SEGMENTED DIMENSIONALITY REDUCTION CODING ON FREQUENCY DOMAIN SIGNAL

MINJE KIM[1], SEUNGKWON BEACK[2], TAEJIN LEE[3], DAEYOUNG JANG[4] KYEONGOK KANG[5]

Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea
[1]mkim@etri.re.kr
[2]skbeack@etri.re.kr
[3]tjlee@etri.re.kr
[4]dyjang@etri.re.kr
[5]kokang@etri.re.kr

This paper proposes schemes of compressing frequency domain signals using dimensionality reduction methods. Dimensionality reduction methods which work on a input matrix result in two dimension-reduced output matrices with usually high compression ratio since they not only allow us to represent the input matrix with smaller amount of data, but exploit intrinsic information of the original data. Frequency domain signals can be seen as a (number of frequency bands) ×(number of total frames) input matrix of dimensionality reduction methods. However, in this case, real-time encoding is not possible and encoder-side delay is inevitable which amounts to the length of whole input signal. To minimize the delay this paper proposes a coding scheme which conducts multiple dimensionality reduction on segments of input data frames serially.

INTRODUCTION

Digitalized representation of audio signals which has been accelerated since the development of CD (Compact Disk) now accomplished both technological and cultural innovation through wide use of MPEG-1 Layer III [1], well known by its extension “MP3”. The MP3 audio signal coding method let us manage the compressed audio signals regardless of physical storage, but, the research on more effective ways of digitalizing audio signals is still in progress. Recent standardization issues, such as USAC (Unified Audio and Speech Coding) in MPEG audio group, aims at compensating the drawbacks of state-of-art coding technologies, HE-AAC v2 (High Efficiency Advanced Audio Coding Version 2) and AMR-WB+ (Adaptive Multi Rate–WideBand plus) whose coding efficiencies decrease in the case of low-bitrate speech-like signals and high-bitrate music-like signals, respectively.

One of the most important goals of audio signal coding is to represent input signals with the smallest data rate while maintaining the acoustic quality of the original ones. Reducing data rate can be helpful for transporting digitalized signals through broadcasting or communication networks where both quality and quantity of information is critical. In addition, the need for more effective compression of signals is growing as broadcasting targets are expanding to mobile users and multichannel environments are commonly adapted to end users.

Audio coding schemes are based on two widely believed principles: using statistical characteristics of signals and perceptual coding which relies on psychoacoustic features.

Dimensionality reduction methods have usually been used for expressing data as more efficient forms. There are method-specific characteristics of dimension-reduced data which can enhance the performance of systems, for example, pattern recognition systems, if they use the dimension-reduced data instead of original ones. Our work tries to eliminate intrinsic redundancy of input signals by using dimensionality reduction methods similarly to the traditional ways of using dimensionality reduction methods except that we regard the dimension-reduced input signals as another form of bitrate-reduced ones. We finally propose an effective representation of frequency domain audio signals.

1 FREQUENCY DOMAIN REPRESENTATION OF AUDIO SIGNALS

Most audio coding techniques earn coding gain by transforming time domain signals into frequency domain. There are two kinds of time to frequency mapping methods: filter bank and block transform [3].

1.1 Filter Bank

Time to frequency transform using filter bank basically convolves a time domain signal with filter bank and disassembles it into K frequency bands. Then, each frequency band signal is quantized with limited number of bits. During the quantization process it allocates quantization noises to the most inaudible frequency bands considering masking curves drawn from psychoacoustic model. When decoder gets the quantized signals they are merged to recover the full-band signal. One of the biggest problems of filter bank methods using K parallel bands is that data rate increases by K times when input signal is partitioned into K bands. To tackle this problem, encoder down-samples the band-pass-filtered signals by selecting one sample out of K samples. On the other hands, decoder up-samples the band pass filtered signals to compensate the loss of sampling rate. PQMF (Pseudo Quadrature Mirror Filter) [4][5] is a representative example of filter bank methods which is used in commercial audio codecs (MPEG 1,2 Layers I, II, III) [6][1].

1.2 Block Transform

Another important time-frequency mapping method is the realm of block transform. Block transform and filter bank have different histories, but, their internal process can be regarded as identical except that the block transform uses more number of frequency bands. This fact causes audio-oriented codecs which need many frequency channels, such as MPEG AAC [2], Dolby AC-2 and AC-3[7], and AT&T/Lucent PAC [8], adapt block transform. MDCT (Modified Discrete Cosine Transform) is one of the most popular block transform methods since it resolves data rate increment problem caused by overlap-and-add procedure of block transform which is essential in eliminating aliasing cancelation.

