Cross-Layer Rate-Distortion Optimization for Scalable Advanced Audio Coding

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

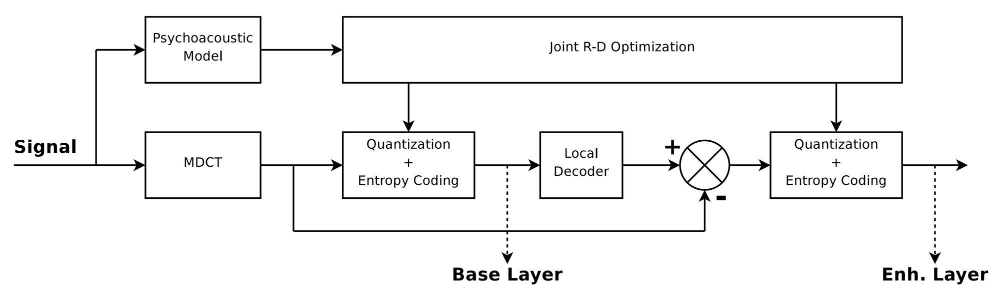
Signal Compression Lab, Department of ECE, University of California, Santa Barbara

Two-layer Scalable AAC

- Common encoding approach: First, optimize base-layer. Subsequently, optimize enhancement-layer.
- Separate optimization ignores inter-layer dependence.
- $\Rightarrow$ Enhancement-layer significantly worse than a non-scalable AAC encoder.

Proposed encoding approach

- Joint optimization of both layers.
- Allows to trade-off performance of the two layers.



Rate-Distortion optimization problems

- Prob. 1: main problem with bitrate constraints

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} (1-\beta)\mathcal{D}^{(b)}(\mathcal{P}) + \beta\mathcal{D}^{(e)}(\mathcal{P})$$
$$\text{s.t. } \mathcal{R}^{(b)}(\mathcal{P}) \leq \mathcal{R}_t^{(b)}$$
$$\text{and } \mathcal{R}^{(e)}(\mathcal{P}) \leq \mathcal{R}_t^{(e)}$$

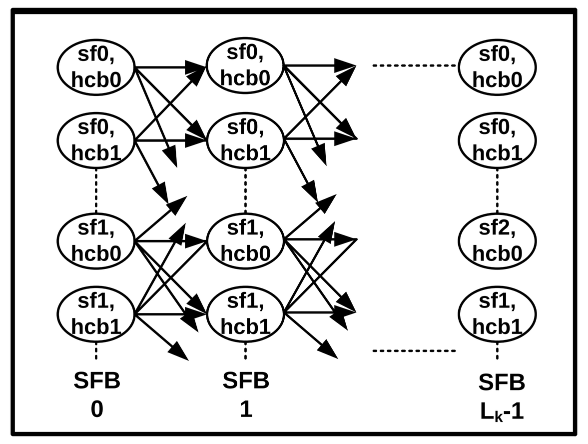
- Prob. 2: intermediate problem with distortion constraints

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} (1-\alpha)\mathcal{R}^{(b)}(\mathcal{P}) + \alpha\mathcal{R}^{(e)}(\mathcal{P})$$
$$\text{s.t. } \mathcal{D}^{(b)}(\mathcal{P}) \leq \mathcal{D}_t^{(b)}$$
$$\text{and } \mathcal{D}^{(e)}(\mathcal{P}) \leq \mathcal{D}_t^{(e)}$$

- where
  - $\mathcal{P}$: set of AAC parameters - scalefactors (SFs) + Huffman codebooks (HCBs) - for both layers and all frames
  - $\mathcal{R}^{(b)}(\mathcal{P})$, $\mathcal{R}^{(e)}(\mathcal{P})$: bitrate in each layer
  - $\mathcal{D}^{(b)}(\mathcal{P})$, $\mathcal{D}^{(e)}(\mathcal{P})$: distortion in each layer, the distortion measure is the Maximum-Maximum NMR (MMNMR).
  - $\mathcal{R}_t^{(b)}$, $\mathcal{R}_t^{(e)}$: target bitrate in each layer
  - $\mathcal{D}_t^{(b)}$, $\mathcal{D}_t^{(e)}$: target distortion in each layer
  - α, β : parameters that control the performance trade-off

Optimal solution for the single layer case

- Reference: Melkote et al, IEEE TASLP 2010.
- Solve Prob. 2 with trellis-based search:
  1. Stages of the trellis: SF bands.
  2. Nodal parameters : SCFs + HCBs.
  3. Nodal costs : spectral data bits.
  4. Transition costs : differential bits for encoding SCFs and HCBs.
  5. Best path via Viterbi algorithm.



