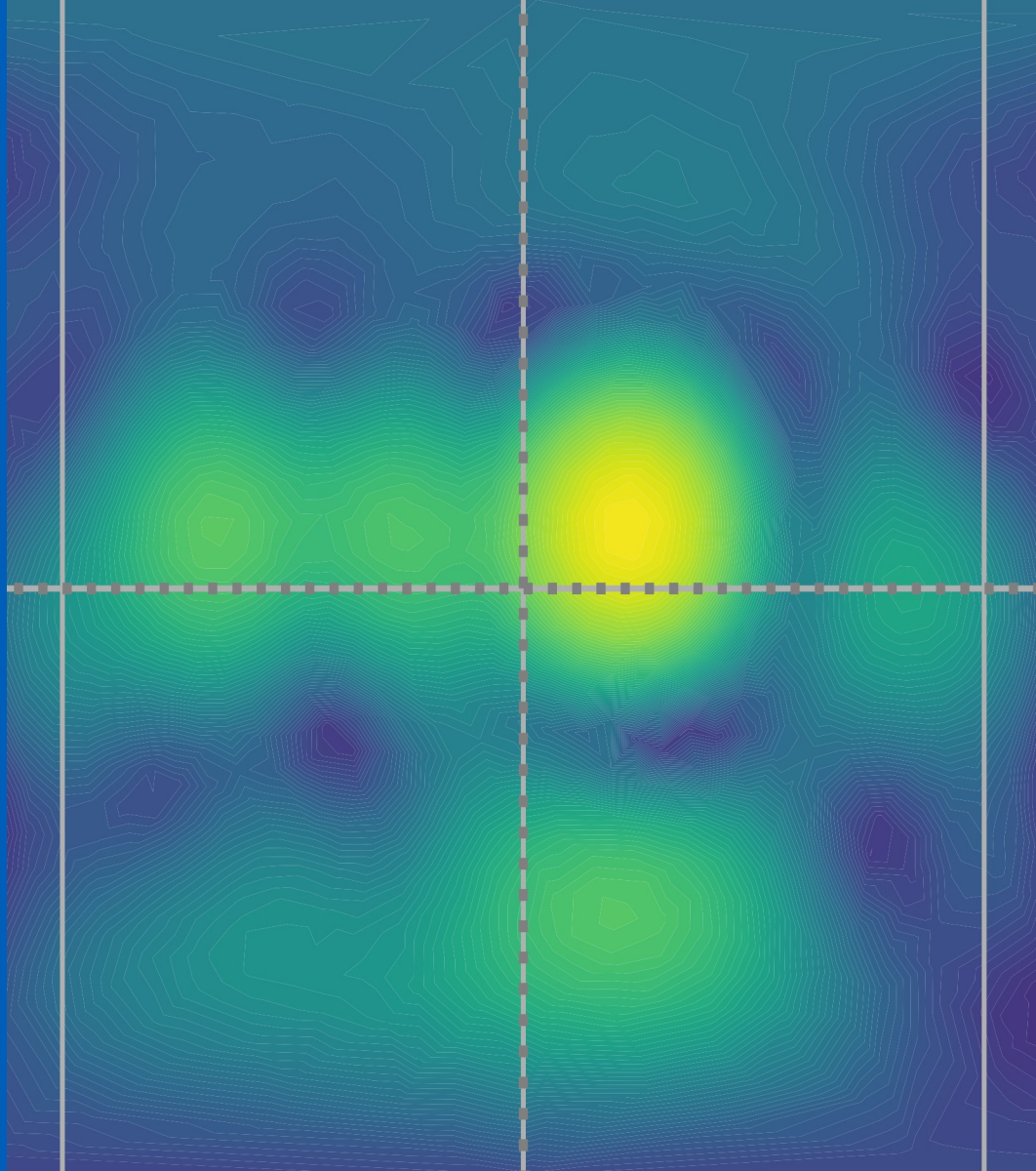
Parametric Spatial Audio Compression

Using Higher-Order-Directional Audio Coding

Christoph Hold

25.03.2024

Aalto-yliopisto
Aalto-universitetet
Aalto University



Outline

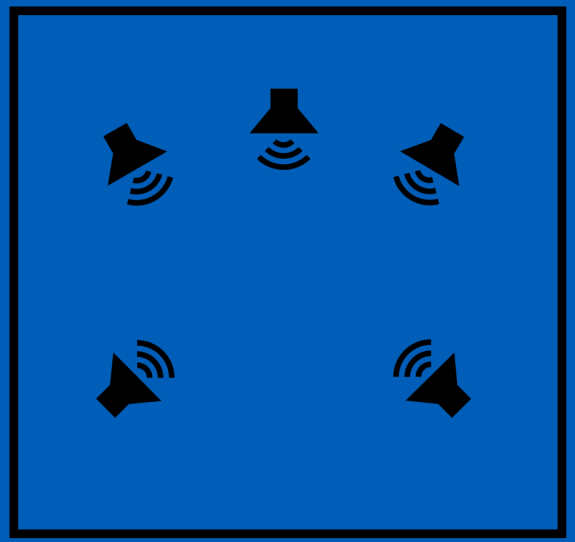
Introduce Parametric Spatial Audio

Introduce a Spatial Audio Codec

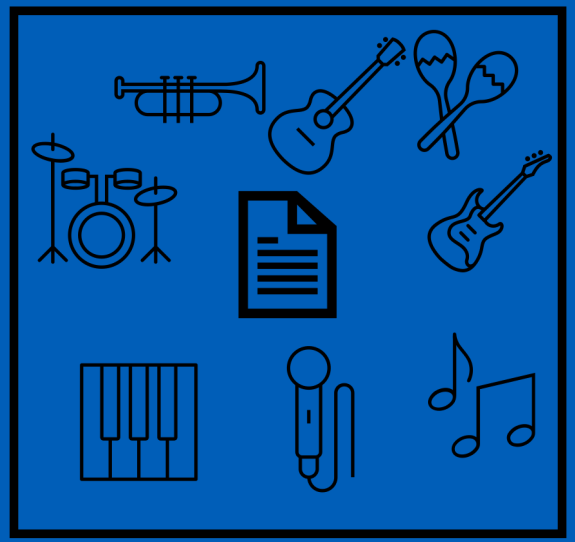
Evaluation

Spatial Audio Formats

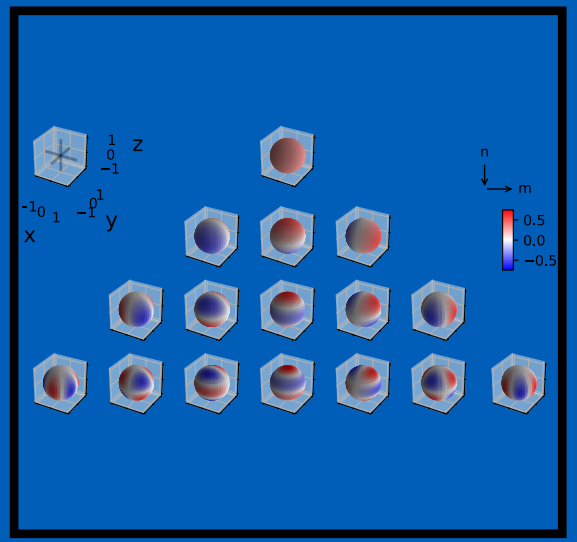
Channel



Object



Scene

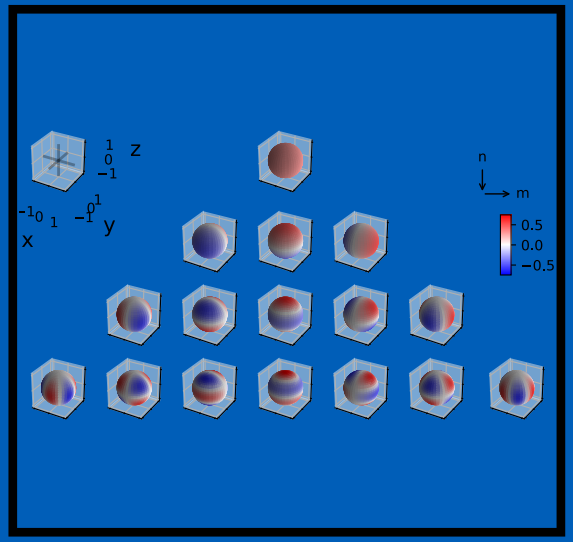


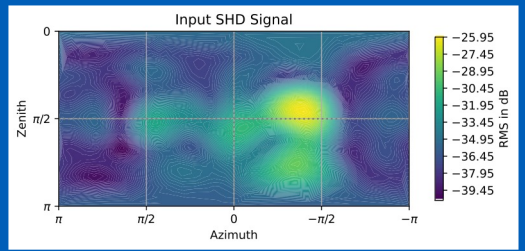
Aalto-yliopisto
Aalto-universitetet
Aalto University

Spatial Audio - Ambisonics

Input

Scene



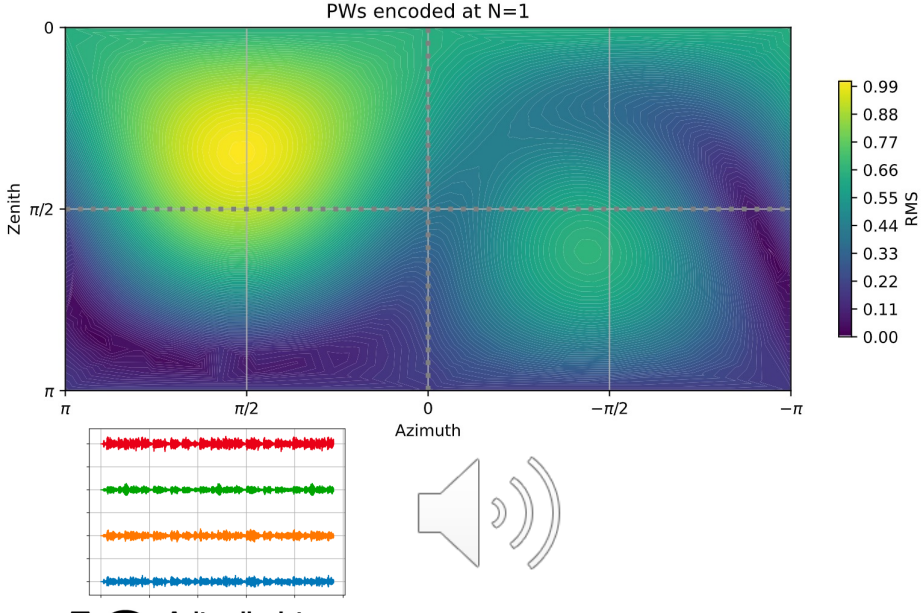


Output

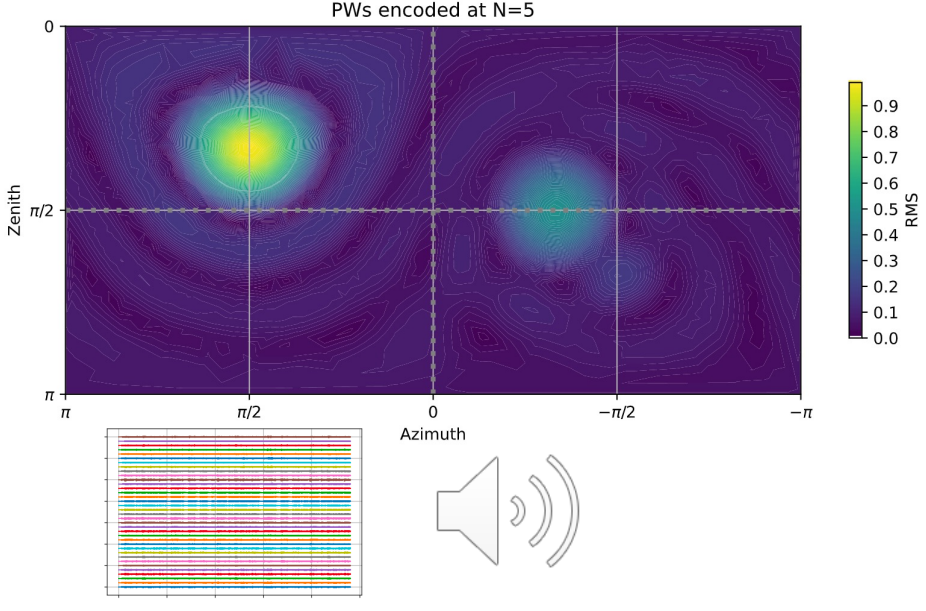
