ACCOUNTING FOR COMPANDING NONLINEARITIES IN LOSSLESS AUDIO COMPRESSION

Florin Ghido and Ioan Tăbuş

Institute of Signal Processing, Tampere Univ. of Technology, Finland

ABSTRACT

This paper introduces a novel prediction structure for improving the lossless compression ratio, by accounting for companding nonlinearities of different sample-based audio formats. This applies to a wide class of formats including a-law, μ-law, DAT-LP (Digital Audio Tape recorders), DV-LP (Digital Video camcorders), and HDCD (High Definition Compatible Digital). The proposed prediction structure obtains significant compression improvements (8-12%) over traditional linear prediction for a-law, μ-law, DAT-LP, DV-LP and also small compression improvements for HDCD. The improvement in compression can also be used for the detection of nonlinearities in HDCD format, making possible to play HDCD CDs at improved audio resolution in ordinary public domain players.

Index Terms— companding, audio compression, nonlinear prediction, lossless compression, audio formats

1. INTRODUCTION

The state of the art in lossless audio compression [1] [2] [3] is undoubtedly achieved by predictive *linear* methods. There have been a few previous attempts to consider various nonlinear prediction techniques, which could improve the results of the prediction stage but in the overall compression ratio (which includes the model costs additionally to the error costs) the improvements were not conclusive.

Nonlinear *dynamical* relationships are usually difficult to capture with prediction methods of reasonable complexity. If the allowed complexity is high, it becomes possible to consider the powerful multilayer perceptron predictors (MLP), and such a study in [4] found it useful to switch between nonlinear and linear prediction, depending on the local features of the speech material. As a result, segmental SNR improvements of 1 to 2 dB can be obtained when compared to linear prediction alone. In an attempt to utilize a similar nonlinear modeling for lossless audio compression, a less powerful model than MLP was considered in [5], where the predictor is updating two models, a linear one and a simple nonlinear one (a single layer perceptron), and is switching between them according to the best achieved results. The complexity of the method in [5] is not prohibitive for lossless audio compression, but the results shown presented no improvement in the compression rates of typical audio CDs compared to Monkey's Audio [1], although some improvements where shown for the "MPEG test" samples, which are single instrument or voice samples recorded in controlled environments.

In this paper, we consider a specific class of nonlinear *memoryless* models, and our precise goal is to account in the prediction stage for the companding nonlinearities which are present in some audio formats used in transmitting and storing audio files. As an effect of these nonlinearities, the essentially linear predictor used in most audio compressors is only suboptimal, and the compression results can be improved significantly by modifying the structure of the predictor.

Companding is a method of reducing the effects of limited dynamic range of a channel or storage format (the word "companding" was created as a combination of *comp*ressing and exp*anding*) in order to achieve better signal-to-noise ratio or higher dynamic range for a given number of bits. In contrast to audio level compression used in audio recording and sound volume leveling, which is based on a variable gain amplifier and is locally a linear process (quasi constant amplification for short time periods), companding is a nonlinear transformation, applied in the same way at any point in a given recording. In companding, each value is compressed before transmission or storage and is expanded at the receiver or retrieval part.

1.1. A-law and μ-law formats

First introduced as a way to increase signal-to-noise ratio in the analog domain, and later extended to the digital domain, μ-law [6] and a-law [6] are logarithmic companding schemes. μ-law maps 14 bit (or scaled 16 bit) signed integers into one sign bit and 7 bits magnitude (which may be regarded as a floating-point representation having 3 bits exponent and 4 bits mantissa), as in Figure 1. A-law maps 13 bit (or scaled 16 bit) signed integers into one sign bit and 7 bits magnitude (which may be regarded as a floating-point representation having 3 bits exponent and 4 bits mantissa).

1.2. DAT-LP / DV-LP formats

Digital Audio Tape recorders [7] have a Long Play mode, which converts 16 bit linear samples to 12 bit non-linear samples, using a logarithmic companding scheme. DAT-LP maps 16 bit signed integers into 1 sign bit and 11 bits magnitude (which may be regarded as a floating-point representation with 3 bits for the exponent and 8 bits for the mantissa), as in Figure 1. Identically to DAT-LP, new generation personal digital video (DV) camcorders [8], offer a Long Play option to increase the recording time of a cassette from 1 hour to 1.5 hours, and use the 12 bit non-linear sample format of DAT-LP.

