LOSSLESS AUDIO CODING WITH BANDWIDTH EXTENSION LAYERS

Stephen Voran
Institute for Telecommunication Sciences
325 Broadway, Boulder, Colorado, 80305, USA
svoran@its.bldrdoc.gov

ABSTRACT

Layered audio coding typically offers reduced distortion as bit rate is increased, but that distortion is spread across the entire band until the lossless coding bit rate is reached and distortion is eliminated. We propose a layered audio coding paradigm of bandwidth extension, rather than distortion reduction. For example, a core layer can provide lossless coding of a 24 kHz bandwidth signal (f_s=48 kHz), then first and second bandwidth extension lossless layers can extend that signal to losslessly coded 48 and then 96 kHz bandwidths (f_s=96 and 192 kHz).

1. INTRODUCTION

Consideration of the Nyquist theorem ($f_s=2 \cdot f_N \geq 2 \cdot f_{max}$, where f_N is the Nyquist frequency) and the conventional upper limit of human hearing (f_{max}=20 kHz) suggests that audio sampling rates f_s =44.1 or 48 kHz would be sufficient for creating a transparent digital representation of any audible signal. Yet double and quadruple sampling rates (f_s=88.2, 96, 176.4, and 192 kHz) have found favor and use among audio professionals. Possible explanations are discussed in [1]-[3] and include filter design issues as well as the possible value of signal components above 20 kHz for some signals and some listeners.

The disadvantage of higher sample rates is the higher data rates of the resulting digital audio streams. This can be partially mitigated with lossless coding [4]. A more flexible approach is layered, scaleable, or hierarchical coding. These approaches allow for increasing fidelity as bit rate is increased and the endpoint of this relationship is lossless coding. A prime example is scaleable to lossless (SLS) coding [5] included in the MPEG-4 audio standard and based on successive refinement of integer modified discrete cosine transform (IntMDCT) coefficients. But for any given sample rate and corresponding audio bandwidth, SLS shows some loss across the entire band at all bit-rates below the final lossless coding rate. A layered approach that further mixes lossy coding, lossless coding, sample rate reduction, and amplitude resolution reduction is given in [6].

We propose a different layered coding paradigm for audio signals sampled at double and quadruple rates. It is motivated by the fact that different users make different choices when weighing the costs and benefits of sampling at 48, 96, or 192 kHz, coupled with the fact that sending or storing multiple versions is inefficient. It is further motivated by the expectation that users limited by data rates and storage space may prefer a lossless coding of a 24 kHz version of a signal (f_s=48 kHz) over a lossy coding of the 96 kHz original signal (f_s=192 kHz).

The proposed paradigm is one of bandwidth extension, rather than distortion reduction. A core or base coding layer can losslessly encode the band from 0 to 24 kHz and the resulting signal has f_s=48 kHz and the full (e.g., 24 bit) amplitude resolution. A first bandwidth extension lossless layer (BELL) can extend the core signal to provide lossless coding of the 48 kHz bandwidth version of the signal (f_s=96 kHz). A second BELL can then extend that signal to provide lossless coding of the original, 96 kHz bandwidth signal (f_s=192 kHz). The core layer and the first BELL are formed through lossless coding of bandlimited versions of the original signal with full amplitude resolution. For the term "lossless coding" to have meaning, the notion of "bandlimited versions" (i.e., filtering and subsampling) must be agreed to and precisely defined.

In the following we present integer-input, integer-output filters that create suitable bandlimited versions of audio signals. We evaluate the BELL approach for a single lossless coding scheme comprised of normalized least mean squares (NLMS) prediction with entropy coding. Through this process we also see how the information content of the signal spectrum decomposes and observe that this is very different from how the energy in the spectrum decomposes.