Aalto-yliopisto
Aalto-universitetet
Aalto University

First Order vs. Higher Order

- 4 Audio channels



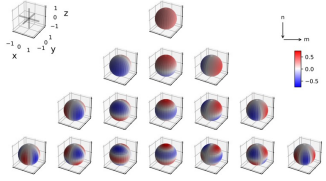
- 36 Audio channels

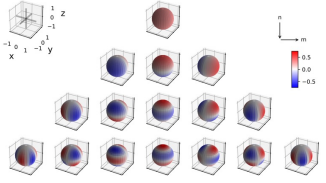


Aalto-yliopisto
Aalto-universitetet
Aalto University

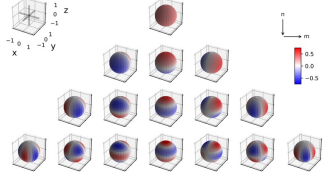
Opus Ambisonics Codec

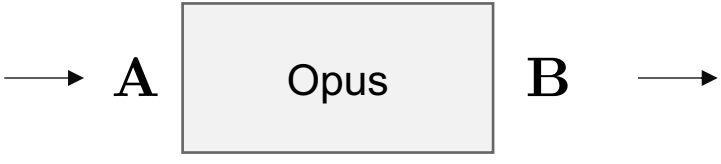
Opus CMF 2

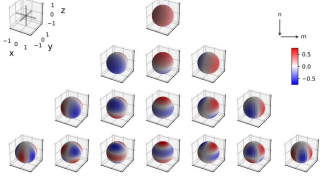




Opus CMF 3







https://www.opus-codec.org/

Transmitting Higher-Order-Ambisonics

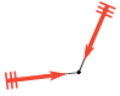
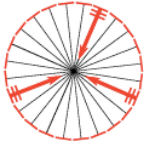
Input (MagLS5)

Opus @ 768kbit/s (MagLS5)

Parametric Spatial Audio

Extract and utilize additional information from the input signals.