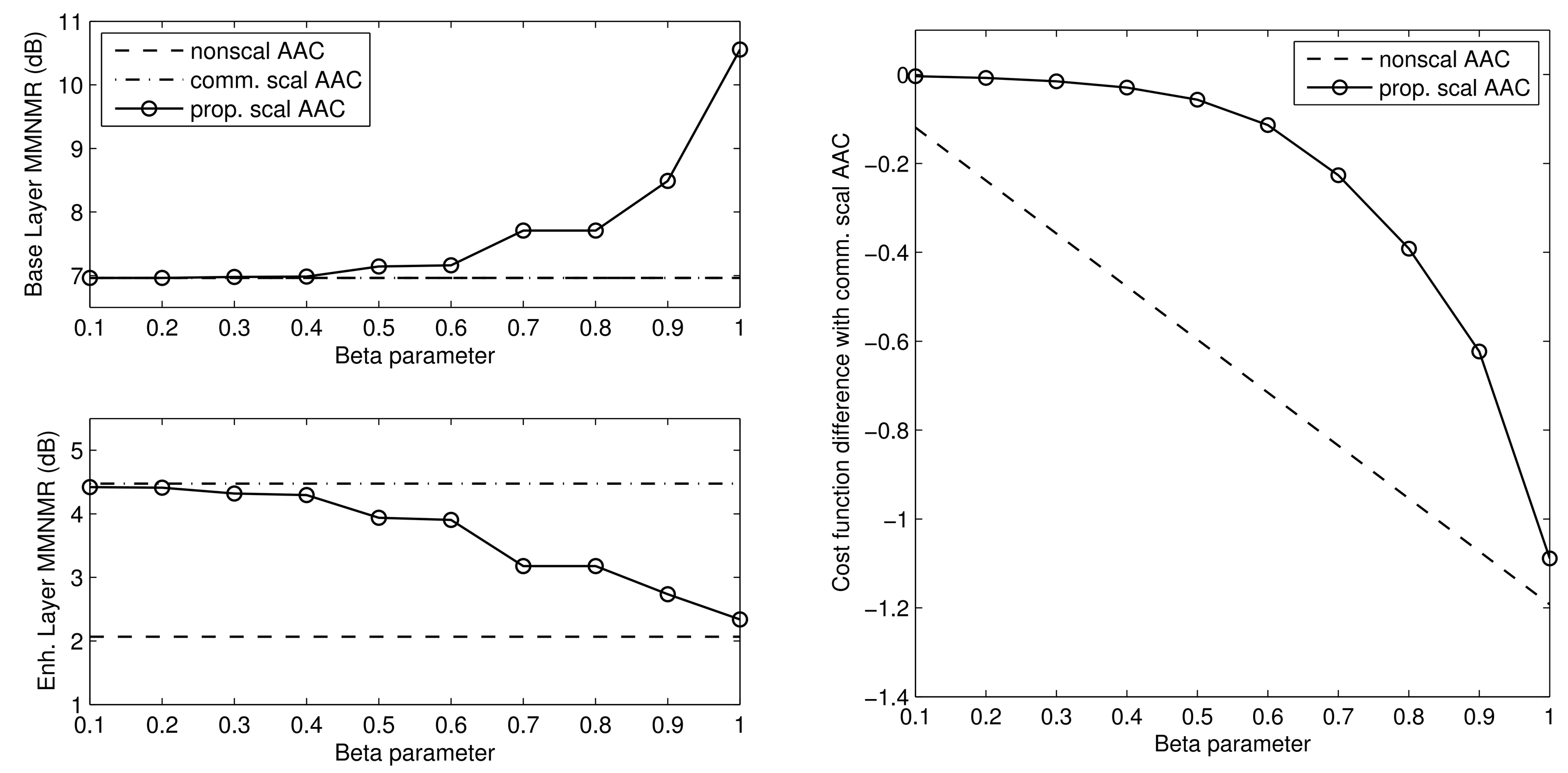
- Solve Prob. 1: repeat search for different distortion constraints, until bitrate constraint achieved.

Proposed solution for the two-layer case

- Direct extension of the optimal trellis-based method: solve Prob. 2 theoretically possible, but impractical complexity.
- Proposed method: iterative approach, computationally efficient alternative, though suboptimal.
- Proposed solution for Prob. 2:
  1. Perform a preliminary base layer optimization.
  2. Optimize the enhancement layer, given the base layer.
  3. Optimize the base layer, given the enhancement layer.
  4. Repeat steps 2. and 3. till a stopping criterion is met.
- Note that steps 2. and 3. only marginally refines the initial guess calculated in step 1. $\Rightarrow$ only few iterations needed.
- Preliminary base layer optimization:
  1. For each node in the base-layer trellis: choose SCF and HCB in the enhancement layer s.t. spectral data bits is minimized, and distortion constraint satisfied.
  2. Integrate corresponding nodal and transition costs, weighted by the parameter α.
- Proposed solution for Prob. 1:
  1. Repeat iterative optimization for different distortion constraints and α values.
  2. Keep solution which satisfy bitrate constraints, and minimize distortion-based cost function.

