Ambisonic Coding with Spatial Image Correction

January 2021

Pierre MAHE - Orange Labs and L3i, University of La Rochelle, France
pierre.mahe@orange.com

Stéphane RAGOT - Orange Labs, Lannion, France
stephane.ragot@orange.com

Sylvain MARCHAND - L3i, University of La Rochelle, France
sylvain.marchand@univ-lr.fr

Jérôme DANIEL - Orange Labs, Lannion, France
jerome.daniel@orange.com

orange™

EUSIPCO2020
28th European Signal Processing Conference,
August 24 - 28, 2020
Amsterdam

La Rochelle Université

Context and Motivations

- Telephony codecs are mostly limited to mono
- Emergence of devices supporting spatial audio
- New needs for immersive audio compression
- Extend existing codecs

Context and Motivations

- Telephony codecs are mostly limited to mono
- Emergence of devices supporting spatial audio
- New needs for immersive audio compression
- Extend existing codecs

Immersive communication, what for ?

Context and Motivations

- Telephony codecs are mostly limited to mono
- Emergence of devices supporting spatial audio
- New needs for immersive audio compression
- Extend existing codecs

Immersive communication, what for ?

- Call with 3D ambiance sharing

Context and Motivations

- Telephony codecs are mostly limited to mono
- Emergence of devices supporting spatial audio
- New needs for immersive audio compression
- Extend existing codecs

Immersive communication, what for ?

- Call with 3D ambiance sharing
- Immersive content broadcasting (360 Video, VR…)

Context and Motivations

- Telephony codecs are mostly limited to mono
- Emergence of devices supporting spatial audio
- New needs for immersive audio compression
- Extend existing codecs

Immersive communication, what for ?

- Call with 3D ambiance sharing
- Immersive content broadcasting (360 Video, VR…)
- Spatial audio conferencing

3D Audio and Ambisonics

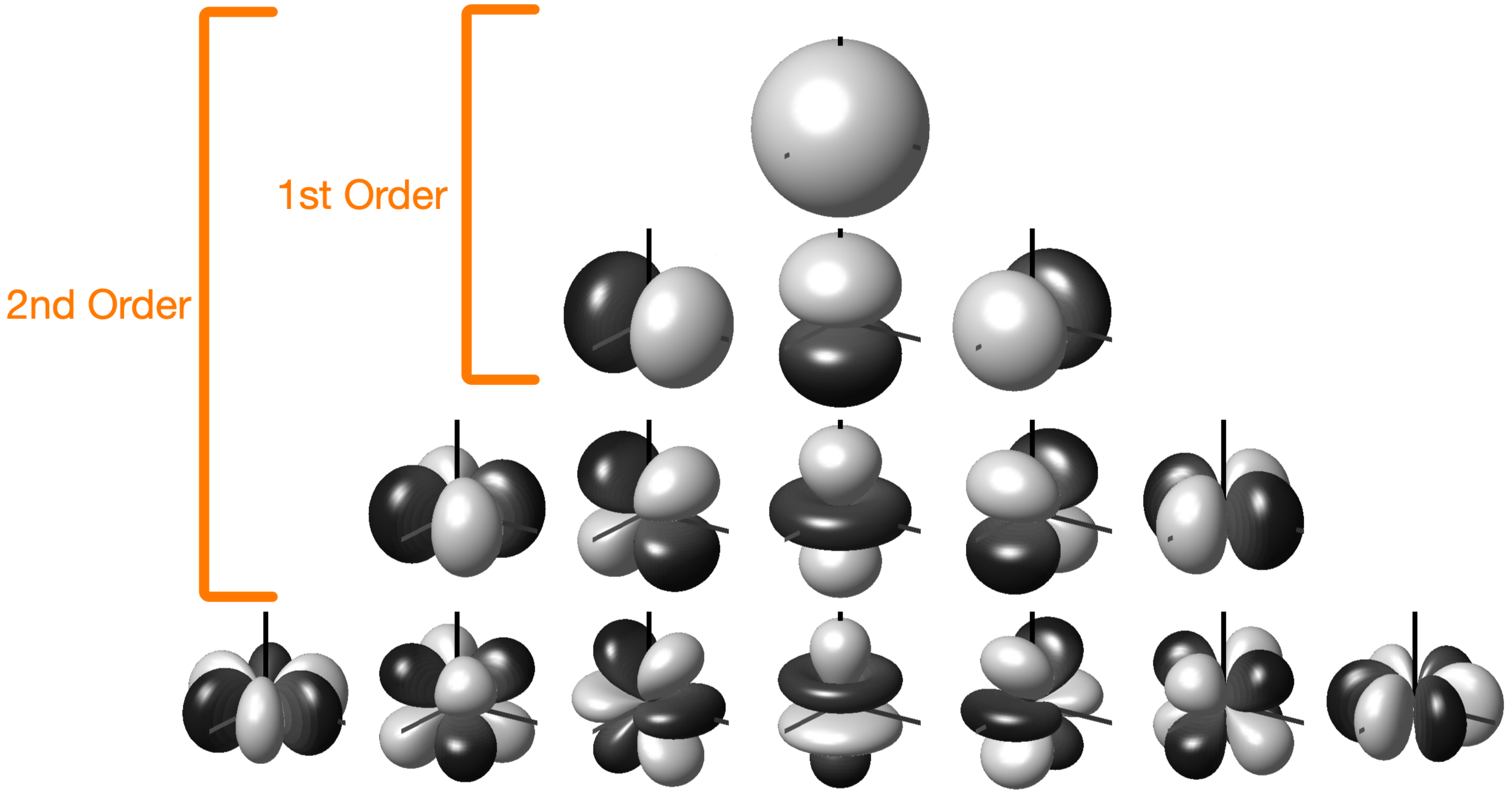
Ambisonics is a sound field decomposition into a spherical harmonics basis.

[1] M. A. Gerzon, "Periphony: With-height sound reproduction," AES Journal, Volume 21 Issue 1 pp.2–10, 1973.
[2] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D., Université Paris 6, 2000.

3D Audio and Ambisonics

Ambisonics is a sound field decomposition into a spherical harmonics basis.