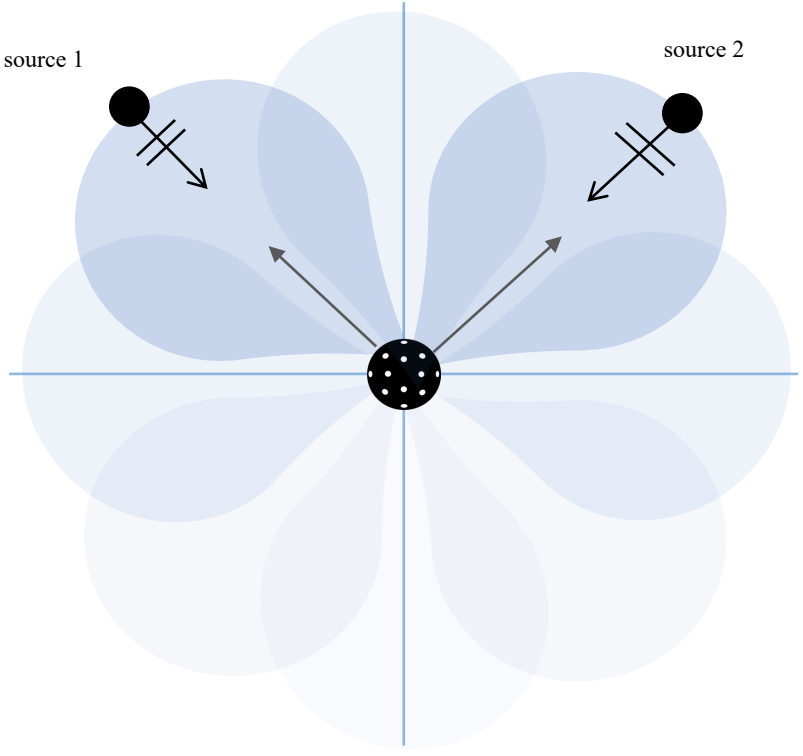
Spatial Parameterization Models

Method	Input	Model	
DirAC (Pulkki, 2006)	FOA (4ch)		1 source component + 1 iso. diffuse component
HARPEX (Berge, 2010)	FOA (4ch)		2 source components
HO-DirAC (Politis et.al., 2015)	HOA (9+ch)		~M sector source + ~M sector diffuse components
Sparse Recovery (Wabnitz, Jin, 2012)	FOA/HOA (4+ch)		≤M/2 source components
COMPASS (Politis et.al., 2018)	FOA/HOA (4+ch)		≤M/2 source components + spatial ambient component

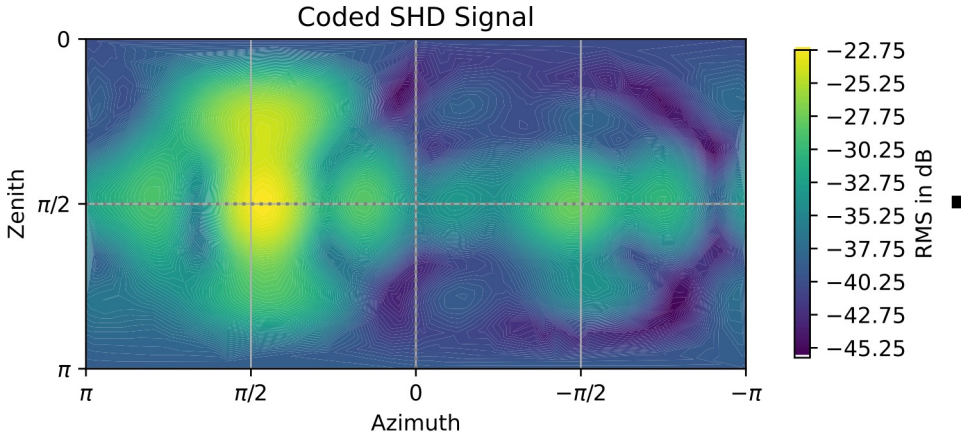
→ Typically used for enhanced HOA rendering (to loudspeakers)

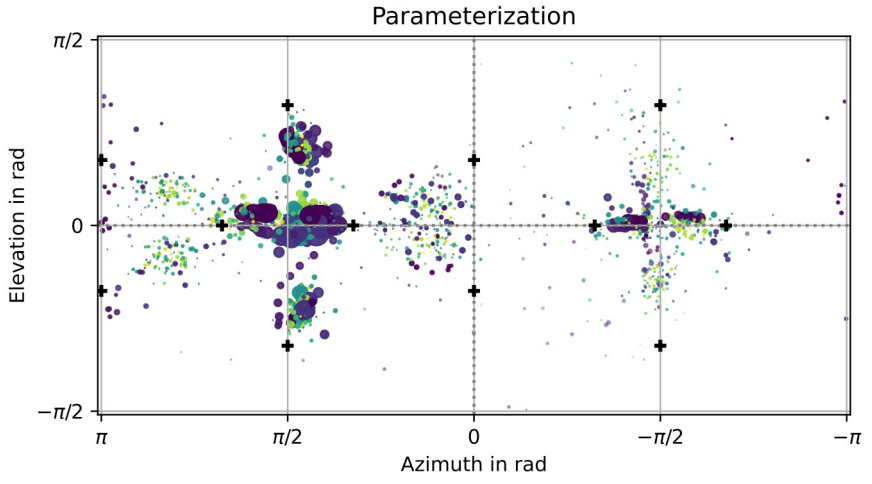
Higher-Order Directional Audio Coding (HODirAC)