2 DIMENSIONALITY REDUCTION

Dimensionality reduction expresses input matrix $X^{(N\times M)}$ as a multiplication of the two output matrix, $A^{(N\times R)}$ and $B^{(R\times M)}$, based on the method-specific conditions. In that, a smaller value is usually assigned to the reduction rate R than original dimension M. The reconstruction error, which can be defined as difference between the input matrix X and recovered matrix $A\times B$, depends on the method-specific condition and the reduction ratio M/R. Figure 1 shows the representation of matrix inner product during general dimensionality reduction methods.

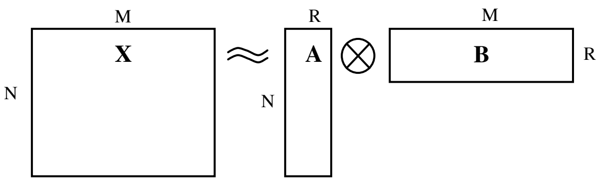

Figure 1. Matrix decomposition during dimensionality reduction

A variety of dimensionality reduction methods have been developed so far, for example, PCA (Principle Component Analysis) [10], ICA (Independent Component Analysis) [11] and NMF (Non-negative Matrix Factorization) [12]. In addition to them, this work can implicate every other dimensionality reduction methods whose reconstruction process can be viewed as a multiplication of two output matrices and when the reduction ration R is smaller than both N and M.

PCA provides projection vectors ordered by their degree of contribution to minimize the correlation of projected samples. The priority of projection vectors let us choose first R projection vectors and corresponding encoding values as the output matrices A and B of our dimensionality reduction scheme.

ICA learns independent probability distributions that lie in the sample data. Since it relaxes the orthogonal property of basis (projection) vectors of PCA to be just independent, it can fit to the sample data more flexibly. ICA is usually said to be better in its compression efficiency than PCA, and primarily used in blind source separation field. Furthermore, audio data compression using ICA produces better sound quality than quantization based on psychoacoustic model [9].

NMF finds out a matrix factorization which preserves non-negativity by restricting its input and output matrices to be non-negative during its multiplicative update process. NMF can only be applied to the non-negative matrices, but, the output matrix A and B are sparser than above methods. The sparseness is a property of human brain when it process data; NMF not only just reflects more about human brain, but it usually surpass other methods in its performance.

3 SEGMENTED DIMENSIONALITY REDUCTION ON FREQUENCY DOMAIN SIGNALS

Our work is based on frequency domain signals whose transformation was made by various time to frequency mapping methods including the ones described in section 1. Those frequency domain signals can make input matrix of dimensionality reduction methods if we regard each frame as an N-length column vector. However, in this case, it is impossible to process the input signals until whole input signal is acquired. To tackle this problem, we propose a segmented dimensionality reduction method which gradually takes parts of total-length matrix as input matrix. In other words, when a predefined number of frames are gathered the encoder processes them with appropriate dimensionality reduction method. The smaller the predefined number is, the less encoder delay is.

An encoder equipped with the dimensionality reduction function first collects M frames of frequency domain signal, each of whose length is N, as the input matrix $X^{N\times M}$. We call those M frames a *segmented* input. Then X is decomposed into $A^{N\times R}$ and $B^{R\times M}$ using a proper dimensionality reduction method. A and B can be stored or transported as soon as they are formed depending on its application. When decoder gets input matrices A and B, it reconstructs original matrix X by multiplying the input matrices. The encoder can gradually take successive M frames and encodes them until the end of the input signal.

