Parameter Analysis for the Generalized LZ Compression of Audio

Darko Kirovski† and Zeph Landau‡
† Microsoft Research, One Microsoft Way, Redmond, WA, USA
‡ Department of Mathematics, The City College of New York, New York, NY, USA

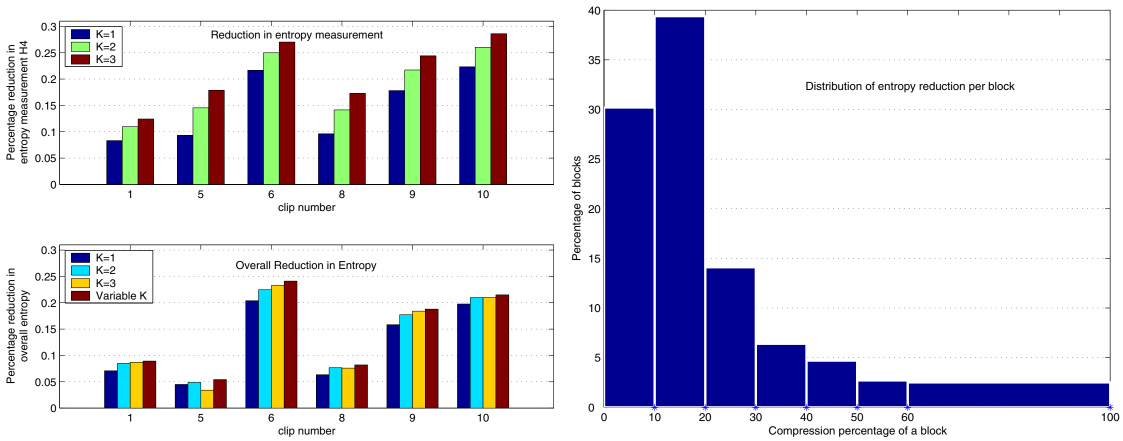

Figure 1: **(left) Entropy reduction performance.** The top left graph represents the reduction in $H(\varepsilon)$ achieved after 1-3 rounds. The bottom left graph includes the cost of pointer encoding. In this graph, the first three bars are the results when using 1-3 rounds. The fourth bar is the performance when a variable round scheme is used. **(right) Distribution of entropy reduction.** The histogram gives the distribution of the reduction in the entropy per block.

Recently, we introduced a memory-based model of the source signal **[1]**, which explores multimedia repetitiveness to improve upon compression rates achieved by classic memoryless or simple prediction-based audio compression algorithms such as MP3. We consider an input signal $\mathbf{x}$ of N samples, where each sample $x_i \subset \mathbf{x}$ is normalized $x_i \in [-1, 1]$. The signal is partitioned into 50% overlapped and windowed blocks of $n = 2^k$ samples where $k \in [9, 12]$. For a given signal block $\mathbf{x}_i = \{x_i, \ldots, x_{i+n-1}\}$, we establish a search window $\mathbf{s}_i = \{x_{\text{start}(i)}, \ldots, x_{i-1}\}$ which either represents the full signal history (i.e., $\text{start}(i) = 1$) or has a certain predetermined length S (i.e., $i - \text{start}(i) = S$). We search for a subset of blocks $W = \{\mathbf{w}_j \subset \mathbf{s}_i, j = 1 \ldots K\}$ and a set of corresponding scalars $A = \{\alpha_j \in \mathbb{A} \subset \mathbb{R}, j = 1 \ldots K\}$ and transforms $F = \{f_j(\mathbb{R}^n) \to \mathbb{R}^n, f_j \subset \mathbb{F}, j = 1 \ldots K\}$, where $\mathbb{F}$ is the set of all considered transforms. Sets F, A, and W satisfy the following optimization goal: $\arg\min_{W,A,F} H\left\{m\left[\mathbf{x}_i, \sum_{j=1}^{K} \alpha_j f_j(\mathbf{w}_j), \mathbf{b}_i\right]\right\}$. The representation error $\varepsilon = \mathbf{x}_i - \sum_{j=1}^{K} \alpha_j f_j(\mathbf{w}_j) = \mathbf{x}_i - \mathbf{r}_i$ is masked using a psycho-acoustic filter $m()$ as follows. We first compute the psycho-acoustic mask $\mathbf{b}_i$ of the source block $\mathbf{x}_i$. The mask $\mathbf{b}_i \in \{0, 1\}^n$ distinguishes audible from inaudible frequency coefficients and can be computed using well known psycho-acoustic models **[2]**. We now define the masking function $m(x, r, b)$ on a single signal coefficient x and its reconstruction r and masking bit b: $m(x, r, b) \equiv \{b = 1 \Rightarrow m = x - r, b = 0 \wedge |r| \leq T \Rightarrow m = 0, b = 0 \wedge |r| > T \Rightarrow m = r - T \cdot \text{sign}(r)\}$ where T denotes the hearing threshold for sample x. The goal of the masking function $m()$ is to set the error such that reconstruction of audible samples is exact whereas the reconstruction of inaudible samples is such that the absolute magnitude of the error is minimized. Finally, function $H()$ computes the entropy of: the quantized pointers to all blocks in W, the quantized pointers to the applied transforms F, the quantized scalars in A used to create the linear combination of transformed blocks, and the error vector returned by function $m()$. By encoding this data we obtain the final compressed stream. The optimization goal is to find a set W of K blocks which occur prior to $\mathbf{x}_i$ and a linear combination of their transforms $\mathbf{r}_i$, which represents $\mathbf{x}_i$ as close as possible in the sense of minimizing the entropy of the remaining error vector $m(\mathbf{x}_i, \mathbf{r}_i, \mathbf{b}_i)$. On a select audio benchmark, our scheme produced a total entropy reduction of 8.9%, 5.4%, 24.1%, 8.1%, 18.7% and 21.5% respectively. For certain electronic music pieces, however, gains were reaching 90%. Details about the experiments can be found in **[1]**.

[1] D. Kirovski and Z. Landau. Generalized Lempel-Ziv Compression for Audio. IEEE *MMSP*, 2004.

[2] T. Painter and A. Spanias. Perceptual coding of digital audio. *Proceedings of the* IEEE, pp.451–513, April 2000.

Proceedings of the 2005 Data Compression Conference (DCC'05)
1068-0314/05 $ 20.00 **IEEE**