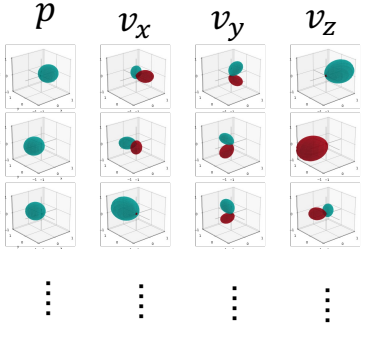
Parametric Spatial Audio - HODirAC

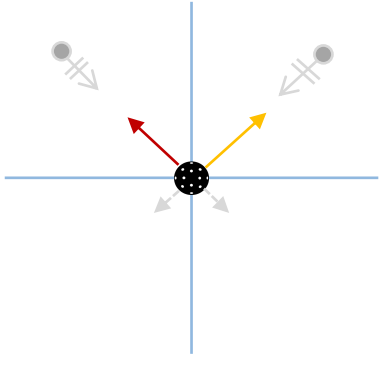


Parameter Estimation







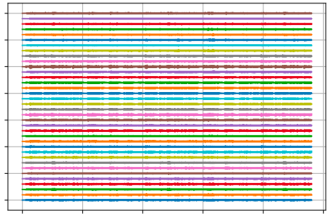


$$\boldsymbol{i}_\xi \propto \Re\{p_\xi^{\mathsf{H}} \boldsymbol{v}_\xi\}$$

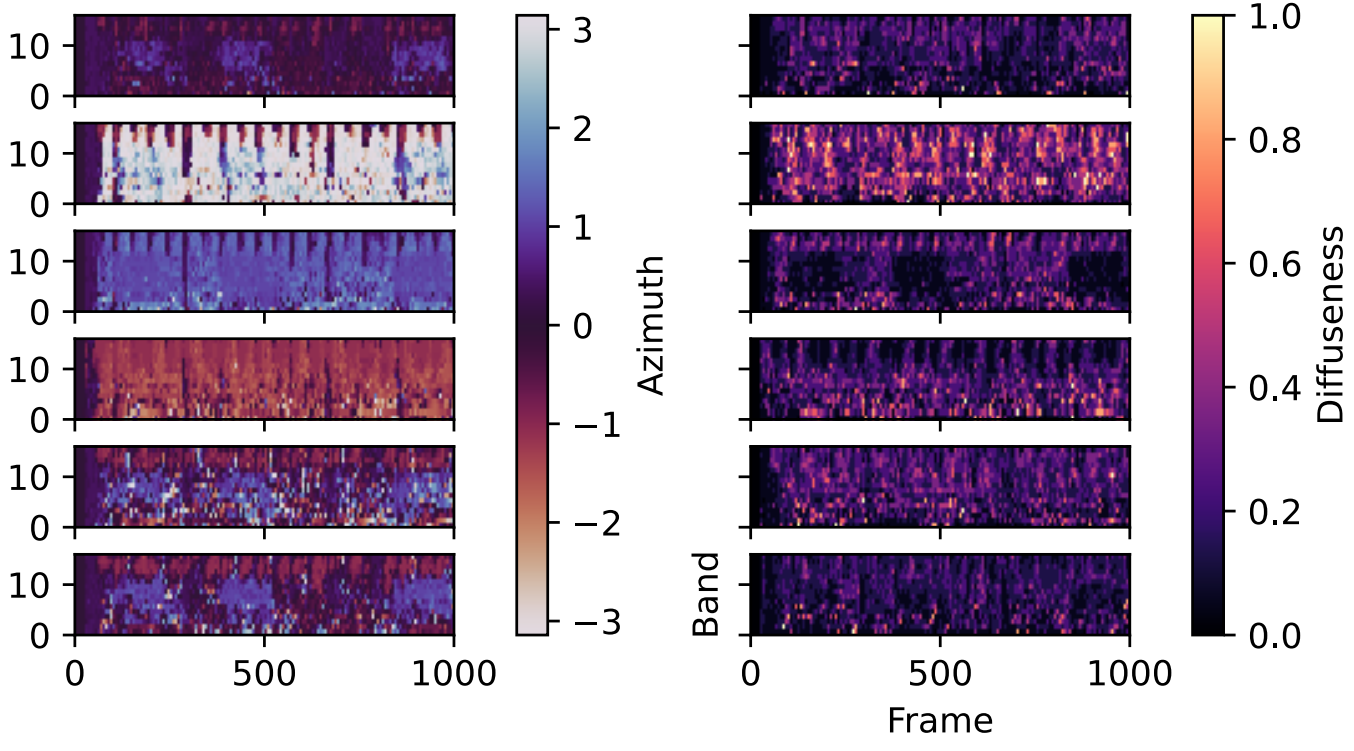
$$\Omega_\xi^{\mathrm{DoA}} = \angle \boldsymbol{i}_\xi$$

$$\psi_\xi = 1 - \frac{||\boldsymbol{i}_\xi||}{E_\xi} = 1 - \frac{2||\boldsymbol{i}_\xi||}{|p_\xi|^2 + \boldsymbol{v}_\xi{}^{\mathsf{H}} \boldsymbol{v}_\xi}$$

Parameter Estimation - Stream



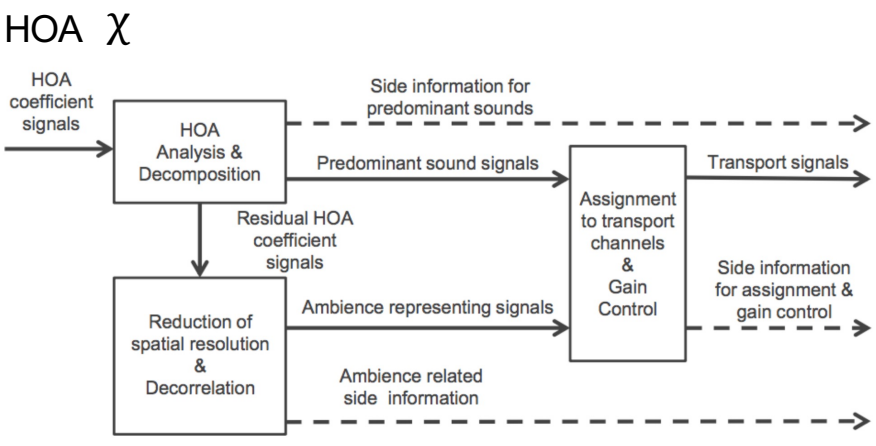
HOA χ



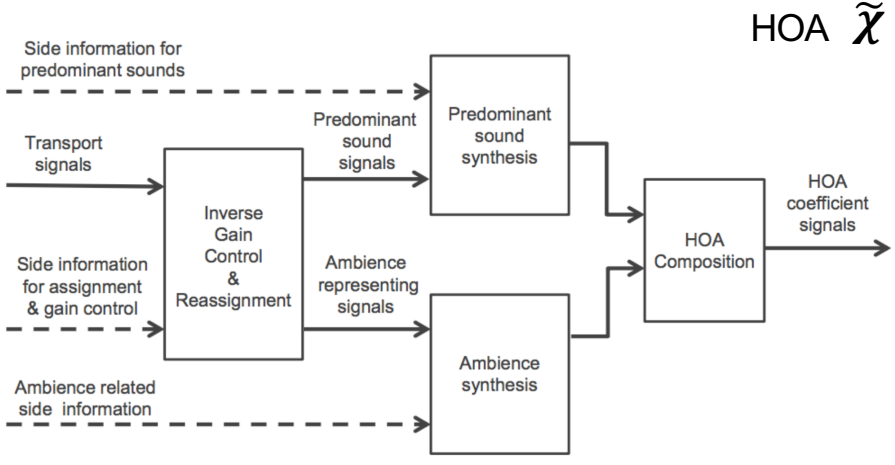
Making a codec

State of the Art: MPEG-H 3D Audio

- designed to provide universal coding of channel-based, object-based and Higher Order Ambisonics input



HOA - Encoder



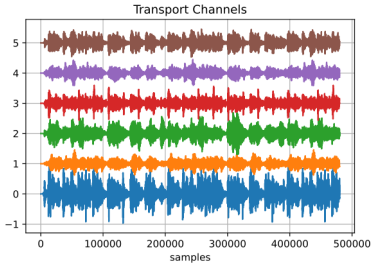
Decoder

Sen, D., Peters, N., Kim, M. Y., & Morrell, M. (2016). Efficient compression and transportation of scene based audio for television broadcast. *Proceedings of the AES International Conference*, *2016-July*.