- First-order ambisonics (FOA) was invented in 1970s by Gerzon et al. [1]
- Later extended to higher orders by Daniel et al. [2]
- For an order N, the number of components is $n = (N+1)^2$



[1] M. A. Gerzon, "Periphony: With-height sound reproduction," AES Journal, Volume 21 Issue 1 pp.2–10, 1973.
[2] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D., Université Paris 6, 2000.

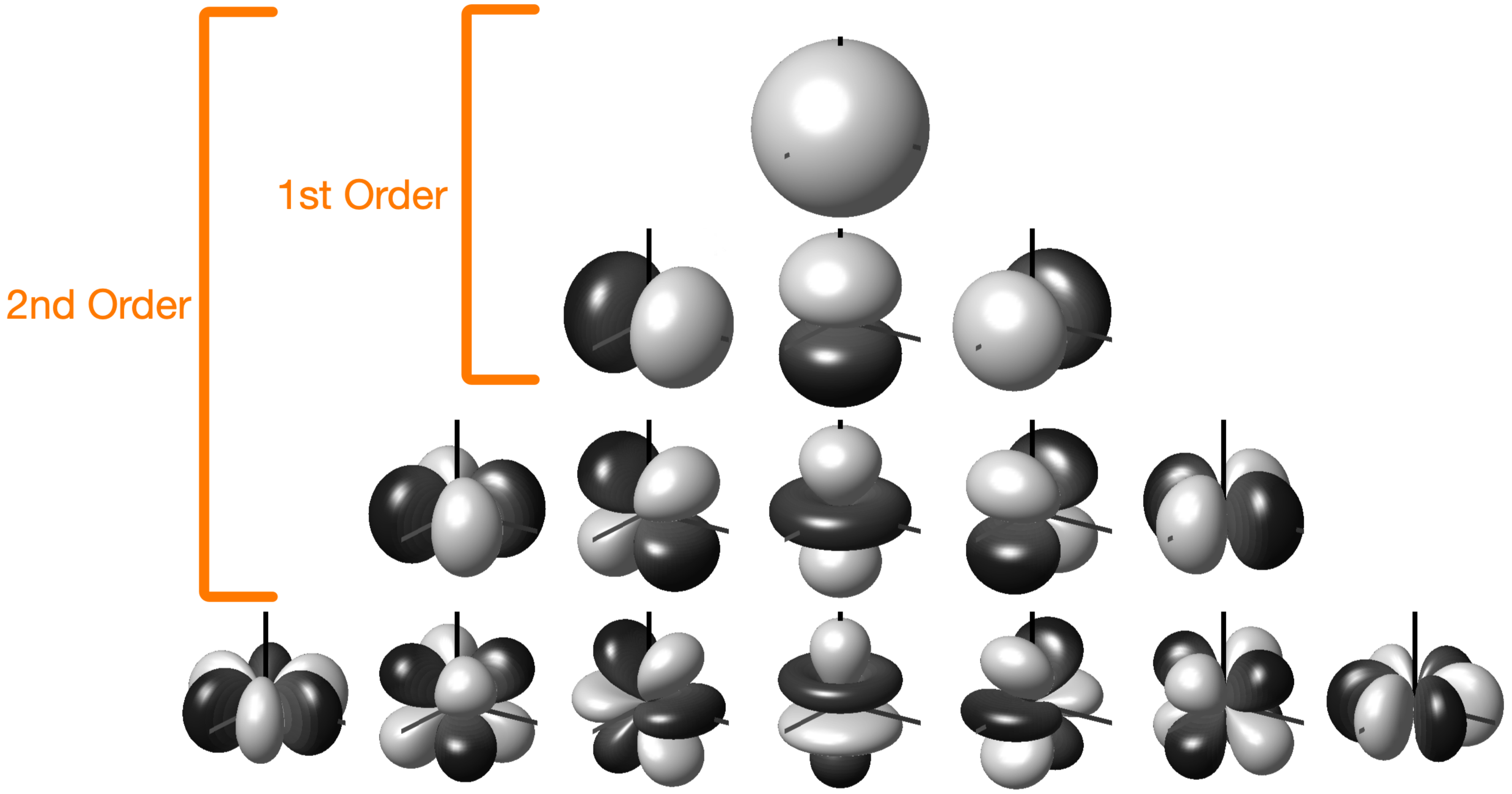
3D Audio and Ambisonics

Ambisonics is a sound field decomposition into a spherical harmonics basis.

- First-order ambisonics (FOA) was invented in 1970s by Gerzon et al. [1]
- Later extended to higher orders by Daniel et al. [2]
- For an order N, the number of components is $n = (N+1)^2$

Source encoding for First-Order Ambisonic (FOA)

$$\mathbf{b}(t) = \begin{bmatrix} w(t) \\ x(t) \\ y(t) \\ z(t) \end{bmatrix}^T = \begin{bmatrix} 1 \\ \cos\,\theta\ \ \cos\,\phi \\ \sin\,\theta\ \ \cos\,\phi \\ \sin\,\phi \end{bmatrix}^T s(t)$$



