Lossless Compression of Speech and Audio Signals, and Its Application

著者	原田 登
内容記述	この博士論文は内容の要約のみの公開（または一部非公開）になっています
year	2017
その他のタイトル	音声・音響信号の可逆符号化とその応用に関する研究
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2016
報告番号	12102甲第8081号
URL	http://hdl.handle.net/2241/00148254

Lossless Compression of Speech and Audio Signals, and Its Application

（音声・音響信号の可逆符号化とその応用に関する研究）

システム情報工学研究科　コンピュータサイエンス専攻

原田　登

２０１７年　３月

Abstract

Speech and audio coding technologies have broad application, such as speech communication, radio and TV Broadcasting, portable music players, and storage on optical discs.

The first speech coding technologies were introduced in early 1970s and have widely been used in many applications. In order to accommodate more number of end user devices, high-efficient speech coding schemes are applied. Those coding schemes are capable of archiving compression ratios up to 12:1 and higher, by applying a psycho-acoustic model. The information that human ear is insensitive can be dropped without degrading the subjective audio quality. This type of codec is so called "perceptual lossy", or simply called "LOSSY" coding schemes.

On the other hand, some applications require perfect reconstruction. This type of codec is so called "LOSSLESS" coding schemes. Recently, high-resolution audio service is getting popular. High-resolution audio is specified as sampling frequency of 48 kHz or higher, and bit depth is 24 bit. As for the high-resolution audio, LOSSY coding scheme is of no use because high-resolution audio signal contains frequency beyond that of human hearing and lossy coding schemes will remove information at the frequency. If any kind of compression is applied, it should come with no loss of information (i.e., no fidelity loss). Thus a LOSSLESS technique must be used.

The goal of this study is to provide efficient lossless coding schemes that can be used in real world application. For lossless compression of audio signals represented in IEEE 754 floating-point, a new coding scheme, comprising Approximate Common Factor (ApxCF) coding and the Masked Lempel-Ziv (Masked-LZ) compression, is introduced. In the proposed scheme, an input sequence X is decomposed into three parts: a common multiplier A, a multiplicand sequence Y, and a difference sequence Z. Instead of re-inventing a brand new coding tool, proposed scheme makes use of existing efficient encoding tool for integer input sequences. Experimental test results using professional music recording data show that the ApxCF coding can reduce the bit rates considerably, especially when the input values in a frame are constructed by multiplication of the sequence of integer values and a floating-point constant. In

addition, the Masked-LZ compression scheme has the potential to reduce bit rates of the difference mantissa. The scheme has been accepted as a part of an ISO/IEC standard, MPEG-4 Audio Lossless Coding (ALS).

For lossless compression of log-companded speech signals, the input target is ITU-T G.711 encoded sequence sampled with 8 kHz, 8 bit, 64 Kbit/s. Plus Minus Zero (PMZ) mapping is proposed for the prediction residual calculation in Mapped Domain Linear Prediction (MDLP) and Escaped-Huffman (E-Huffman) coding combined with adaptive recursive Rice coding is proposed for the prediction residual compression. It is shown that the PMZ mapping improves the compression performance by 0.2% for mu-law input. The E-Huffman coding combined with adaptive recursive Rice coding improves the compression by 0.16% averaged for all test conditions, compare to the conventional Rice coding scheme. Average computational complexity is 1.071 WMOPS for the encoder/decoder pair and the worst-case complexity is 1.667 WMOPS in total. These proposed schemes are approved as a part of ITU-T Recommendation G.711.0. The G.711.0 standard provides more than 50% average compression in service provider environments while keeping low computational complexity for the encoder/decoder pair (1.0 WMOPS average, <1.7 WMOPS worst case) and low memory footprint (about 5k octets RAM, 5.7k octets ROM, and 3.6k basic operators).

In addition, in order to apply those proposed coding schemes to long-term preservation and media data exchange, an archival information package format complies with the Open Archival Information System (OAIS) reference model is designed. The proposed archiving format is approved as an ISO/IEC standard: MPEG-A Professional Archival Application Format (PA-AF).

MPEG-A PA-AF is applied to archiving of recorded audio project. Two standard-compliant implementations of the PA-AF packaging and un-packaging tool are introduced. The implementations made use of MPEG-4 ALS for lossless compression of audio files and Gzip for other input files. Proposed implementation 1 is an open-source version and Proposed implementation 2 is optimized for audio archiving applications in terms of processing speed. An optimized MPEG-4 ALS codec library is applied to the implementation 2. Experimental test results show that the proposed archiving tool with the optimized implementation performs much better than widely used archiving tools such as Tar-gz, MacDMG and WinZip. Compression performance of the proposed PA-AF implementation is equivalent to or much better than other tools while keeping processing speed much faster. The devised PA-AF tool is used in commercial archiving systems in music industry.

Another example application is proposed. MPEG-A PA-AF combined with ITU-T G.711.0 is applied to archiving of speech data for telephone customer support system. In the proposed system, speech data is efficiently preserved and easily accessed for improving end user experience therefore customer satisfaction.

By applying the proposed enhancement of lossless coding schemes and the proposed package format, long-term preservation of time domain signals along with related contents and metadata has been made possible.