2. PRELIMINARIES

2.1. Filters

When losslessly coding a limited bandwidth, very high quality filters are required. Figure 1 shows the magnitude response of three filters designed with a frequency domain least-squares method. Each of these is a linear phase, order 310, FIR filter. The lowpass filter (LPF) has an attenuation of at least 120 dB for frequencies at or above $0.5 \cdot f_N$ and this is critical to minimize aliasing when subsampling. The −3dB point is at $0.4702 \cdot f_N$, and the −0.01 dB point is at $0.4539 \cdot f_N$. These −0.01 dB points translate to 20,017 Hz and 21,787 Hz when f_s=88.2 and 96 kHz respectively. Passband ripple is negligible. These passband characteristics are critical to the preservation of full fidelity within the passband. The highpass filter (HPF) has analogous specifications as it is the mirror image of the LPF about $0.5 \cdot f_N$. The bandpass filter is complementary to the LPF and HPF in the sense that when used in parallel their combined z-domain transfer function is a pure delay of k=155 samples:

$$H_{LPF}(z) + H_{BPF}(z) + H_{HPF}(z) = z^{-k}. \quad (1)$$

Integer filter outputs are required for lossless coding. For this initial work we accomplish this by simply rounding real signal values to integer signal values at the filter output. This causes filter stop-band performance to depart from the behavior depicted in Fig. 1 and this departure is more dramatic for smaller signals. Examples for the LPF are included with broken lines and correspond to a maximum level signal (below) and a signal 80 dB below maximum level (above). In a more

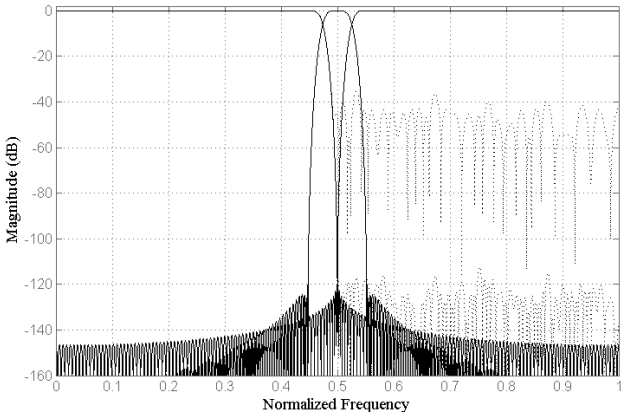

Figure 1: *Magnitude responses for three filters and two examples of lowpass response due to rounding.*

sophisticated implementation, one might seek to impart some spectral shaping to this rounding error [7]. For these filters to be used in conjunction with lossless coding across multiple platforms, bit-exact filter operation must be specified so that repeatable results for every possible signal are assured. One might also consider the apodizing filters described in [8].

2.2. Audio Recording

Our audio recording system uses studio-grade professional equipment exclusively. The system consists of a pair of small-diaphragm condenser cardioid microphones in a coincident (X-Y) configuration, a stand-alone microphone preamplifier, delta-sigma A/D conversion at 7 MHz that produces 24 bit linear samples at f_s=192 kHz, and hard drive recording. The measured −0.5 dB bandwidth of the electrical portion of the system extends from 11 Hz to 89 kHz. The high frequency −3 dB point is at 91.5 kHz. The manufacturer's specification shows microphone response is down by just 1 dB at 20 kHz (relative to 1 kHz) but no response values above 20 kHz are available. The unweighted noise floor (96 kHz bandwidth) of the electrical portion of the system is 86.9 dB below the peak unweighted signal level that we recorded.

We recorded a recital in a church sanctuary. The recital contained a broad range of musical sources including vocals, piano, violin, hammer dulcimer, recorder, guitars, banjo, and electric bass. These different sources appeared in a variety of solo and ensemble configurations. After editing, the final recording contained 56 minutes of stereo music and applause. We segmented this into 112, 30-second stereo segments and saved each segment in a separate file. Half of these files were selected at random and assigned to a training database and the other half became the testing database.

2.3. Lossless Stereo Matrix

We exploit correlation between the stereo channels with a sum-and-difference lossless stereo matrix. Given that the left and right signals x^L and x^R are each represented with b=24 b/smp, most such representations will require b+1 b/smp for the sum signal, x^S, and b+1 b/smp for the difference signal, x^D. We have adopted a more economical lossless representation that requires just b b/smp for x^S and b+1 b/smp for x^D:

$$x^S = \left\lfloor \left(x^L + x^R\right)/2 \right\rfloor \quad , \quad x^D = x^L - x^R$$
$$x^L = x^S + \left\lceil x^D/2 \right\rceil \quad , \quad x^R = x^S - \left\lfloor x^D/2 \right\rfloor , \tag{2}$$

where $\lfloor \bullet \rfloor$ and $\lceil \bullet \rceil$ represent floor and ceiling functions.

