A Two-Stage MLP+NLMS Lossless Coder
For Stereo Audio

Emmanuel Ravelli, Philippe Gournay and Roch Lefebvre

Abstract
We propose in this paper a two-stage lossless audio coder for stereo signals. The first stage is based on a hybrid approach that uses either a stereo linear prediction or a multilayer perceptron-based (MLP) nonlinear prediction. The second stage is based on two cascaded Normalized Least Mean Square (NLMS) filters that remove the remaining redundancy in the residuals of the stereo prediction. The obtained compression ratios are equal or superior to the best state-of-the-art coders.

Background
State-of-the-art lossless audio coders use linear prediction to remove redundancy from an audio signal. Three good references are:
1. Cambridge et al, AES, 1993 with a technique called joint-channel linear prediction that takes advantage of the correlation that exists between the two channels of a stereo signal.
2. Schuller et al, ICASSP, 2001 with a lossless audio coder based on cascaded adaptive linear prediction. In this paper, the prediction was performed directly on an input mono signal.
3. Garcia et al, AES, 2004 in which the authors showed that the coefficients of a joint-channel linear prediction can’t be quantized using efficient techniques. To get around the transmission of the coefficients, a backward approach was proposed.

In our approach we take ideas from previous methods but combine them in novel ways. We also introduce a new approach inspired by (Faundez, EUSIPCO, 2000) and based on neural networks.

First stage: hybrid linear/MLP stereo prediction
The first stage is based on a hybrid approach using, for each frame and for each channel, either a joint-channel linear prediction or a joint-channel nonlinear prediction; the model which maximizes the prediction gain is chosen.

Linear prediction
The linear prediction used is the same as in (Garcia, AES, 2004)—the analysis is performed on an extended frame weighted with an asymmetric window. The Cholesky decomposition is used to invert the correlation matrices.

Nonlinear prediction
The joint-channel nonlinear prediction is based on two Multi-Layer Perceptron networks, one for each channel. To initialize the weights and bias of the neural network, we use the neural network parameters computed from the previous frame. The Levineberg-Marquardt is used to train the neural networks.

Hybrid approach
For each frame and for each channel, two analyses are performed (one linear and one nonlinear), the prediction gains obtained with each analysis are then calculated, and finally, for each channel, the best predictor is chosen. In order for the decoder to know which prediction was used for each channel, two bits per frame are sent to the decoder.

Backward analysis
The predictor coefficients are estimated from the past decoded signal, which is available at both the encoder and the decoder. The backward analysis has the advantage to avoid the quantization and the transmission of the prediction coefficients as in the forward approach. To better adapt the model to the signal, a “subframe” analysis is performed.

Experimental results: prediction gains
To limit the complexity of the system, we found that the following values are a good compromise: an autoregression order of \( p = 4 \) and an autoregressive order of \( p = 6 \). The frame length is 20 ms (N=960 at 48 kHz). The analysis is performed four times per frame, leading to a 5 ms subframe analysis. The signals used for these experiments are taken from an MPEG test base, and all are sampled at 48 kHz.

First stage: hybrid linear/MLP stereo prediction

<table>
<thead>
<tr>
<th></th>
<th>Speech</th>
<th>Linear</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left channel</td>
<td>26.54 dB</td>
<td>26.78 dB</td>
<td>27.20 dB</td>
</tr>
<tr>
<td>Right channel</td>
<td>26.14 dB</td>
<td>26.49 dB</td>
<td>26.85 dB</td>
</tr>
<tr>
<td>Pop music</td>
<td>24.64 dB</td>
<td>25.15 dB</td>
<td>25.55 dB</td>
</tr>
<tr>
<td>Classical music</td>
<td>24.12 dB</td>
<td>24.71 dB</td>
<td>25.11 dB</td>
</tr>
</tbody>
</table>

Second Stage: Cascaded NLMS Filtering

Algorithm | Left channel | Right channel | Left channel | Right channel |
----------|--------------|---------------|--------------|---------------|

Experimental results: compression ratios
We compared the compression ratios of four lossless codecs. Two state-of-the-art codecs, LPAC and Monkey’s Audio. A third one, Codec A, based only on linear prediction. And the algorithm proposed in this paper, named Codec B. The tests were performed with 6 tracks of different styles of music and also an MPEG test base.

Conclusion
- First stage: We have shown that our hybrid approach clearly outperforms the backward joint-channel linear prediction used in (Garcia et al, AES, 2004).
- Second stage: We have shown that this second stage greatly improves the performance of the first stage. The obtained compression ratios are equal or superior to the best state-of-the-art coders.
- Future work: Possible future work would be to investigate other neural network architectures, such as radial basis function networks (RBF) or pipelined recurrent neural networks (PRNN).