1.3. HDCD Compact Disc format

High Definition Compatible Digital [9] is an enhanced Compact Disc compatible format for increasing the resolution of the CD to an equivalent of 19-20 bits, by use of companding for peaks, gain adaptive compression for low amplitude signals, dither with noise shaping, and switchable anti-alias filters.

The format uses in-band signaling to instruct the decoder about the parameters of the restore operations (for the companding and gain adaptive compression) by inserting encrypted command packets in the LSB of the signal. The hidden command packets are around one millisecond long, and are inserted at intervals of several tens of milliseconds. For each HDCD CD, the production engineer decides if the "Peak Extend" (PE) feature will be used (the feature is

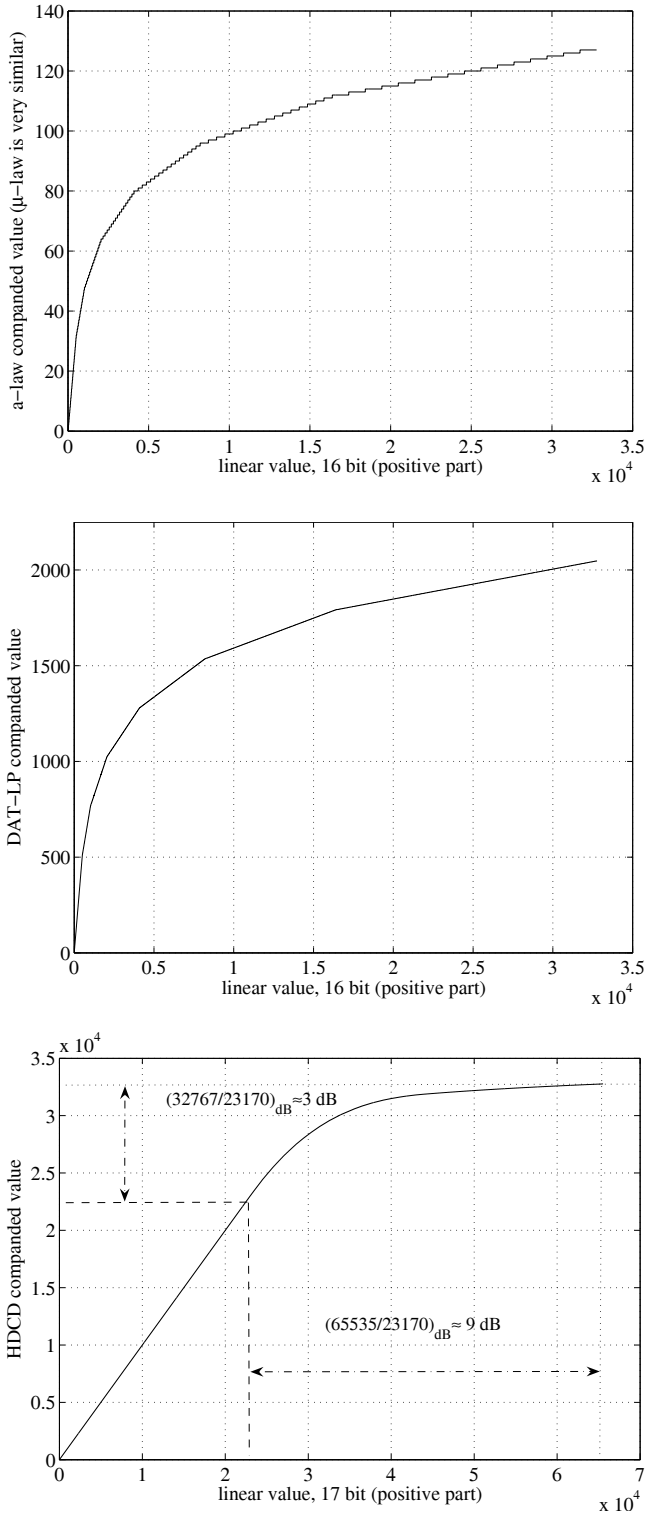

Fig. 1. Companding functions: (Top) a-law and μ-law, (Middle) DAT-LP / DV-LP, and (Bottom) HDCD