2.4. Predictor

Throughout this work we perform lossless audio coding via a single NLMS predictor [9] followed by entropy coding. Selecting a single coding scheme allows us to make meaningful comparisons in the tables that follow. For an optimized BELL scheme one would also consider nesting multiple predictors [10],[11] and integer transform coding [5],[7]. We note that IntMDCT is used effectively in SLS and might also seem like a natural mechanism for implementing BELLs. Lower-frequency transform coefficients might form a core layer and higher-frequency transform coefficients would then form extension layers. Unfortunately, the band-limiting operation that is implicitly invoked when forming a core layer by this means suffers from several undesirable properties including very poor stopband attenuation.

2.5. Power, Entropy, and Mutual Information

Table 1 shows the results of an initial analysis of the training data. Full-band results are based on the original data. Low- and high-band results are based on the LPF and HPF outputs, subsampled by a factor of two. The results shown are for x^S and identical trends were observed for x^D.

	Full Band	Low Band	High Band
Bandlimits (kHz)	0-96	0-48	48-96
Bandwidth (kHz)	96	48	48
f_s (kHz)	192	96	96
Mean Power (dB)	0.0	0.0	−63.3
Entropy (b/smp)	19.3	19.3	8.6
Resid. Entropy (b/smp)	10.7	11.7	8.6
Appx. Rel. Coding Rate	0.446	0.244	0.179
Appx. Rel. Coding Rate	0.446	0.423 (combined)	

Table 1: *Power, entropy, and relative rate results for three signals; original sample rate is 192 kHz.*

The mean power row in Table 1 shows the time-averaged power in each band, normalized for 0 dB in the full band. The low band has nearly the same power as the full band, consistent with the high band having power 63 dB below the low band. This reduced high-band power is consistent with the measured high-band entropy reduction reported in the fifth row (a 6 dB per bit relationship).

The sixth row gives the entropy of the prediction residual after 8th order NLMS prediction with adaptation parameter λ=6. Comparing rows five and six reveals that the full- and low-band signals have significant predictable components (8.6 or 7.6 b/smp can be saved) but the high-band signal does not (no entropy reduction measured). The final two rows of the table provide approximate coding rates (prediction residual entropy × f_s) normalized to the original full rate (24 b/smp × 192,000

smp/sec). Throughout this paper, "relative coding rate" indicates a coding rate normalized by $24 \times f_s$ b/sec (mono case) or $48 \times f_s$ b/sec (stereo case). The close relative coding rates on the final row of Table 2 indicate the possibility of losslessly coding the low and high bands separately at a total rate that is similar to the full-band-lossless coding rate.

We further split the low band of Table 1 into high and low bands and these results are given in Table 2. On this second splitting we again observe a vast power difference and a consistent entropy difference between the low and high bands. Here we find a very small (rather than zero) entropy reduction in the high band due to NLMS prediction. As before, the final row indicates the potential of coding the low and high bands separately at a total rate that is similar to the full-band-lossless coding rate.

	Full Band	Low Band	High Band
Bandlimits (kHz)	0-48	0-24	24-48
Bandwidth (kHz)	48	24	24
f_s (kHz)	96	48	48
Mean Power (dB)	0.0	0.0	−63.4
Entropy (b/smp)	19.3	19.3	8.1
Resid. Entropy (b/smp)	11.7	13.6	7.9
Appx. Rel. Coding Rate	0.488	0.283	0.165
Appx. Rel. Coding Rate	0.488	0.448 (combined)	

Table 2: *Power, entropy, and relative rate results for three signals; original sample rate is 96 kHz.*

Next we ask if the coding rate of either high band might yield significantly to *any* lossless coding. This would be contingent upon the existence of some form of statistical dependence within or among the signals to be coded. To assess the existence of any possible statistical dependence, we turn to mutual information (MI). The MI between the random variables x and y is given by

$$I(x;y) = H(x) + H(y) - H(x,y), \quad (3)$$

where $H(\cdot)$ denotes the entropy of a random variable or pair of random variables. Table 3 shows results of five different MI calculations across all training data. Where necessary, subscripts L and H denote low and high band signals.