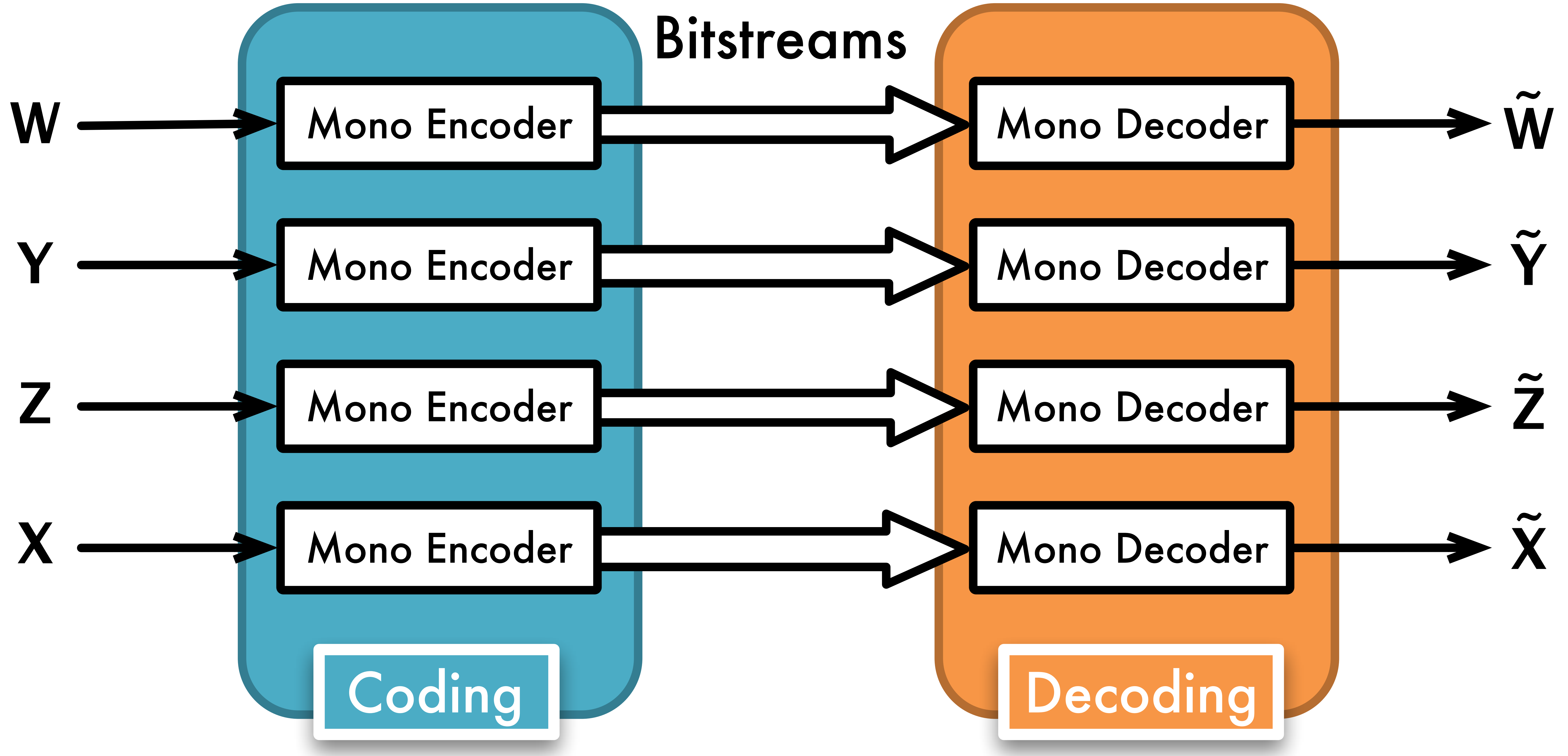
[1] M. A. Gerzon, "Periphony: With-height sound reproduction," AES Journal, Volume 21 Issue 1 pp.2–10, 1973.

[2] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D., Université Paris 6, 2000.

Baseline

Multi-mono approach

- Each ambisonic components is coded independently by a mono codec
- Bitrate is uniformly distributed between components



Baseline

Issues

Baseline

Issues

- Several artefacts and spatial distortions
  - Diffuse noise
  - Phantom sources

Baseline

Issues

- Several artefacts and spatial distortions
  - Diffuse noise
  - Phantom sources
- Highly visible on spatial power map

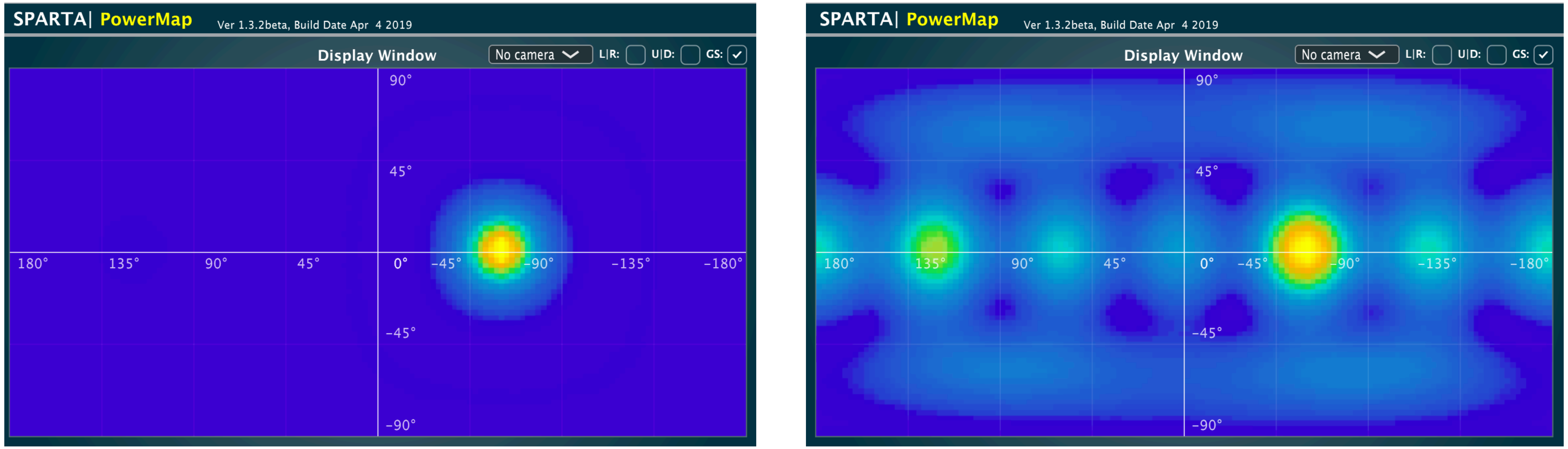


Fig. 1: Power map of original (left) and coded (right) ambisonic signals

Our approach

Observations

- Spatial artefacts are visible on power map
- If the decoder knew the original power map, it could correct the spatial distortions

Our approach

Observations

- Spatial artefacts are visible on power map
- If the decoder knew the original power map, it could correct the spatial distortions