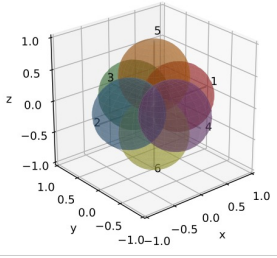
Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015). MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio. *IEEE Journal of Selected Topics in Signal Processing.*

Encoder

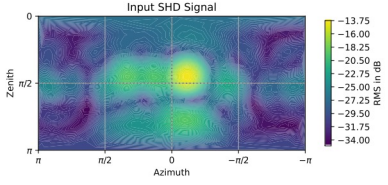
Encoder



Audio Signals



HOA χ

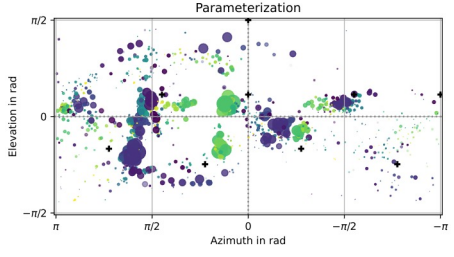


Sector Beamformers

Perceptual Coder

Statistical Coder

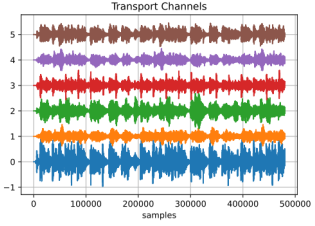
Parameters



*.hoac

Decoder

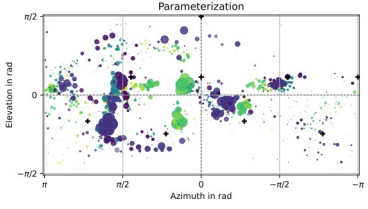
Decoder



*.hoac

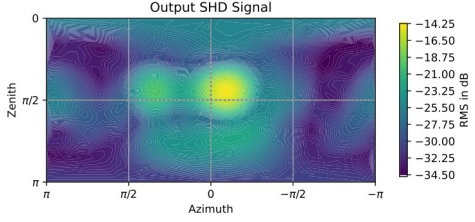
Audio $\mathbf{X}$

Parameters



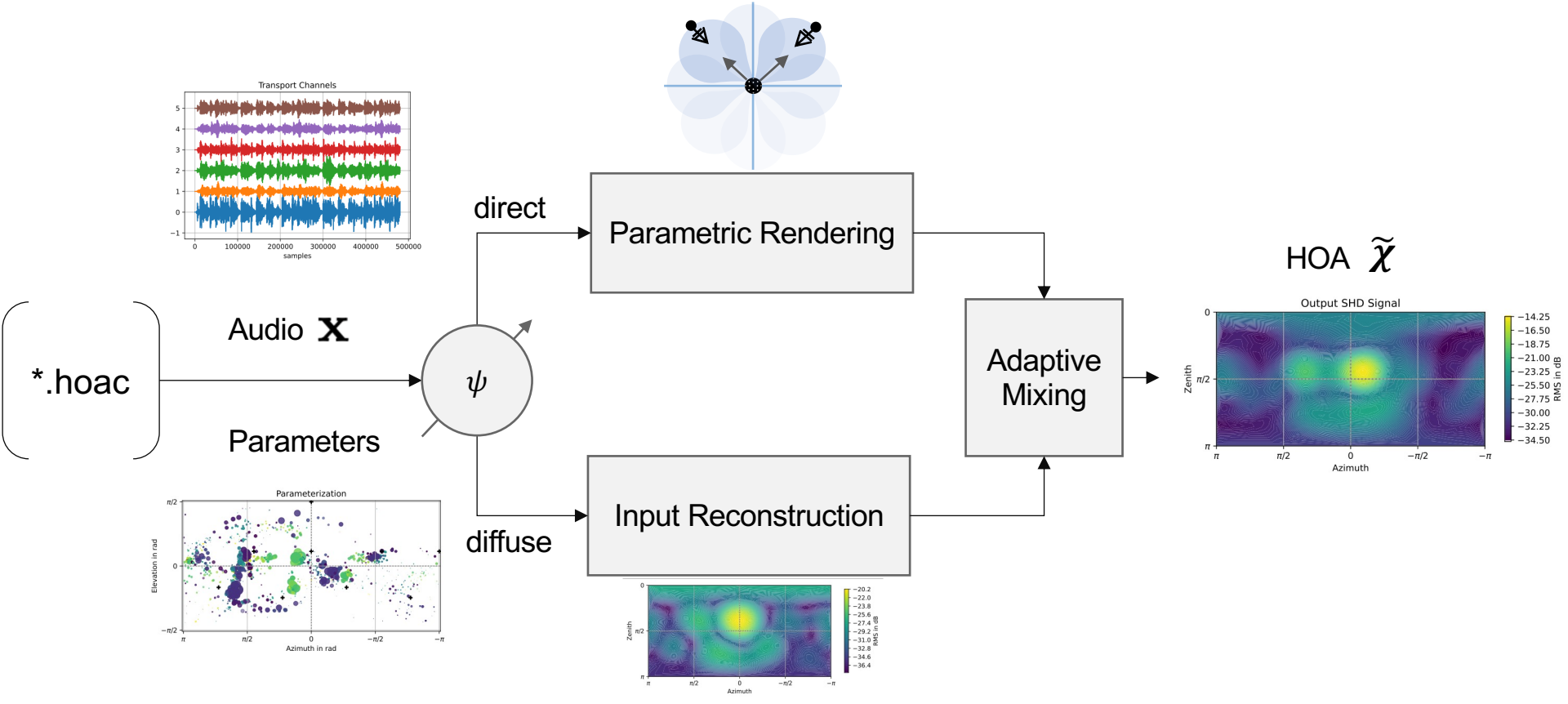
Spatial Reconstruction

HOA $\tilde{\chi}$

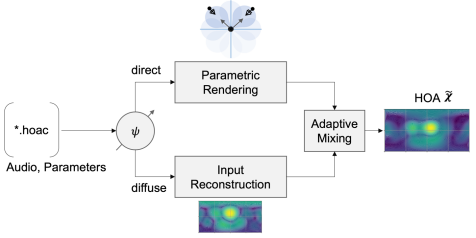


Aalto-yliopisto
Aalto-universitetet
Aalto University

Decoder - Spatial Reconstruction

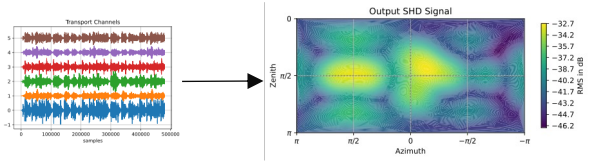


Recover and Resynthesize

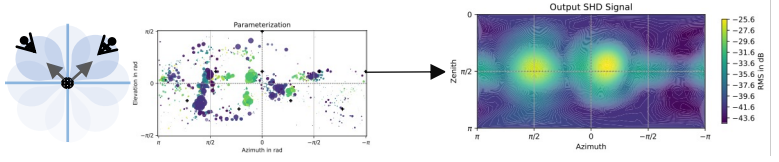


- Reconstruction (low orders) :

$$\text{SFB } \mathbf{x} = \mathbf{A}\boldsymbol{\chi} \text{ , and } \tilde{\boldsymbol{\chi}} = \mathbf{B}\mathbf{x} \rightarrow$$
$$\mathbf{B} = \lfloor \mathbf{A} \rfloor^{\dagger}$$