turned on or off for the entire CD), and he also adjusts the global volume level, which determines how much companding will be applied. When the PE feature is on, the peaks are shrunk, because the range of top 9 dB of 17 bit signed integers is mapped to the range of top 3 dB of the 16 bit signed integers, as presented in Figure 1, effectively adding up an extra bit of dynamic range.

2. PREDICTION STRUCTURE

In order to simplify the notation, we will work in the following description with a single channel (monophonic) audio signal, extension to stereo or multiple channels being straightforward.

We have access to a digital signed integer signal x_n, represented on CB bits, therefore $-2^{CB-1} \leq x_n \leq 2^{CB-1} - 1$. We assume that x_n was created by applying a companding function, defined as

$$f : [-2^{LB-1}, 2^{LB-1} - 1] \rightarrow [-2^{CB-1}, 2^{CB-1} - 1] \quad (1)$$

to the ideal (linearly generated) audio integer signal y_n, represented on LB bits, with $LB \geq CB$. The inverse of f is defined as

$$f^{-1} : [-2^{CB-1}, 2^{CB-1} - 1] \rightarrow [-2^{LB-1}, 2^{LB-1} - 1], \quad (2)$$

which maps back CB bits signed integers into LB bits signed integers. The two functions may be implemented algorithmically if there exist simple operations to compute the result (as for μ-law, a-law, DAT-LP / DV-LP), or by means of look-up tables (as for HDCD).

Modeling of y_n (the signal in the ideal linear domain) may be done efficiently using an order K linear predictor, estimated with Levinson-Durbin or the (recursive) least squares method, to obtain the predicted value $\tilde{y}_n$ as

$$\tilde{y}_n = \sum_{i=1}^{K} c_{i,n} y_{n-i}. \quad (3)$$

However, for x_n (the signal in the companded domain), using a linear predictor will not produce good results, because the linear relations between neighboring samples are severely affected by the companding procedure. Since we assume that the companding function is known, we can map x_{n-i} to the original linear domain to obtain y_{n-i}, compute the linear prediction $\tilde{y}_n$, and then map back the prediction to the companded domain prediction $\tilde{x}_n$ as

$$\tilde{x}_n = f\left(\sum_{i=1}^{K} c_{i,n} f^{-1}\left(x_{n-i}\right)\right). \quad (4)$$

The most relevant criterion to be minimized is

$$J_1(\mathbf{x}, \mathbf{c_n}) = \sum_n \left(x_n - f\left(\sum_{i=1}^{K} c_{i,n} f^{-1}(x_{n-i})\right)\right)^2, \quad (5)$$

which can be easily utilized in a coder based on gradient methods for updating, like the one in [1]. For coders based on Levinson-Durbin or the (recursive) least squares method, as it is [3], the alternative is to use a different criterion, which is linear in parameters

$$J_2(\mathbf{x}, \mathbf{c_n}) = \sum_n \left(f^{-1}(x_n) - \sum_{i=1}^{K} c_{i,n} f^{-1}(x_{n-i})\right)^2, \quad (6)$$

and works entirely with values in the linear domain.

For μ-law and a-law, a special preprocessing step had to be done, in order to make compression possible. To facilitate clock recovery (which may be impeded by transmission of long periods of silence, therefore zeros), the procedure specified in the standard was to apply a bitwise XOR operation for the companded result (composed of sign s and magnitude M) with 0xFF and 0xD5 respectively, so that silence periods will not produce a sequence of zeros (difficult for synchronization), but values containing non-zero bits instead.