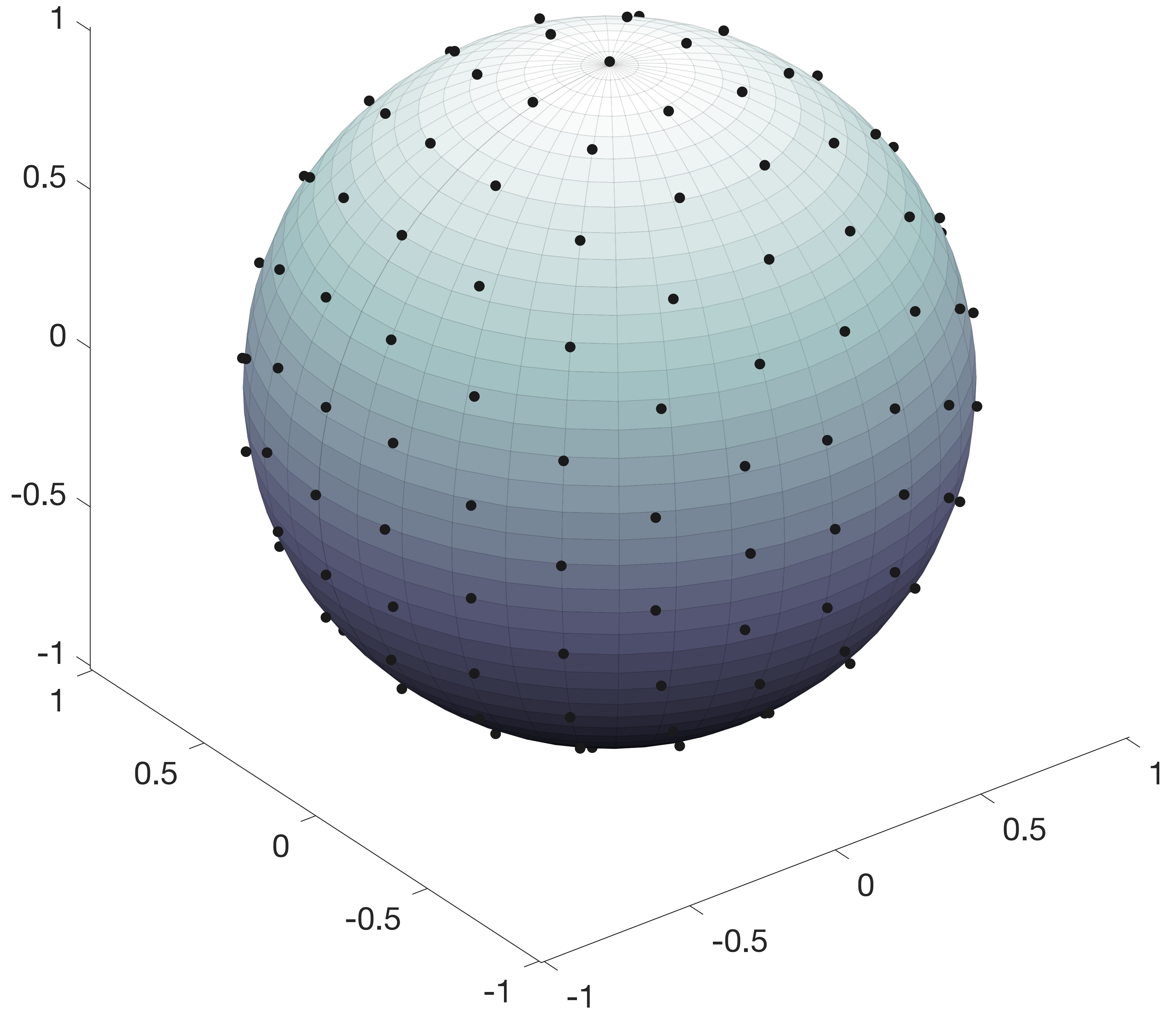
Our proposal

- Create a post-processing to correct multi-mono
- Based on the power map
- Re-using core codecs

Our approach — Power map

How does power map work ?