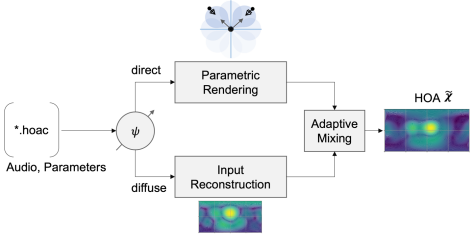
- Resynthesis (high orders) :



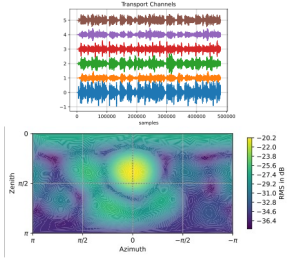
$$\hat{\boldsymbol{\chi}} = \underbrace{\mathbf{B}\,\mathrm{diag}(\boldsymbol{\psi}_s)\mathbf{x}}_{\text{diffuse}} + \underbrace{\beta_A \mathbf{Y}(\boldsymbol{\Omega}_s)\,\mathrm{diag}(1-\boldsymbol{\psi}_s)\mathbf{x}}_{\text{directional}}$$

$$\hat{\boldsymbol{\chi}} = \mathbf{Q}\mathbf{x}$$

Mix and Match



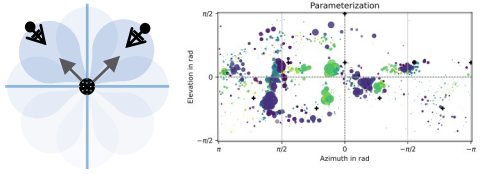
Measurements



$$\mathbf{C}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^{\mathsf{H}}\}$$

$$\tilde{\mathbf{C}}_\chi = \mathcal{E}\{\tilde{\boldsymbol{\chi}}\tilde{\boldsymbol{\chi}}^{\mathsf{H}}\} = \mathbf{B}\mathbf{C}_x\mathbf{B}^{\mathsf{H}}$$

Target + Model



$$\mathbf{C}_\chi = \mathbf{C}_{\mathrm{dir}} + \mathbf{C}_{\mathrm{dif}}$$

$$\mathcal{E}\{\mathbf{Q}\mathbf{x}(\mathbf{Q}\mathbf{x})^{\mathsf{H}}\} = \mathbf{Q}\mathbf{C}_x\mathbf{Q}^{\mathsf{H}}$$

Solution

$$\tilde{\boldsymbol{\chi}} = \mathbf{M}\mathbf{x} + \mathbf{r}$$

Aalto-yliopisto
Aalto-universitetet
Aalto University

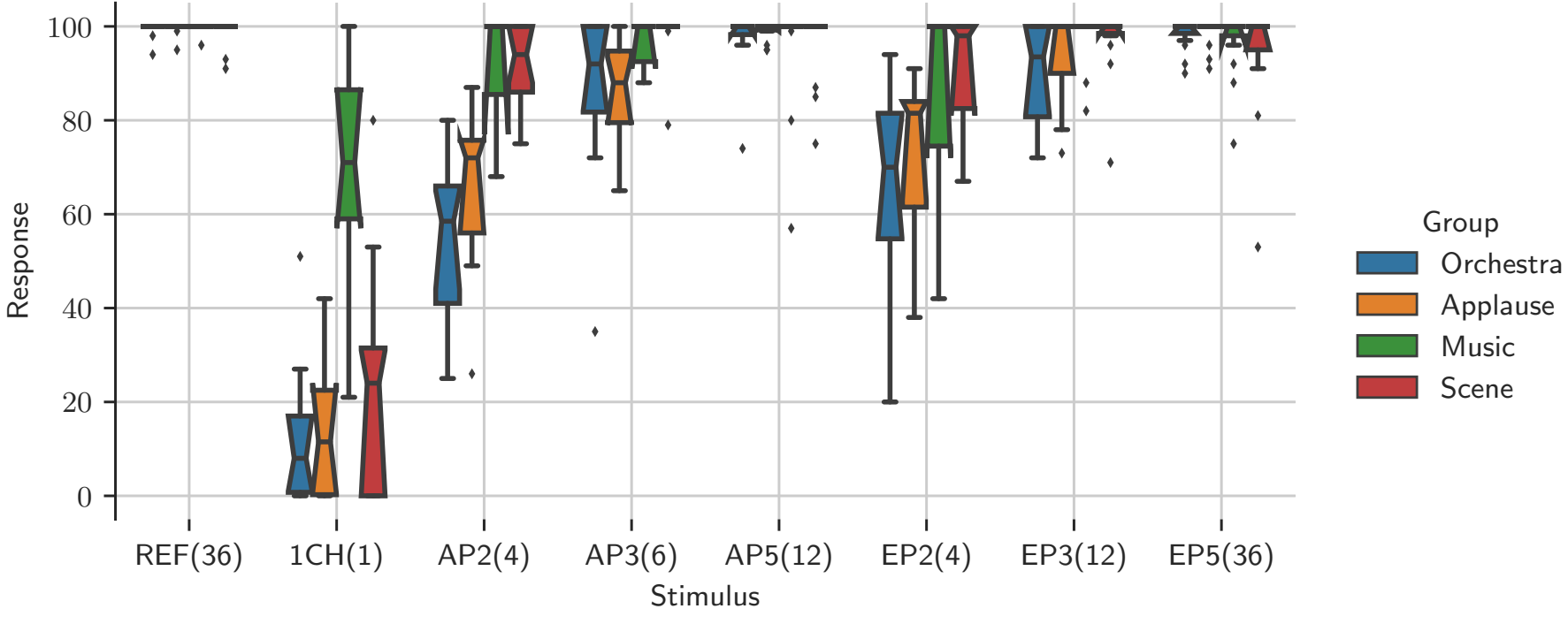
Evaluation

Input (MagLS5)

Opus @ 768kbit/s (MagLS5)

Hoac @ “Opusbitrate” (MagLS5)