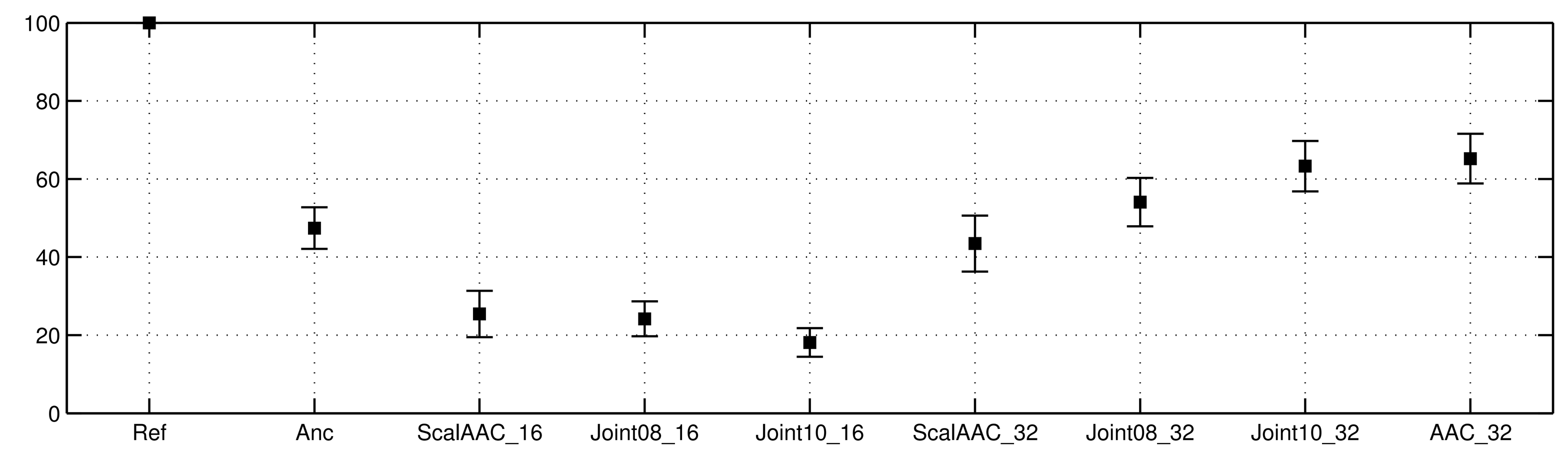
Evaluation

- Coders compared: common scalable AAC coder, proposed scalable AAC coder, non-scalable AAC coder.
- Test set: 4 MPEG audio files, 5s each, mono, 48 kHz.
- Bitrates: base at 16kbps, enhancement at 32kbps.
- Distortions and cost function at different β values:

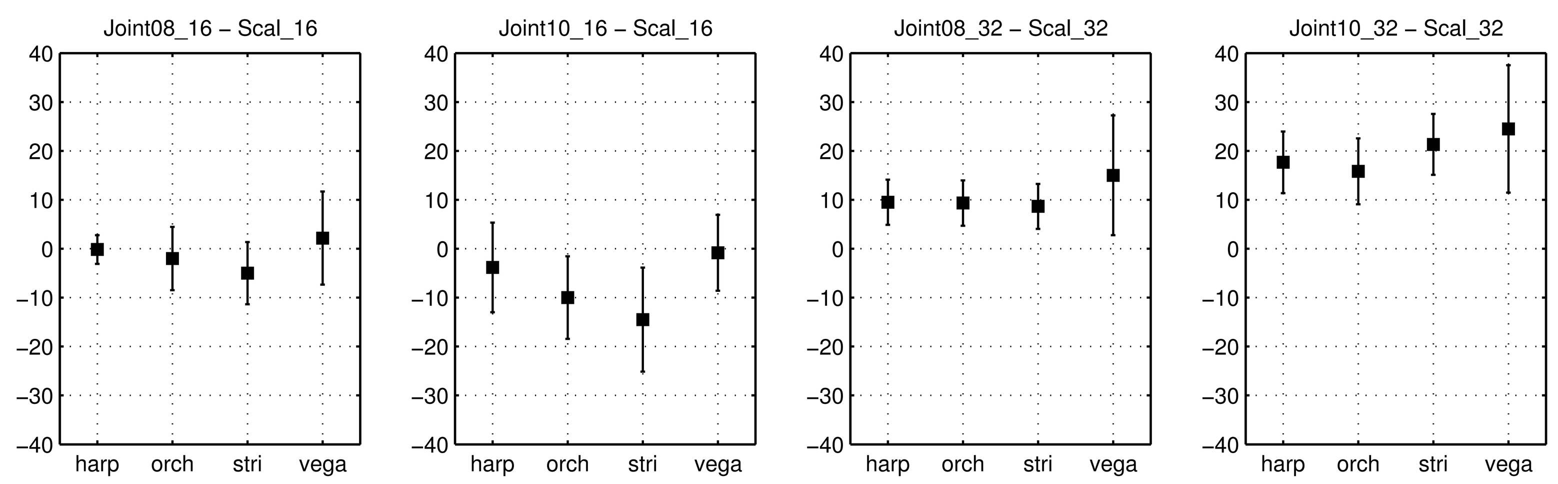


- MUSHRA listening test for 2 different β values (0.8 and 1.0):

Average scores for all items:



Average differential scores per item:



- Conclusion: the proposed approach gives a better trade-off between the two layers performance, than the commonly employed approach.

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