- A set of points discretizing the sphere is taken

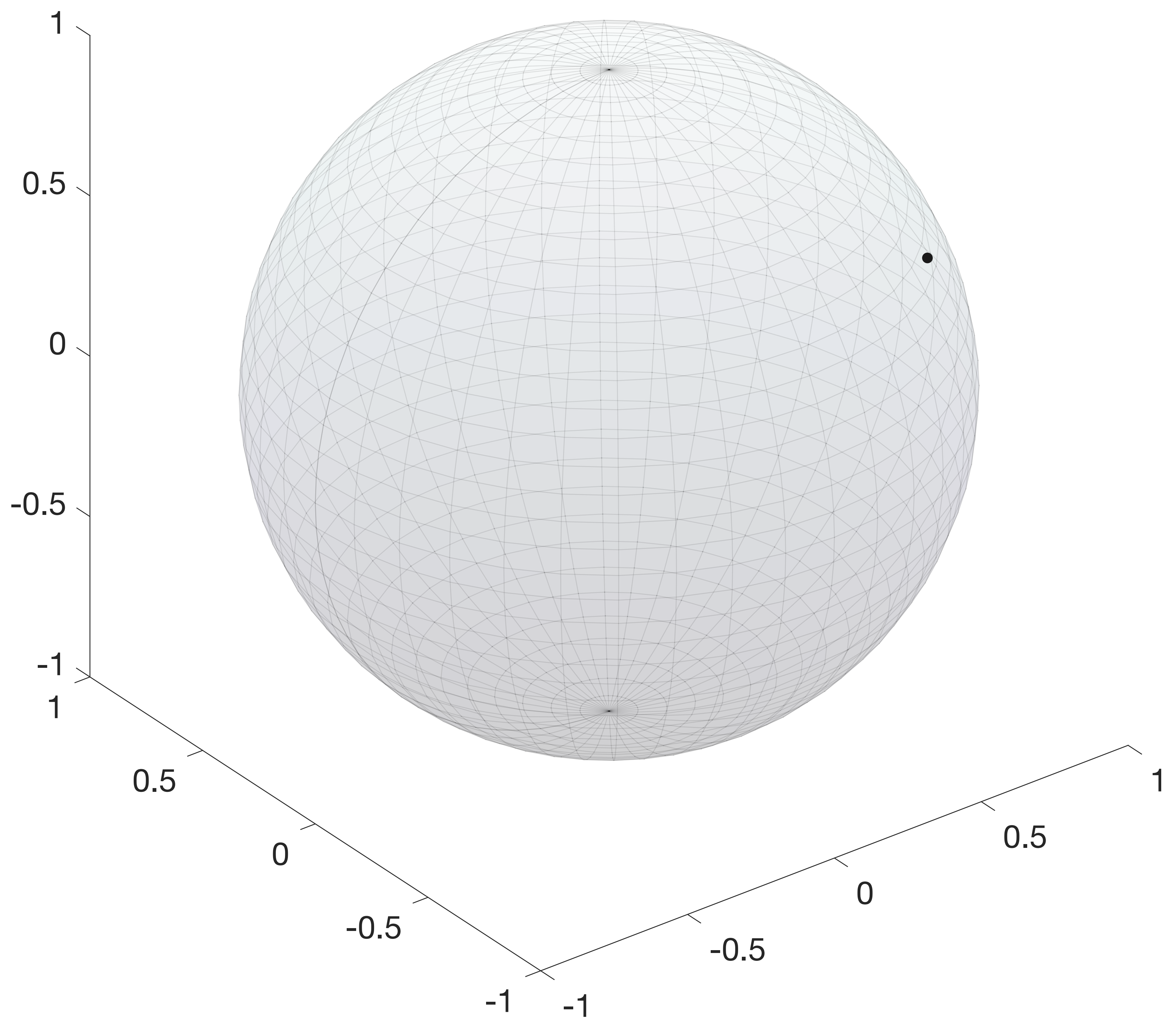


Our approach — Power map

How does power map work ?

- A set of points discretizing the sphere is taken
- For each, a weight vector $\mathbf{d}_i$ is used to create a bream forming in the direction i

$$s_i(t) = \sum_{k=1}^{n} \mathbf{d}_i(k) \cdot b_k(t)$$



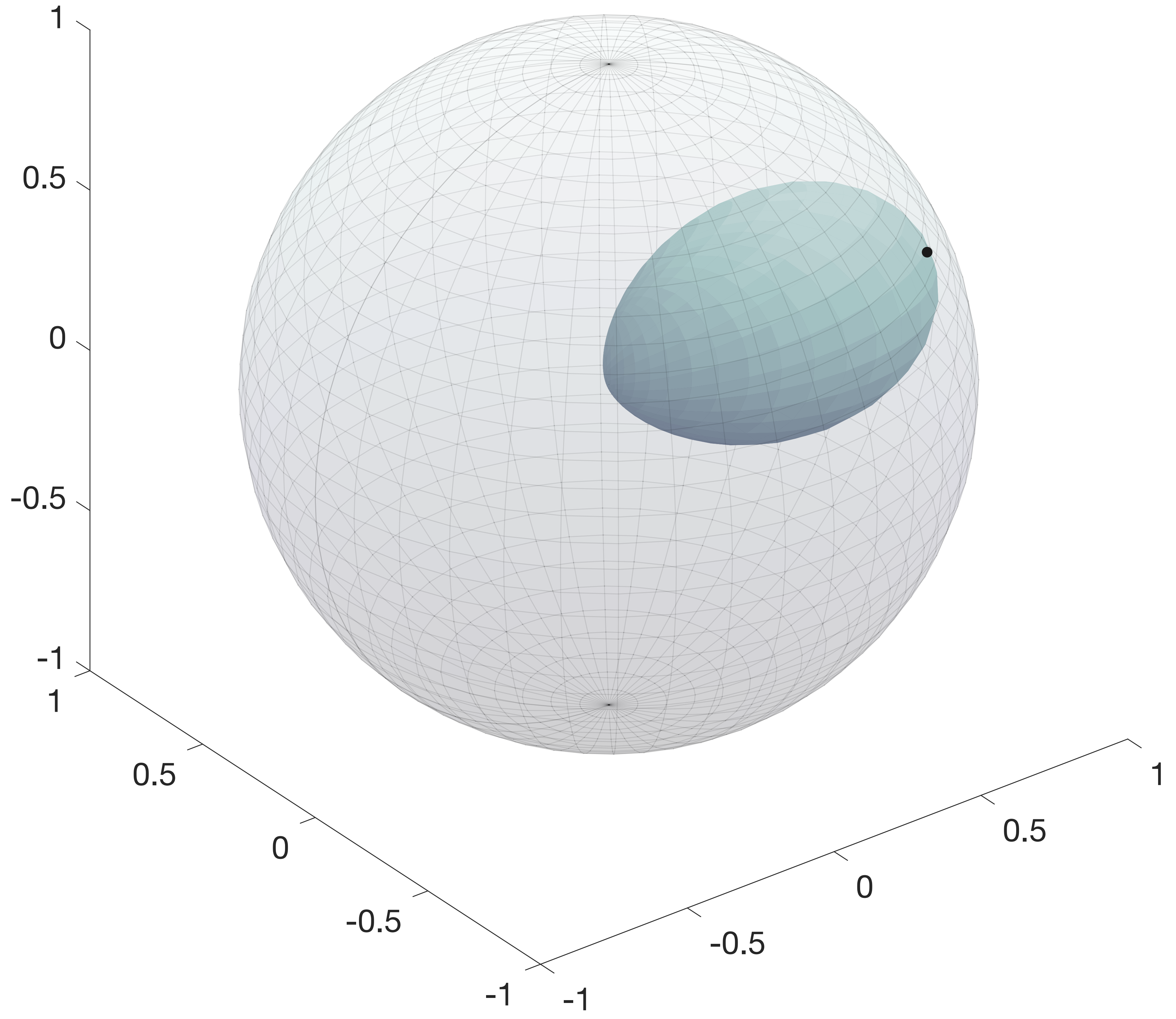
Our approach — Power map

How does power map work ?