After restoring the companded result by applying the bitwise XOR, we further convert it from the sign and magnitude representation to an unsigned 8 bit value val, by the mapping

$$val = \begin{cases} 128 + M & \text{if } s = 0 \\ 128 - M - 1 & \text{if } s = 1 \end{cases}.$$

The value -0 (which is represented by $s = 1$ and $M = 0$) is mapped to 127, -1 is mapped to 126, and so on, in order to keep the mapping bijective. If we would map both $+0$ and -0 to 128, the inverse mapping would lose -0.

For DAT-LP, the preprocessing step was to compand back the 16 bit decoded linear values to 12 bit nonlinear, stored in 16 bit containers. This was necessary, because both DAT and DV devices transfer to the computer 16 bit decoded linear values.

We modified the existing OptimFROG lossless audio compressor [10] in the following way. In the implementation, the code (written here in a pseudocode notation similar to C++) which computed the prediction for the sample x_n and produced the error e_n was

```
e[n] = x[n] - sliding_rls.predict();
sliding_rls.update(x[n]);
```

and was modified to

```
e[n] = x[n] - compand(sliding_rls.predict());
sliding_rls.update(decompand(x[n]));
```

For each sample format, the $compand$ and $decompand$ functions were expanded as look-up tables, therefore the time increase compared to the normal code was negligible.

3. EXPERIMENTAL RESULTS

For the tests on DV-LP, a-law, and μ-law we made use of a reduced audio corpus consisting of 80 files of one minute length, 44100 Hz, 16 bit, stereo (obtained by extracting the middle minute of track 3 of each CD, from a large audio corpus of 80 audio CDs). We converted the samples to DV-LP, a-law, and μ-law, conform to the corresponding standards. For DV-LP, we verified the conformance and exactness of our conversion by digitally transferring one of the samples to a DV camcorder and then back to the computer.

We compared the proposed prediction structure, implemented as a modified OptimFROG encoder (identified by OFR-NEW), with the existing OptimFROG stable version 4.520b1 [3] (identified by OFR-OLD), and with the newly standardized MPEG-4 ALS version RM17 [2] (identified by ALS-V17), on the following cases (tests)

1. ALS-V17 and OFR-OLD for 16 bit decoded DV-LP
2. ALS-V17, OFR-OLD, and OFR-NEW for 12 bit DV-LP
3. ALS-V17, OFR-OLD, and OFR-NEW for 8 bit prep. a-law
4. ALS-V17, OFR-OLD, and OFR-NEW for 8 bit prep. μ-law
5. OFR-OLD and OFR-NEW for 16 bit HDCD (5 CDs)

The compression options used for testing were, for ALS-V17 "-7 -p -t2" (maximum possible asymmetrical compression), and for both OFR-NEW and OFR-OLD "–mode extra" (in order to match the ALS-V17 average compression level for the test 1). For test 1, which is not shown on the graphs, ALS-V17 and OFR-OLD obtain very similar compression, presented in Table 1, but ALS-V17 is 34 times slower at encoding and 1.6 times slower at decoding.

For test 2, presented in Table 2 and detailed in Figure 2, the proposed method obtains an improvement of 8.4%; for test 3, presented in Table 3 and detailed in Figure 3, there is an improvement of 10.8%; for test 4, presented in Table 4 and detailed in Figure 4, there is an improvement of 12.5%. Although ALS-V17 obtains very similar compression compared to OFR-OLD on test 1, its compression deteriorates significantly (and decompression speed increases) compared to OFR-OLD on the other tests, with 2.3% worse compression for test 2, 4.0% for test 3, and 4.8% for test 4.

The only software decoder for HDCD enhanced CDs is Windows Media Player 10, and the detailed specification for decoding HDCD is not publicly available. HDCD CDs without "Peak Extend" (PE) can produce at most half scale peaks, and those with PE are able to produce full scale peaks. We played each of the 5 HDCD test CDs with WMP 10, and recorded analogically its output, to see if the PE was enabled or not. We found two CDs without PE, "The Doors;

Test 1	*Compression (%)*	*Enc. time (s)*	*Dec. time (s)*
ALS-V17	59.7	19911	675
OFR-OLD	59.4	575	417