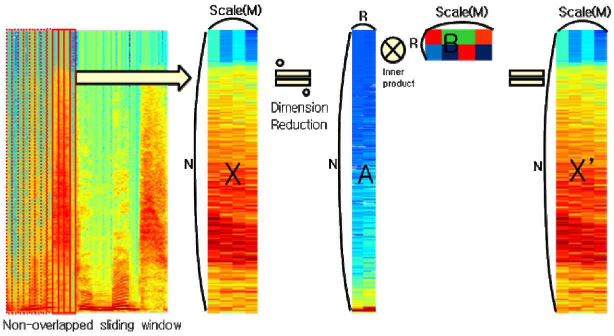

Figure 2. Dimensionality reduction and reconstruction process of a frequency domain signal

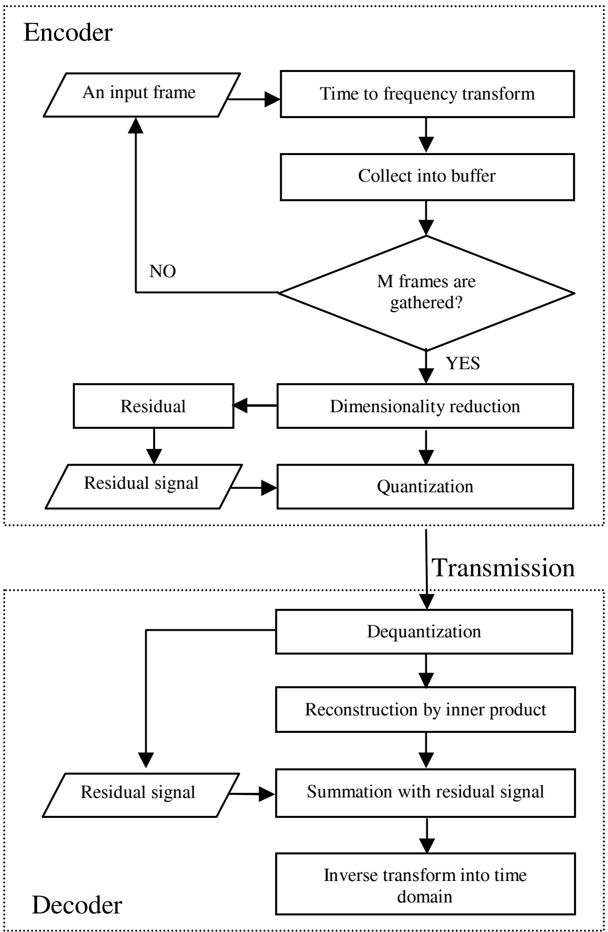

Figure 3. Dimensionality reduction method in a general audio codec.

Figure 2 depicts encoding procedure of the segmented dimensionality reduction on a frequency domain signal when M and R is initialized with 4 and 2 respectively. In this case, $4N$ data samples can be approximately represented with $2N+8$ samples which can recover the original one by simple inner product.

In figure 3, we can see an example where a transform codec collaborate with segmented dimensionality reduction method: *dimensionality reduction* block in the encoder side and *reconstruction by inner product* in the decoder side. Note that a proper residual coding scheme and quantization method are needed.

4 EXPERIMENTAL RESULTS

In our experiments we used NMF and PCA to reduce dimension of input signal. DFT (Discrete Fourier Transform) was used to transform time domain signal into frequency domain because NMF can only take non-negative data samples. In NMF case absolute value of the complex frequency domain sample was used as input while the PCA takes the complex values of DFT results. It is assumed that phase information is known to decoder in NMF case.

Figure 4 shows a comparison among a frequency domain original signal (English female speech) and reconstructions of it using NMF and PCA where M and R were set to 8 and 2 respectively. Figure 5 shows its original time domain version and residual signals of NMF and PCA reconstructions. Residual signals were gotten by subtracting reconstructed signals from original time domain signal. We can decide that the reconstruction result is good if its corresponding residual signal is more noise-like and has less energy. In spite of the quite high compression ratio (8/2), reconstruction result of NMF shows reasonable quality.

We also verified our system with a various types of signals including musical signals.

5 DISCUSSIONS

This paper suggests a segmented dimensionality reduction method as an audio signal compression tool. This method can be used in audio codecs where coding delay is not critical since M frames of encoder-side delay and R frames of decoder-side delay are inevitable. Additional research which tries to enhance the adaptability of this system is in progress: an effective residual coding scheme in order to increase coding gain of the system, figuring out proper quantization methods for this system including the psychoacoustic models for the acoustically impaired dimension-reduced data, and application to various kind of frequency domain transformation including QMF and DCT.

