On the Efficiency Difference between Range and Huffman Coding on CELT Layer of Opus Audio Coder

Shingchern D. YOU[1] , Senior *Member, IEEE,* and Po-Yueh LAI
National Taipei University of Technology, Taipei Taiwan

***Abstract*—This paper compares the coding efficiency between the range coder in the Opus coder and the Huffman coder used in the MP-3 (MPEG-I Layer 3) and MPEG-2 AAC. The results show that the range coder has efficiency advantage of about 9 % at a rate of 128 kbps. The simulation, in a sense, indicates that transcoding from the Opus format to MP-3 or AAC format will lead to quality degradation.**

I. Introduction

With the advances of technology, audio coding technologies have tremendous progress in the past twenty years. The audio coding technology can be divided into lossless and lossy approaches. As the lossy coding approach can easily reach a compression rate of ten or higher without obvious quality impairment, this approach is widely used in all commercial markets, such as online music services.

Among the lossy coders currently (year 2018) in use, most of them were developed either by big companies or by ISO (International Standardization Organization). Thus, using these coders require paying for a license (royalty) fee. To address this issue, the Xiph.org foundation, founded in 1994, devotes to produce free multimedia formats and software tools [1]. The audio coder developed by the Xigh.org was the Opus coder, intended to encode both natural audio and speech. The Opus coder was subsequently standardized by the IETF (Internet Engineering Task Force) [2]. According to Wikipedia, the Opus coder was "designed to efficiently code speech and general audio in a single format, while remaining low-latency enough for real-time interactive communication and low-complexity enough for low-end embedded processors." In terms of coded audio quality, the Opus coder also has better coded quality [3] than many standard audio coders, such as MP-3 (MPEG-1 Layer III) [4] and HE-AAC (MPEG-4 High Efficiency Advanced Audio Coding) [5].

Despite the great success of the Opus coder, there are not too many published papers discussing which blocks (tools) are the main sources leading to the superior coding quality. To this end, we study the efficiency difference between the range coder used in the Opus and the Huffman coder used in MP-3 and HE-AAC. As the range coder is a context-based arithmetic coding, it could be more efficient than the static Hufffman coder is. However, it is uncertain to what extent the range coder outperforms the counterpart. The answer to this problem has two purposes. The first one is to answer whether using the range coder is a good choice for the Opus scheme. The second one is to understand the possible quality degradation if transcoding from the Opus format to AAC or MP-3 format. For example, the Opus coder supports a frame size of 960 samples for 48 ks/s audio. Accidentally, one variation of the LC (low complexity) AAC coder used mainly in broadcasting also uses 960 samples per frame at 48 ks/s. Thus, it might be possible to directly transcode the Opus quantized MDCT coefficients to the AAC format without going through the entire decoding and then encoding process.

Corresponding author. E-mail: you@csie.ntut.edu.tw

II. Basics of CELT Coder

A. Overview of Opus

The Opus coder consists of two parts: SILK for speech coding and CELT (Constrained-Energy Lapped Transform) [6] for natural audio. Fig. 1 shows the signal flow diagram of the CELT coder. The first step is pre-filtering, used mainly for noise shaping and for finding the harmonic structure of the signal. The filtered PCM samples, after windowing, are converted to frequency domain by the MDCT (modified discrete cosine transform) with the MDCT coefficients grouped into frequency bands based on the Bark scale. The spectral values within the same frequency band share a common coarse gain. The residual values are sent to the pyramid vector quantizer (PVQ) [7], followed by the range coder [8] to encode the indices of the codewords.

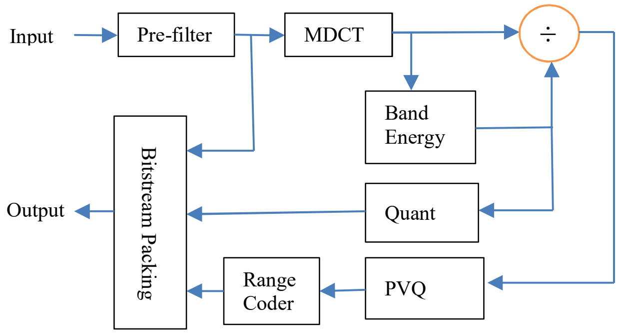

Fig. 1. Signal processing flow of the CELT coder.

B. Quantization

The quantization of the coarse energy follows the coarse, fine, and final fine procedure. If necessary, the coder can use time or frequency prediction to improve coding efficiency. Once the coarse energy values are determined, the MDCT coefficients are normalized accordingly. As the CELT coder has no psychoacoustic model inside, it determines the number of available bits per band based on pre-determined tables with some adjustments. The quantization of the MDCT coefficients are conducted by a special type of vector quantization, called PVQ. The size of the PVQ codeword, $N(N,K)$, is recursively computed based on the number of samples N and the number of pulses K. In the CELT coder, K is a function of the available bits in a band and N is a fraction of the number of MDCT coefficients in a band. The index of the PVQ codeword, after a conversion, is coded by the range coder.

