio than the conventional one (dashed line and square dots). At the same compression ratio, e.g. 35.6%, the new algorithm is five times as fast as the original. Figure 3 compares the searching space of channels and processing-time ratio. The conventional method needs more time than real time; in other words the processing-time ratio over 100%, when the number of channels to search is larger than 64. In contrast, the new method works about in real time even if the search is performed for all 512-channels. Our method can encode 1000 times as fast as the original one.

The results for ECG signals are shown in Table 2. The devised method processes just a little faster than the conventional one, similar to the result in Fig.3 for 8- or 16-channels. In addition, ECG data can be encoded and decoded more than 100 times as fast as real time.

In summary, the ALS can efficiently compress biomedical signals to around 1/3 losslessly. And the codec software with the new method has enough speed for practical use.

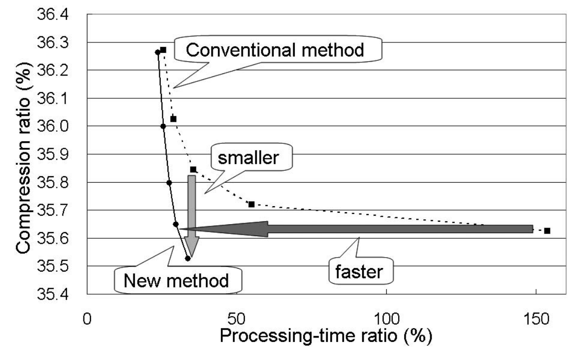

Fig. 2. Encoding time and compression ratio for MEG signals (left is faster, bottom is smaller).

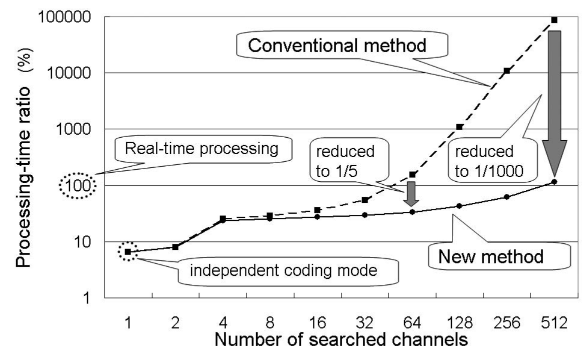

Fig. 3. Searching space of channels and encoding time for MEG signals.

5. CONCLUSION

A new quick searching algorithm for interchannel relationships between a coding channel and a reference channel has been devised, with the goal of enhancing the compression performance of the MPEG-4 ALS. This algorithm is compliant with the international standard and it can reduce data size with significantly smaller computation load than that of the conventional one. We confirmed that this method maintains the compression performance and encodes 1000 times as fast as the conventional one for 512-channel MEG signals.

The MPEG-4 ALS can practically compress the biomedical data utilizing the new algorithm, and the tree structure of the interchannel dependency may be used for purposes other than compression. This international standard will continue to be maintained so that compressed files can be decoded losslessly even 100 years from now.

Table 2. Results for ECG data

ECG signals	Compression ratio	Encoding-time ratio	Decoding-time ratio
Conventional	31.1%	0.36%	0.09%
Devised	31.1%	0.30%	0.09%

6. REFERENCES

[1] ISO/IEC 14496-3:2005/AMD.2:2006, *Information technology — Coding of audio visual objects — Part 3 Audio, 3rd Ed. AMENDMENT 2: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions*, March 2006.

[2] Y. Kamamoto, T. Moriya, T. Nishimoto, and S. Sagayama, "Lossless compression of multi-channel signals using inter-channel correlation," *IPSJ Journal (in Japanese)*, vol. 46, pp. 1118–1128, May 2005.

[3] http://www.nue.tu-berlin.de/forschung/projekte/lossless/mp4als.html

[4] T. Moriya, N. Hadara, and Y. Kamamoto, "Performance-complexity tradeoffs of the mpeg-4 als lossless coding standard," in *Proc. IEEE 40th Asilomar conference, WA7a-4*, October 2006.

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms (second edition)*, The MIT Press, 2001.

[6] http://www.physionet.org