- A set of points discretizing the sphere is taken
- For each, a weight vector $\mathbf{d}_i$ is used to create a bream forming in the direction i

$$s_i(t) = \sum_{k=1}^{n} \mathbf{d}_i(k) \cdot b_k(t)$$

$$\mathbf{P}_i = \sum_{t=1}^{L} s_i(t)^2$$



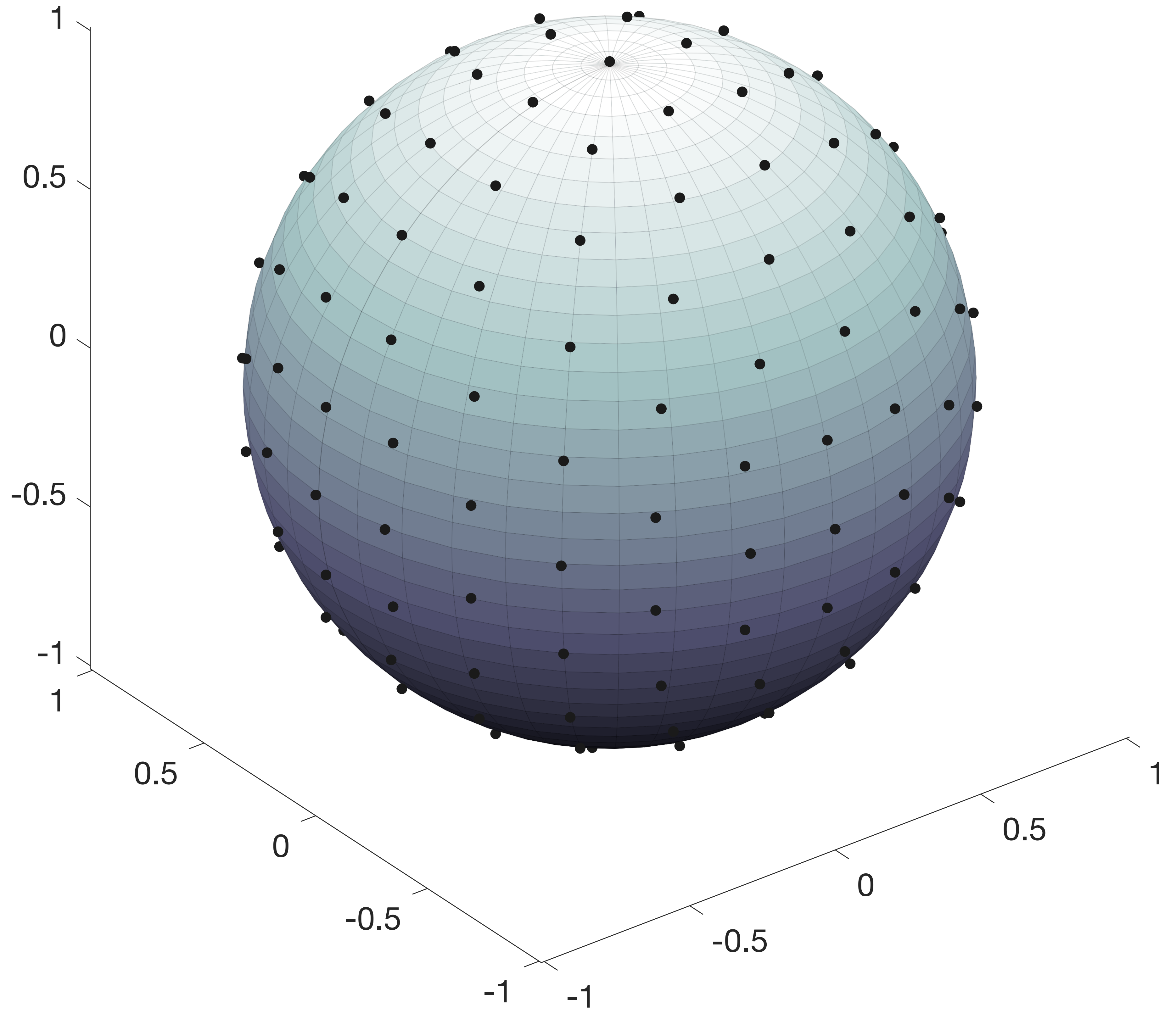
Our approach — Power map

How does power map work ?

- By re-writing equations, the power of each direction can be compute directly

$$\mathbf{P}_i = \mathbf{d}_i \mathbf{C} \mathbf{d}_i$$

Where $\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^{\mathbf{T}}$



Our approach — Spatial correction

- Problem formulation: $$\forall i,\ d_i \mathbf{C}_{cor} d_i = d_i \mathbf{C} d_i$$

Our approach — Spatial correction

- Problem formulation: $$\forall i,\ d_i \mathbf{C}_{cor} d_i = d_i \mathbf{C} d_i$$
- Let: $$\mathbf{B}_{cor} = \mathbf{T}\widetilde{\mathbf{B}}$$
- Try to find the matrix $\mathbf{T}$ such as: $$\mathbf{C}_{cor} = \mathbf{T}\widetilde{\mathbf{C}}\mathbf{T}$$

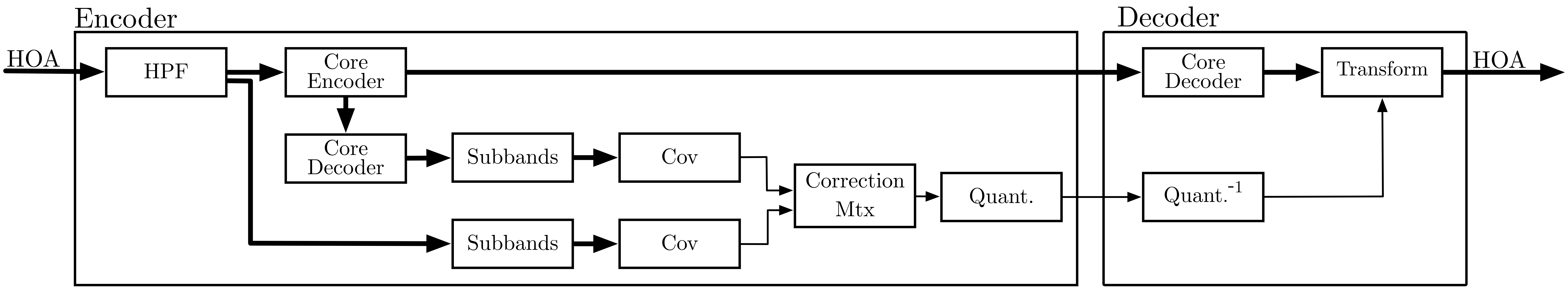
Our approach — Spatial correction

- Problem formulation: $$\forall i,\ d_i \mathbf{C}_{cor} d_i = d_i \mathbf{C} d_i$$
- Let: $$\mathbf{B}_{cor} = \mathbf{T}\widetilde{\mathbf{B}}$$
- Try to find the matrix T such as: $$\mathbf{C}_{cor} = \mathbf{T}\widetilde{\mathbf{C}}\mathbf{T}$$
- Cholesky Factorisation: $$\mathbf{C} = \mathbf{L}\mathbf{L}^t$$