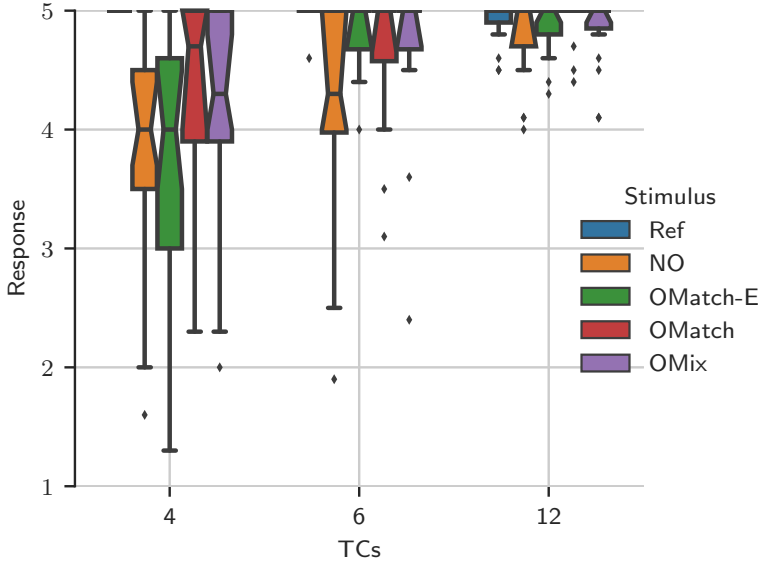
Influence of Transport Channels

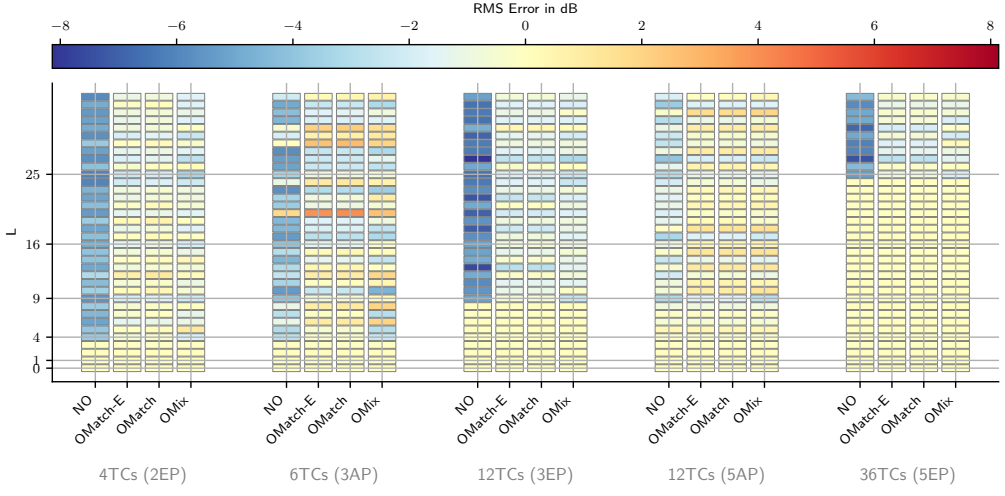


Aalto-yliopisto
Aalto-universitetet
Aalto University

Compression of Higher-Order Ambisonic Signals using Directional Audio Coding,
Christoph Hold, Ville Pulkki, Archontis Politis, Leo McCormack, IEEE TASLP, 2023