Table 1. Overall results for 16 bit decoded DV-LP

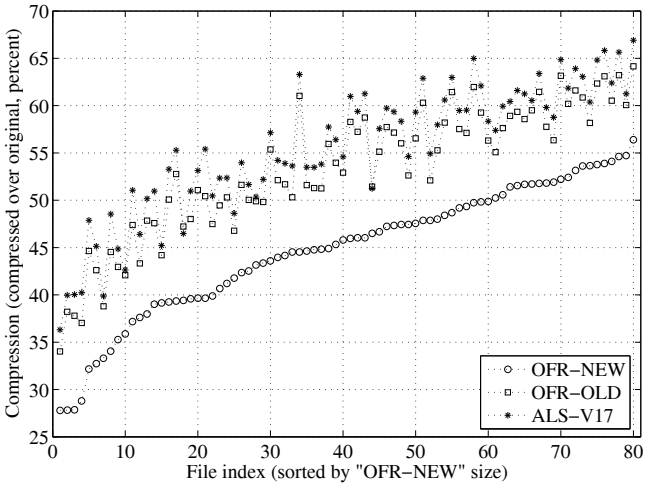

Fig. 2. Compression for 80 files in format 12 bit DV-LP

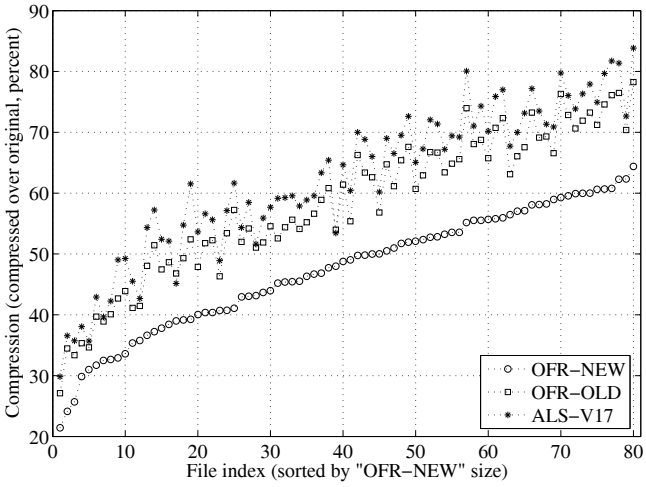

Fig. 3. Compression for 80 files in format 8 bit a-law

Test 2	*Compression (%)*	*Enc. time (s)*	*Dec. time (s)*
ALS-V17	55.5	16698	328
OFR-OLD	53.2	578	416
OFR-NEW	44.8	614	439

Table 2. Overall results for 12 bit DV-LP

Test 3	*Compression (%)*	*Enc. time (s)*	*Dec. time (s)*
ALS-V17	62.2	15244	299
OFR-OLD	58.2	551	403
OFR-NEW	47.4	561	408

Table 3. Overall results for 8 bit a-law

Test 4	*Compression (%)*	*Enc. time (s)*	*Dec. time (s)*
ALS-V17	64.6	15054	270
OFR-OLD	59.8	552	402
OFR-NEW	47.3	563	406

Table 4. Overall results for 8 bit μ-law

The Soft Parade; remastered 2000" (CD 1) and "The Doors; L.A. Woman; remastered 2000" (CD 2), and three CDs with PE "Yes; Yessongs; remastered 2001", pack of three CDs (CD 3/4/5).

Figure 5 presents the compression improvements of OFR-NEW compared with OFR-OLD. As expected, the CDs without PE gave worse results and the CDs with PE gave better results. Even in this case when the improvements are only marginal, an extra benefit is the possibility to detect the use of the hidden format HDCD and subsequently to improve the quality of the restored audio, even for normal decompressors, which can not decipher the proprietary information about PE being on or off.

4. CONCLUSION

We have presented a novel prediction structure for improving the compression of files in a variety of sample-based audio formats. The proposed prediction structure obtains significant compression improvements (in the range of 8-12%) compared with the existing state of the art, with very small complexity increase, for audio in formats like a-law, μ-law, and DV-LP, and also small compression improvements for HDCD. We are also investigating adaptive algorithms to estimate an unknown parametric companding curve for compression of digital audio signals recorded from analog support (like tape, where strong nonlinearities are induced by the magnetic saturation at extreme levels).

