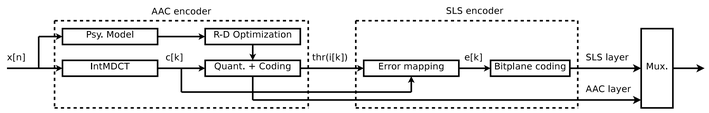
Joint Optimization of the Perceptual Core and Lossless Compression Layers in Scalable Audio Coding

Emmanuel Ravelli, Vinay Melkote, Tejaswi Nanjundaswamy, and Kenneth Rose

Signal Compression Lab, Department of ECE, University of California, Santa Barbara

MPEG-4 HD-AAC

- A scalable-to-lossless (SLS) codec with perceptual (lossy) AAC core + fine-grained enhancements (cumulatively lossless)
- Integrated solution for playback, streaming, archiving, etc.



Current encoding approach

- First, optimize AAC base-layer. Subsequently, choose SLS encoding parameters.
- Separate optimization ignores inter-layer dependence.
- Optimal AAC layer $\Rightarrow$ good perceptual quality, but at cost to lossless compression
- SLS non-core $\Rightarrow$ optimal lossless compression, but poor perceptual quality at low bit-rates

Proposed method

- Optimal selection of AAC parameters with an explicit accounting of effect on subsequent SLS coding
- A framework for trade-off between AAC perceptual quality and SLS lossless compression
- R-D optimization formulated as:

$$\mathcal{P}^* = \arg \min_{\mathcal{P}:\mathcal{D}_b(\mathcal{P})\leq\mathcal{D}_t} \mathcal{R}_b(\mathcal{P}) + \alpha\mathcal{R}_e(\mathcal{P})$$

- where
  - $\mathcal{P}$: set of AAC parameters - scalefactors (SFs) + Huffman codebooks (HCBs) - for all frames
  - $\mathcal{R}_b(\mathcal{P})$: AAC bit-rate given parameters $\mathcal{P}$
  - $\mathcal{R}_e(\mathcal{P})$: bit-rate for the optimal SLS layer when the base-layer is encoded with parameters $\mathcal{P}$
  - $\mathcal{D}_b(\mathcal{P})$: achieved distortion in the AAC layer
  - $\mathcal{D}_t$: target distortion in the AAC layer
  - α : parameter that controls the AAC-SLS performance trade-off
- Definition of $\mathcal{R}_e(\mathcal{P})$ subsumes optimization over SLS parameters $\Rightarrow$ *jointly optimized AAC and SLS layers*.

Distortion measures

- Average-Average NMR (AANMR):

$$\mathcal{D}_b^A(\mathcal{P}) = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{L}\sum_{l=0}^{L-1} d_n^l(s_n^l)$$

- Maximum-Maximum NMR (MMNMR):

$$\mathcal{D}_b^M(\mathcal{P}) = \max_{0\leq n<N}\max_{0\leq l<L} d_n^l(s_n^l)$$

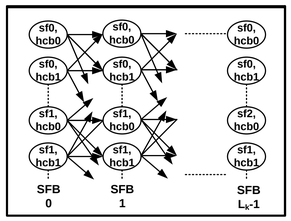
- $d_n^l(s_n^l)$: NMR in scalefactor band *l* of frame *n*, when corresponding scalefactor is s_n^l

MMNMR solution

- Equivalent per SFB distortion constraint: $d_n^l(s_n^l) \leq \mathcal{D}_t$
- Equivalent problem (for each frame n):

$$\mathrm{p}_n^* = \arg \min_{\mathrm{p}:d_n^l(s^l)\leq\mathcal{D}_t} \mathrm{R}_b^n(\mathrm{p}) + \alpha\mathrm{R}_e^n(\mathrm{p})$$

- where p, $\mathrm{R}_b^n(\mathrm{p})$, and $\mathrm{R}_e^n(\mathrm{p})$ represent, respectively, per-frame AAC parameters, AAC bit-rate, and SLS bit-rate.
- Computationally efficient search for p_n^* via trellis-based approach
- Stages of the trellis: scalefactor bands
- Nodal parameters : AAC SCFs and HCBs, and SLS MSB-planes and optimal lazy bit-planes



- Nodal costs: AAC distortion, and bits for encoding AAC spectral data, SLS residual, and lazy bit-plane parameters
- Transition costs: bits for encoding SCFs and HCBs (in AAC), and MSB-planes (in SLS)
- Best path via Viterbi algorithm

AANMR solution

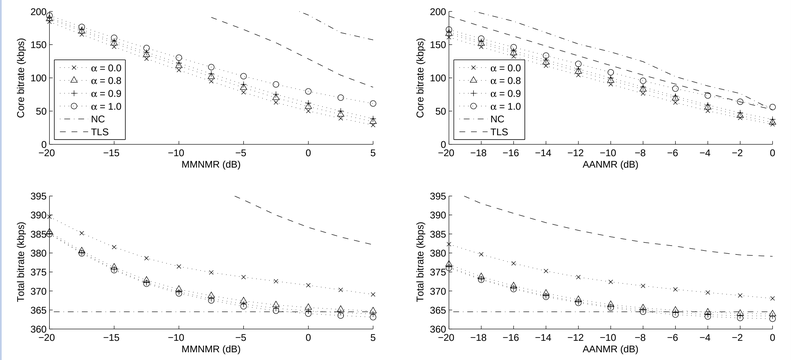
- Follows the Lagrangian technique: minimize

$$\mathrm{J}_n(\mathrm{p}_n, \lambda) = \mathrm{R}_b^n(\mathrm{p}_n) + \alpha\mathrm{R}_e^n(\mathrm{p}_n) + \lambda\left(\frac{1}{L}\sum_{l=0}^{L-1} d_n^l(s_n^l)\right)$$

- Solution via similar trellis-based search
- Lagrange parameter λ adjusted till $\mathcal{D}_t$ achieved

Evaluation

- Codecs compared: SLS non-core (NC), TLS-based HD-AAC, proposed coding scheme
- Test set: 15 audio files, 30s each, mono, 48 kHz
- R-D curves with distortion-constrained (AANMR or MMNMR) base-layer at different α values



- Objective quality (PEAQ), total bit-rate, and compression ratio evaluated for rate-constrained base-layer at 64kbps

	PEAQ ODG	Total bit-rate	Compr. ratio
MMNMR α=0.00	-0.942	368.7	2.083
MMNMR α=0.92	-0.957	361.3	2.125
MMNMR α=1.00	-1.919	360.0	2.133
AANMR α=0.00	-0.885	367.0	2.093
AANMR α=0.79	-1.125	361.4	2.125
AANMR α=1.00	-2.060	359.7	2.135
TLS	-1.676	376.3	2.041
SLS non-core	-1.923	361.4	2.125

Conclusions

- A novel trellis-based algorithm for joint optimization of SLS and AAC encoding parameters in HD-AAC is proposed
- Facilitates a trade-off between lossless compression ratio and perceptual quality at intermediate bit-rates
- Contrary to popular opinion, inclusion of AAC core need not preclude excellent lossless compression performance

http://www.scl.ece.ucsb.edu/

{ravelli, melkote, tejaswi, rose}@ece.ucsb.edu