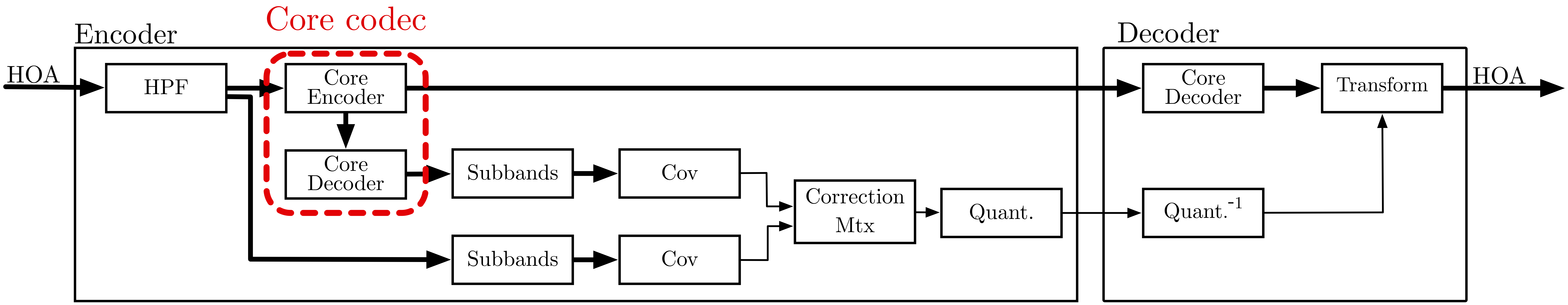
$$(\mathbf{T}\widetilde{\mathbf{L}})(\mathbf{T}\widetilde{\mathbf{L}})^t = \mathbf{L}\mathbf{L}^{-1}$$