5. REFERENCES

[1] M.T. Ashland, "Monkey's Audio lossless audio compressor," on Internet, at *http://www.monkeysaudio.com/*, February 2006, version 4.01b2.

[2] ISO/IEC, "ISO/IEC 14496-3:2005/Amd 2:2006, Audio Lossless Coding (ALS), new audio profiles and BSAC extensions," on Internet, at *http://www.nue.tu-berlin.de/mp4als*, April 2006, reference software version RM17.

[3] F. Ghido, "OptimFROG lossless audio compressor," on Internet, at *http://www.LosslessAudio.org/*, April 2006, version 4.520b1.

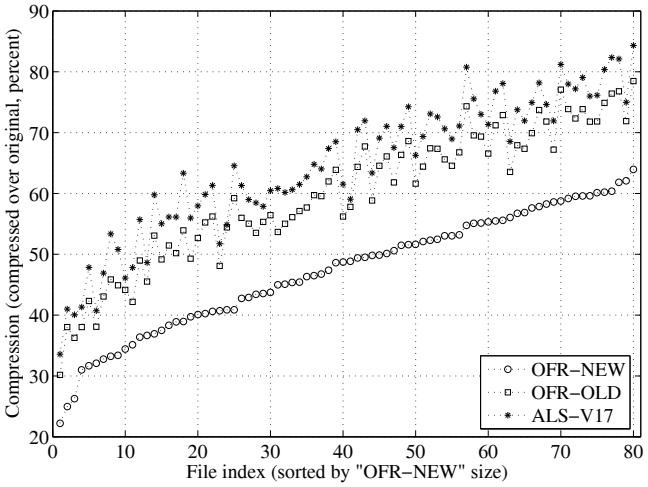

Fig. 4. Compression for 80 files in format 8 bit μ-law

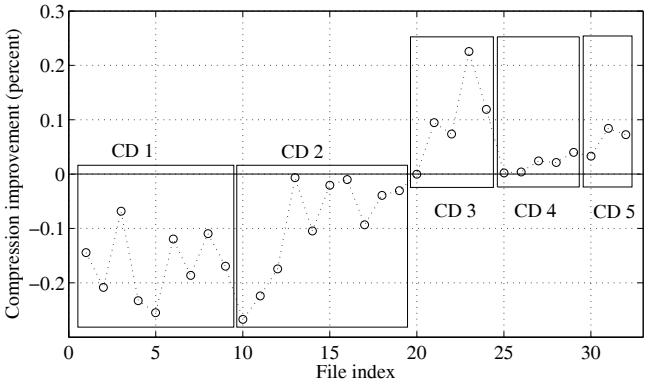

Fig. 5. Comparison for HDCDs (only the CD's labeled 3,4, and 5 are in companded format, with the option Peak Extend set to "on")

[4] M. Faundez-Zanuy, F. Vallverdu, and E. Monte, "Nonlinear prediction with neural nets in ADPCM," in *ICASSP 1998 Proceedings*. IEEE, May 1998, vol. 1, pp. 345–348.

[5] E. Ravelli, P. Gournay, and R. Lefebvre, "A two-stage MLP+NLMS lossless coder for stereo audio," in *ICASSP 2006 Proceedings*. IEEE, May 2006, vol. 5, pp. 177–180.

[6] ITU-T, "Pulse code modulation (PCM) of voice frequencies, Recommendation G.711," November 1988.

[7] IEC, "IEC 61119-1, Digital audio tape cassette system (DAT) - Part 1: Dimensions and characteristics," November 1992.

[8] IEC, "IEC 61834, Helical-scan digital video cassette recording system using 6,35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems)," August 1998.

[9] K.O. Johnson and M.W. Pfaumer, "Compatible resolution enhancement in digital audio systems," in *AES 101th Convention Proceedings*, October 1996, preprint number 4392.

[10] F. Ghido, "An asymptotically optimal predictor for stereo lossless audio compression," in *DCC 2003 Proceedings*. IEEE, March 2003, p. 429.