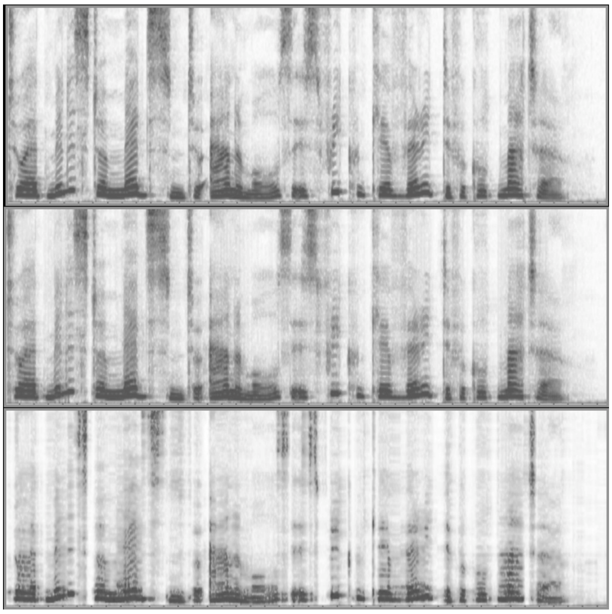

Figure 4. Compressed speech signal using dimensionality reduction (original, NMF reconstruction and PCA reconstruction from top to bottom).

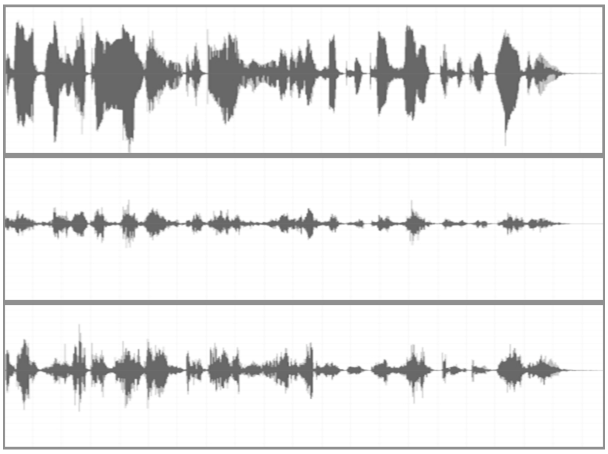

Figure 5. Residual speech signals in time domain (original, NMF residual and PCA residual from top to bottom).

ACKNOWLEDGEMENT

This work was supported by the IT R&D program of MKE/IITA. [2008-F-011-01 Development of Next-Generation DTV Core Technology]

REFERENCES

[1] ISO/IEC 13818-3, Information Technology, "Generic coding of moving pictures and associated audio, Part 3: Audio", 1994-1997.

[2] ISO/IEC 13818-7, Information Technology, "Generic coding of moving pictures and associated audio, Part 7: Advanced Audio Coding", 1997.

[3] M. Bosi, R. E. Goldberg, "Introduction to Digital Audio Coding and Standards", Kluwer Academic Publishers, pp.75-147, 2003.

[4] H. J. Nussbaumer, "Pseudo-QMF Filter Bank", IBM Tech. Disclosure Bull., vol. 24, pp. 3081-3087, November 1981.

[5] J. H. Rothweiler, "Polyphase Quadrature Filters – A new Subband Coding Technique", International Conference IEEE ASSP, Boston, pp. 1280-1283, 1983.

[6] ISO/IEC 11172-3, Information Technology, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s, Part 3: Audio", 1993.

[7] L. D. Fielder, M. Bosi, G. A. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low Complexity Transform-Based Audio Coding," In N. Gielchrist and C. Grewin (ed.), Collected Papers on Digital Audio Bit-Rate Reduction, pp. 54-72, AES 1996.

[8] D. Sinha, J. D. Johnston, S. Dorward and S. R. Quackenbush, "The Perceptual Audio Coder (PAC)", in the Digital Signal Processing Handbook, V. Madisetti and D. Williams (ed.), CRC Press, pp. 42.1-42.18, 1998

[9] A. Ben-Shalom, S. Dubnov and M. Werman , "Improved Low bit-rate audio compression using reduced rank ICA instead of psychoacoustic modeling", IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP2003

[10] I. T. Jolliffe, "Principal Component Analysis", Springer, 2002.

[11] A. HyvÄarinen, J. Karhunen, E. Oja: Independent Component Analysis. John Wiley & Sons, Inc. (2001)

[12] D. D. Lee and H. S. Seung: Learning the parts of objects by non-negative matrix factorization. Nature 401 (1999) 788.791