III. Experiments and Results

A. Experimental Setting

To conduct experiments, we prepare 14 audio clips with different genres as testing clips. The clips are down-mixed to mono from stereo sources. The sample rate of the clips is 48 ks/s. The compression type is CBR (constant bitrate) and the bitrate is 128 kbps and 64 kbps, respectively.

When carefully examining the indices of the PVQ, we notice that their values are typically very large, sometimes larger than 16-bit can represent. This situation makes the use of the Huffman tables very difficult and inefficient. Thus, to improve the efficiency, we encode the codewords, but not the indices, when using the Huffman tables.

As the MP-3 and AAC has many Huffman tables to choose, we then divide the quantized spectral values into several parts (partitions) according to their MDCT indices. For MP-3 tables, the first part uses tables with escape (ESC) values, and the remaining parts use tables with largest absolute value (LAV) of 15, 7, 5, 3, 2, and 1, respectively. For AAC Huffman tables, tables with LAVs of 16+ESC, 12, 7, 4, 2, and 1 are used.

B. Opus Bit Distribution

This experiment is to find out the number of bits used for in various parts of the Opus bitstream. The results show that the range coding actually occupies the majority of the available bits, as shown in Table I where Mis. means miscellaneous. Therefore, it is worthwhile to concentrate on this part.

TABLE I
Average Bits Used in Each Part in Opus Bitstream

Rate (kbps)	Coarse	Fine	Final	Range	Mis.	Overall
128	35	76	3	2409	28	2551
64	35	61	2	1145	28	1271

C. Bits Distribution Using Huffman Tables

Fig. 2 shows the average percentage of used bits in each part for 128 kbps and 64 kbps using MP-3 and AAC Huffman tables. For 128 kbps case, although the AAC table uses more bits in part 1, it uses fewer bits in part 2 because we use a table with LAV of 12 for AAC, but 15 for MP-3. As not a single part consumes the majority of the available bits, the divided parts and chosen tables seem to be reasonable.

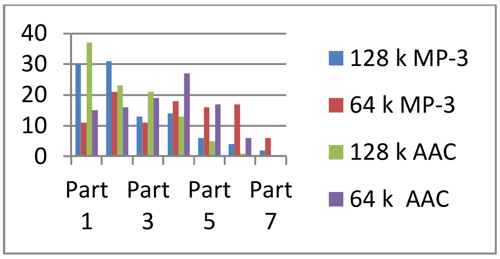

Fig. 2. Percentage of bit consumption using MP-3 and AAC Huffman tables.

D. Comparison of Used Bits in Range and Huffman Codes

The next experiment compares the used bits in range, MP-3 Huffman, and AAC Huffman coders. The results are shown in Table II. From the table we know that the range coder in deed has higher coding efficiency. At 128 kbps bitrate, the savings of bits is around 220 bits compared with Huffman coders. When looking closely at the PVQ quantized spectral values, we notice that the MDCT spectral amplitudes do not smoothly decreasing, an assumption made to design the Huffman coders. In the Opus case, we may see only a few large values in between many 0s and 1s. Consequently, this part of spectral values cannot be coded with LAV = 1 tables, resulting in a lower compression ratio. Thus, to fit into the bitrate constraint, some spectral values may have to be changed.

The efficiency difference for the 64 kbps case is even larger, as also given in Table II. It is also interesting to note that the AAC Huffman coder is less efficient than that of the MP-3 in this case because the chosen Huffman tables for AAC are manually optimized for 128 kbps, not for 64 kbps.

TABLE II
Average Bits Used for Various Approaches

Rate (kbps)	MP-3	AAC	Range
128	2631	2622	2409
64	1281	1314	1145

IV. Conclusion

This paper studies the compression efficiency of the range coder used in the Opus coder. When the quantized spectral coefficients are compressed by the Huffman coders based on the MP-3 and AAC schemes, they require more bits than compressed with the range coder. Therefore, if a direct transcoding in the MDCT domain from the Opus format to the MP-3 or AAC format, some modifications on these values are necessary to account for the lower efficiency of the Huffman coders. Consequently, the quality of a transcoded audio clip is likely lower than the original one.

Reference

[1] See https://en.wikipedia.org/wiki/Xiph.Org_Foundation for more details. (Retrieved on 1/10/2018)

[2] J.-M. Valin, K. Vos, and T. Terriberry, "Definition of the Opus audio codec," Internet Engineering Task Force (IETF), RFC 6716, 2012.

[3] See https://en.wikipedia.org/wiki/Opus_(audio_format) for details. (Retrieved on 1/10/2018)

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

[5] ISO/IEC, Information technology - 14496-3; Coding of audio-visual objects - Part 3: Audio, IS 14496-3, 2009.

[6] Valin, Jean-Marc, et al. "A high-quality speech and audio codec with less than 10-ms delay," *IEEE transactions on audio, speech, and language processing*, vol. 18 no. 1, pp. 58-67, 2010.

[7] T. Fischer, "A Pyramid vector quantizer," *IEEE Trans. on. Information Theory*, vol. 32, pp. 568 – 583, July 1986.

[8] G. Martin, "Range coding: an algorithm for removing redundancy from a digitized message," *Video and Data Recording Conference*, Southampton, UK, July 1979.