Mutual Information Calculation	Low Band (0 to 48 kHz)	High Band (48 to 96 kHz)
$x_\bullet^L(n)$ and $x_\bullet^R(n)$	1.0714	0.0133
$x_\bullet^S(n)$ and $x_\bullet^S(n+1)$	4.0405	0.0162
$x_\bullet^D(n)$ and $x_\bullet^D(n+1)$	3.5038	0.0160
$x_L^S(n)$ and $x_H^S(n)$	0.0020	
$x_L^D(n)$ and $x_H^D(n)$	0.0036	

Table 3: *MI results in bits; original sample rate is 192 kHz.*

The low band shows significant MI between channels and between subsequent samples of x^S and of x^D. In the high band, virtually no MI is found between channels nor between subsequent samples. Table 4 shows results when we split the low band into two additional bands. High-band MI results are larger than in Table 3, but are still very small. Based on this lack of MI, we conclude that there is very little rate reduction to be found when losslessly coding either of these high bands.

Mutual Information Calculation	Low Band (0 to 24 kHz)	High Band (24 to 48 kHz)
$x_\bullet^L(n)$ and $x_\bullet^R(n)$	1.0717	0.0440
$x_\bullet^S(n)$ and $x_\bullet^S(n+1)$	3.2078	0.1248
$x_\bullet^D(n)$ and $x_\bullet^D(n+1)$	2.7402	0.1231
$x_L^S(n)$ and $x_H^S(n)$	0.0096	
$x_L^D(n)$ and $x_H^D(n)$	0.0258	

Table 4: *MI results in bits; original sample rate is 96 kHz.*

3. CODING WITH BANDWIDTH EXTENSION LOSSLESS LAYERS

3.1. Coding Structures

Our initial coding structure uses the filters of Figure 1 and is outlined in Figure 2. The LPF output, subsampled by 2 (y_L) forms the core layer, and this layer can be losslessly compressed. The bandwidth extension layer comprises y_M, y_H, and y_R and none of these three signals can be losslessly compressed. The residual signal y_R is required to compensate for rounding at the filter outputs and this signal cannot be subsampled.

Figure 2 is intended to be conceptual; multiple specific implementations can follow from it. In particular we can extract the three subband signals sequentially so that the rounding error associated with a given subband is passed on to subsequent subbands. We experimented with different subband extraction orders and different locations for the rounding functions. Together these options provide rudimentary forms of error shaping. We also considered different bit slicing arrangements for the BPF and HPF signals. Removing least significant bits from y_M and y_H reduces the entropy there, but increases the entropy of y_R.

The most efficient configuration turns out to be rather simple: disable the BPF and HPF channels and let the residual signal, y_R, carry all of the information not contained in the core signal y_L. Thus the extension signal is comprised of y_R alone. This residual contains mid- and high-band information, as well as rounding error that spans the entire band. This signal cannot be subsampled, but because the bulk of the power is in the upper half of the band, it can be losslessly coded at a reduced rate. For NLMS based coding, we have determined that order 32 with λ=4 is an effective choice for this signal. We use order 8 NLMS with λ=6 for y_{in} and y_L.

3.2. Coding Results

We applied the simplified coding structure (y_L as core layer, y_R as extension layer) to our training data to build up residual histograms for three predictor outputs. The predictor inputs are y_{in}, y_L, and y_R, with $y_{in}=x^S$ or x^D. We used these histograms to

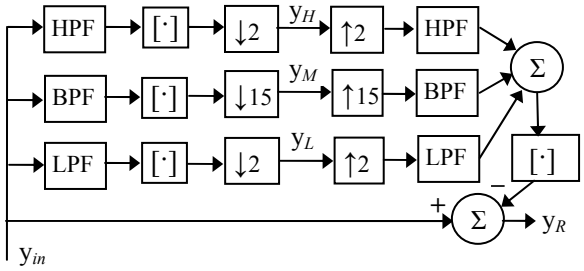

Figure 2: *Conceptual outline of coding structure;* [·] *indicates rounding to an integer value.*

develop a Huffman codebook for each of the three prediction residuals. Finally we applied the coding structure and the trained codebooks to the testing data. The resulting relative coding rates are shown in Tables 5 and 6.