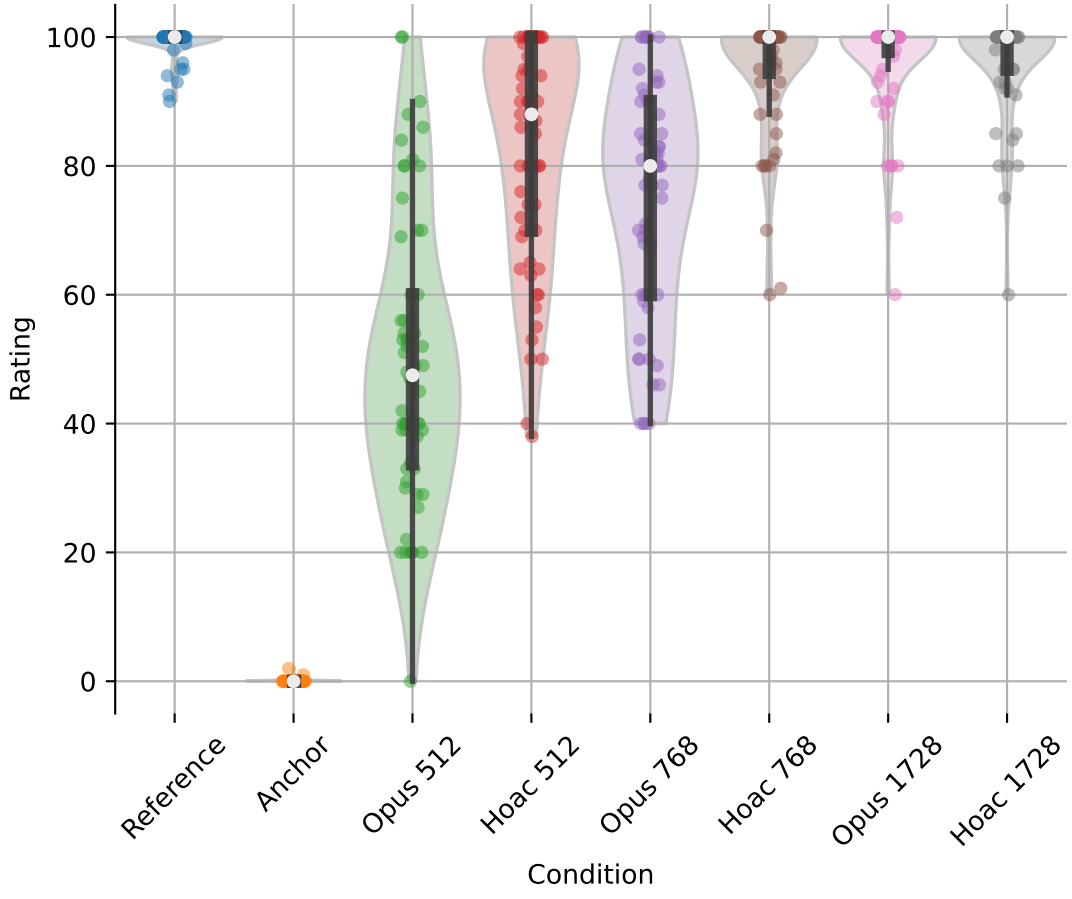
Influence of Adaptive Mixing





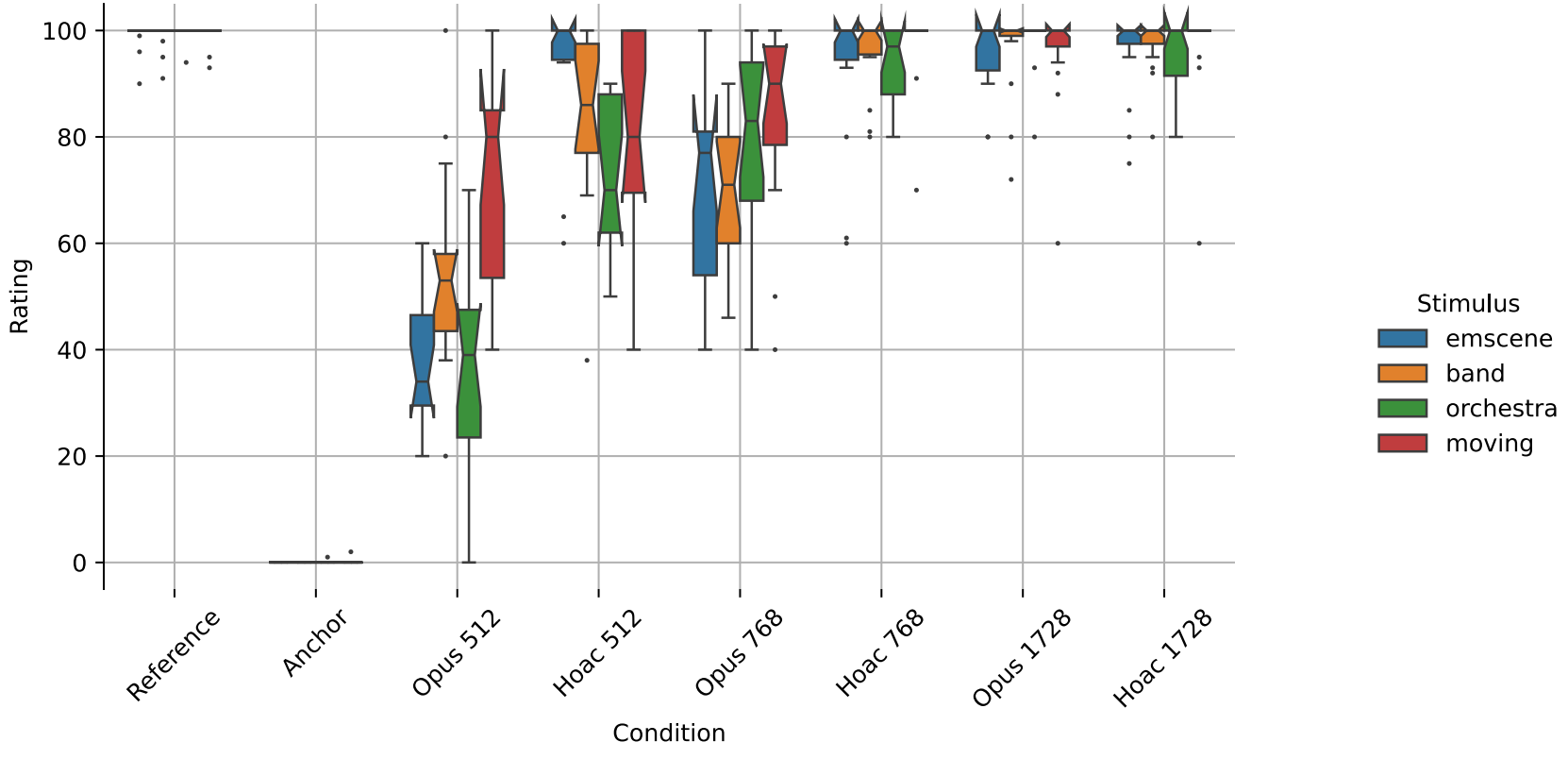
Optimizing Higher-Order Directional Audio Coding with Adaptive Mixing and Energy Matching for Ambisonic Compression and Upmixing,
Christoph Hold, Leo McCormack, Archontis Politis, Ville Pulkki, IEEE WASPAA, 2023

Influence of Coders



Perceptually-Motivated Spatial Audio Codec for Higher-Order Ambisonics Compression,
Christoph Hold, Leo McCormack, Archontis Politis, Ville Pulkki, IEEE ICASSP, 2024

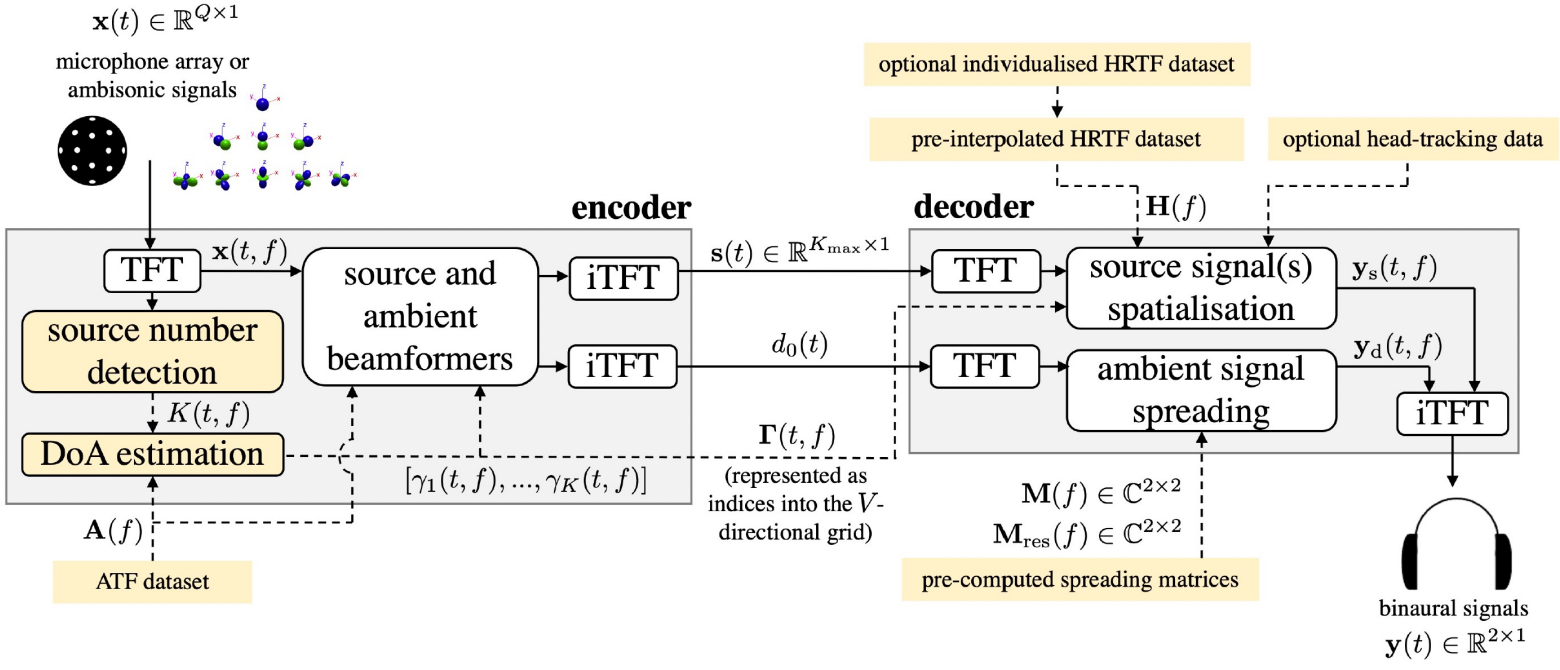
Influence of Coders



Perceptually-Motivated Spatial Audio Codec for Higher-Order Ambisonics Compression,
Christoph Hold, Leo McCormack, Archontis Politis, Ville Pulkki, IEEE ICASSP, 2024

Which model is the best?

It depends!



Parametric architecture for the transmission and binaural reproduction of microphone array recordings, *Leo McCormack, Christoph Hold, Archontis Politis, AES 2023 International Conference on Spatial and Immersive Audio, 2023*

Implications and Outlook

- Parametric spatial audio can benefit compression
- Allows balancing bitrate between audio core-coders and metadata
  - Fewer audio TCs at higher quality, and quantized metadata
- Input parameterization can improve decoding

- Improvements suggested for Opus (Ambi Map 3)
- Implemented low resource Ambisonics layer in 3GPP SA4 IVAS

More:

HOAC - Demo

http://research.spa.aalto.fi/publications/papers/hoac/

aalto.fi

Aalto-yliopisto
Aalto-universitetet
Aalto University