	Full (y_{in})	Core (y_L)	Extension (y_R)
Bandlimits (kHz)	0-96	0-48	0-96
f_s (kHz)	192	96	192
Stereo Coding Rate (b/smp)	22.50	24.39	12.08
Rel. Coding Rate	0.469	0.254	0.252
Rel. Coding Rate	0.469	0.506 (combined)	

	Full (y_{in})	Core (y_L)	Extension (y_R)
Bandlimits (kHz)	0-48	0-24	0-48
f_s (kHz)	96	48	96
Stereo Coding Rate (b/smp)	24.39	28.38	11.69
Rel. Coding Rate	0.508	0.300	0.244
Rel. Coding Rate	0.508	0.544 (combined)	

Tables 5 and 6: *Stereo coding rate results; original signals have sample rates of 192 and 96 kHz.*

4. DISCUSSION

The final rows of Tables 5 and 6 show that information content is approximately evenly distributed between the core and extension layers, even though the powers in the low and high bands are very different (c.f. row 4 in Tables 1 and 2). The relative coding rates of these two tables are summarized and displayed as absolute rates in Table 7. The stereo core (C), first extension (E_1) and second extension (E_2) layers require rates of 1.38, 1.12, and 2.32 Mb/sec respectively. The table shows that decomposition into layers causes a small increase in coding rate (0.16 or 0.50 Mb/s) over simply coding the full signal. But the final column of the table shows that when all three signals are required, the BELL approach saves 3.22 Mb/sec.

SLS provides lossy or lossless coding of the entire signal bandwidth while the BELL approach provides lossless coding of bandlimited versions of the signal. These are fundamentally different approaches and when combined (SLS coding of core and extension layers) they could yield additional flexibility. Finally we note that the BELL approach might be used to also offer "distortion free" low-rate audio with bandwidths narrower than 20 kHz.

Signal Bandwidth (kHz)	24	48	96	All 3 Signals
Layers Used	C	C, E_1	C, E_1, E_2	C, E_1, E_2
BELL Rate (Mb/sec)	1.38	1.38 + 1.12 2.50	1.38 1.12 + 2.32 4.82	4.82
Non-Layered Rate (Mb/sec)	1.38	2.34	4.32	1.38 2.34 + 4.32 8.04

Table 7: *Stereo coding rates for BELL and non-layered lossless coding.*

5. REFERENCES

[1] J. R. Stuart, "Coding for high-resolution audio systems," *J. Audio Eng. Soc.*, vol. 52, no. 3, pp. 117-144, Mar. 2004.

[2] S. Yoshikawa, S. Nome, M. Ohsu**,** S. Toyama, H. Yanagawa, and T. Yamamoto, "Sound quality evaluation of 96 kHz sampling digital audio," Audio Engineering Society 99th Convention, paper 4112, New York, Oct. 6-9, 1995.

[3] K. Hamasaki, T. Nishiguchi, K. Ono, and A. Ando, "Perceptual discrimination of very high frequency components in musical sound recorded with a newly developed wide frequency range microphone," Audio Engineering Society 117th Convention, paper 6298, San Francisco, Oct. 28-31, 2004.

[4] M. Hans and R.W. Schafer, "Lossless compression of digital audio," *IEEE Signal Processing Magazine*, no. 4, pp. 21-32, Jul. 2001.

[5] R. Yu, S. Rahardja, L. Xiao, and C.C. Ko, "A fine granular scalable to lossless audio coder," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 14, no. 4, pp. 1352-1363, Jul. 2006.

[6] T. Moriya, A. Jin, T. Mori, K. Ikeda, and T. Kaneko, "Hierarchical lossless audio coding in terms of sampling rate and amplitude resolution," *Proc. IEEE ICASSP 2003,* vol. 5, pp. 409-412, April 6-10, 2003, Hong Kong.

[7] Y. Yokotani, R. Geiger, G.D.T. Schuller, S. Oraintara, and K.R. Rao, "Lossless audio coding using the IntMDCT and rounding error shaping," to appear in *IEEE Trans. Audio, Speech and Language Proc.*

[8] P.G. Craven, "Antialias filters and system transient response at high sample rates," *J. Audio Eng. Soc.*, vol. 52, no. 3, pp. 216-242, Mar. 2004.

[9] S. Haykin, *Adaptive Filter Theory*, Upper Saddle River, NJ: Prentice-Hall, 2004, ch. 6.

[10] G.D.T. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *IEEE Trans. Speech and Audio Proc.*, vol. 10, no. 6, pp. 379-390, Sep. 2002.

[11] H. Huang, S. Rahardja, X. Lin, R. Yu, and P. Franti "Cascaded RLS-LMS prediction in MPEG-4 lossless audio coding," *Proc. IEEE ICASSP 2006,* vol. 5, pp. 181-184, May 14-19, 2006, Toulouse, France.