- Then: $$\boxed{\mathbf{T} = \mathbf{L}\widetilde{\mathbf{L}}^{-1}}$$

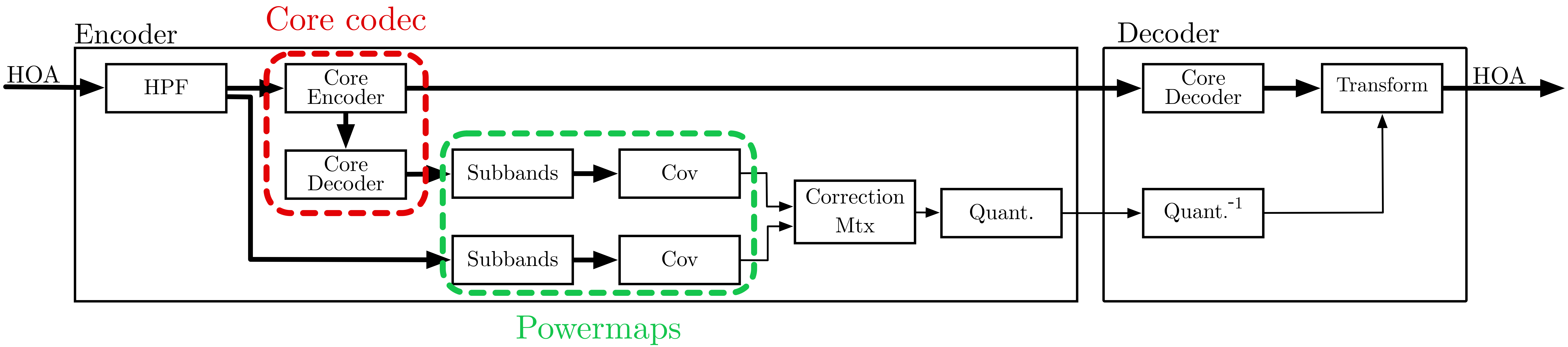
Our approach — Codec overview



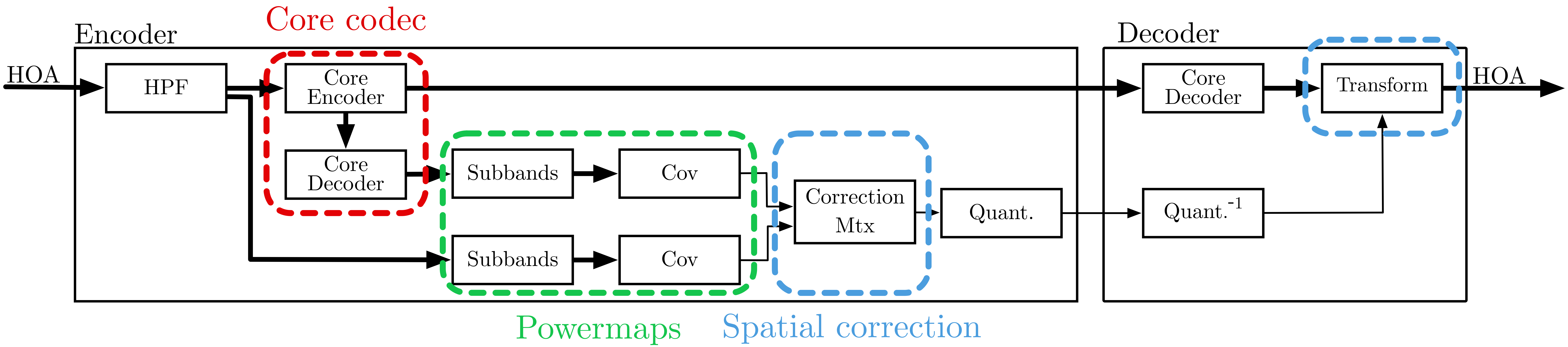
Our approach — Codec overview



Our approach — Codec overview



Our approach — Codec overview



Our approach — Power map after post-processing

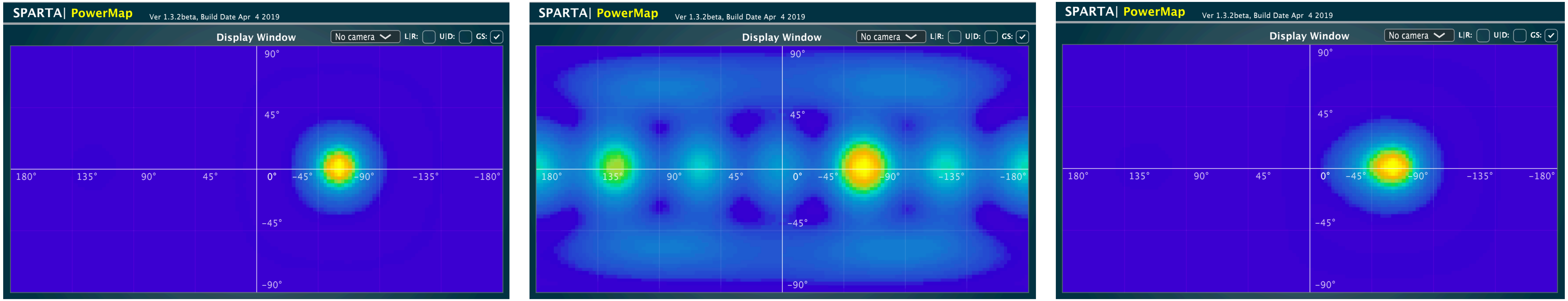
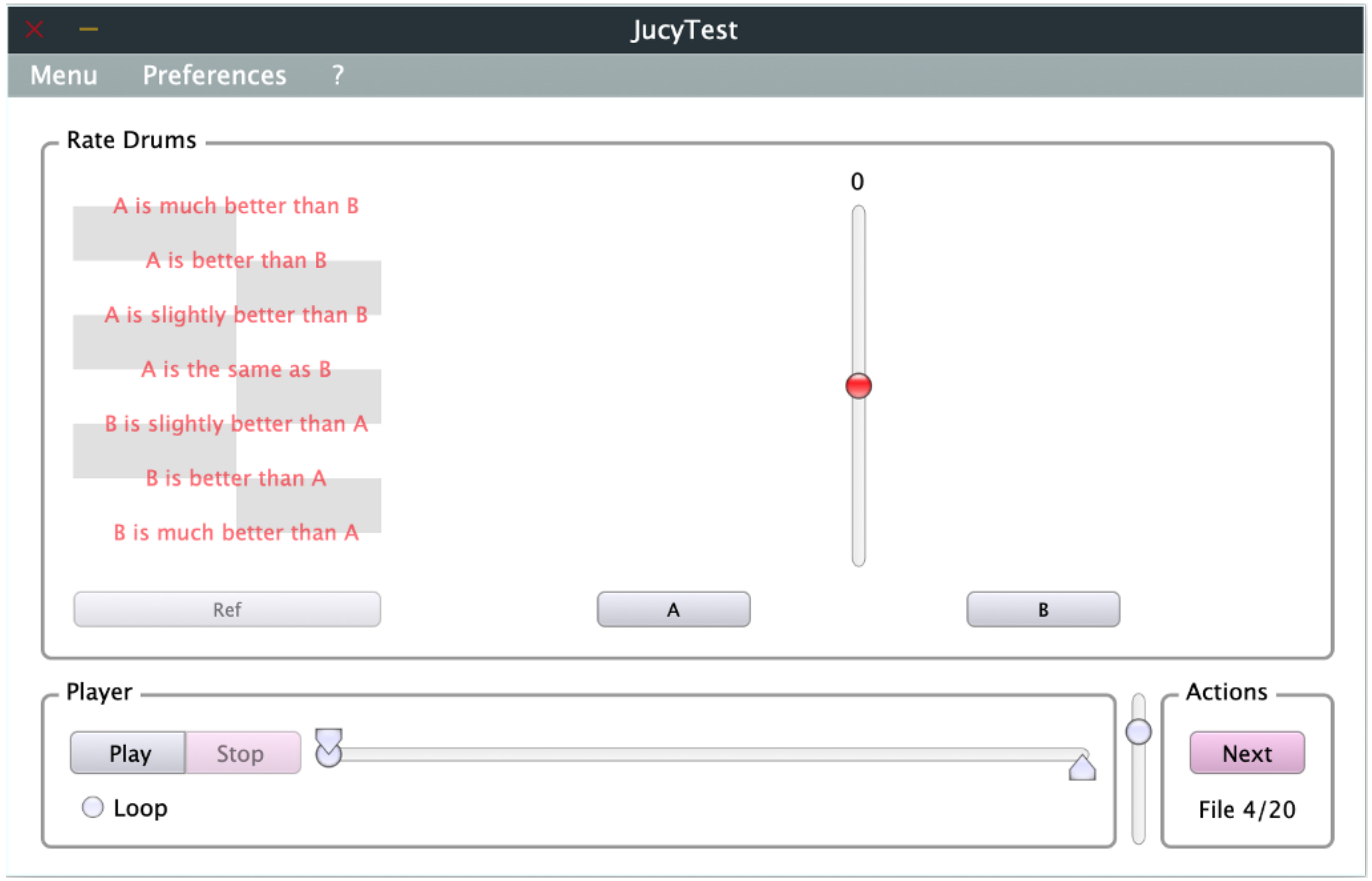


Fig. 2: Power map of original (left), coded (middle) and post-processed (right) signals

Subjective test and results

Test conditions

- Ref AB test
- Comparison with and without the post-processing
- 2 tested core codecs: EVS and OPUS
- 2 subject panels: naive and expert
- 2 tested bitrates: 97.6 kbit/s and 128.0 kbit/s

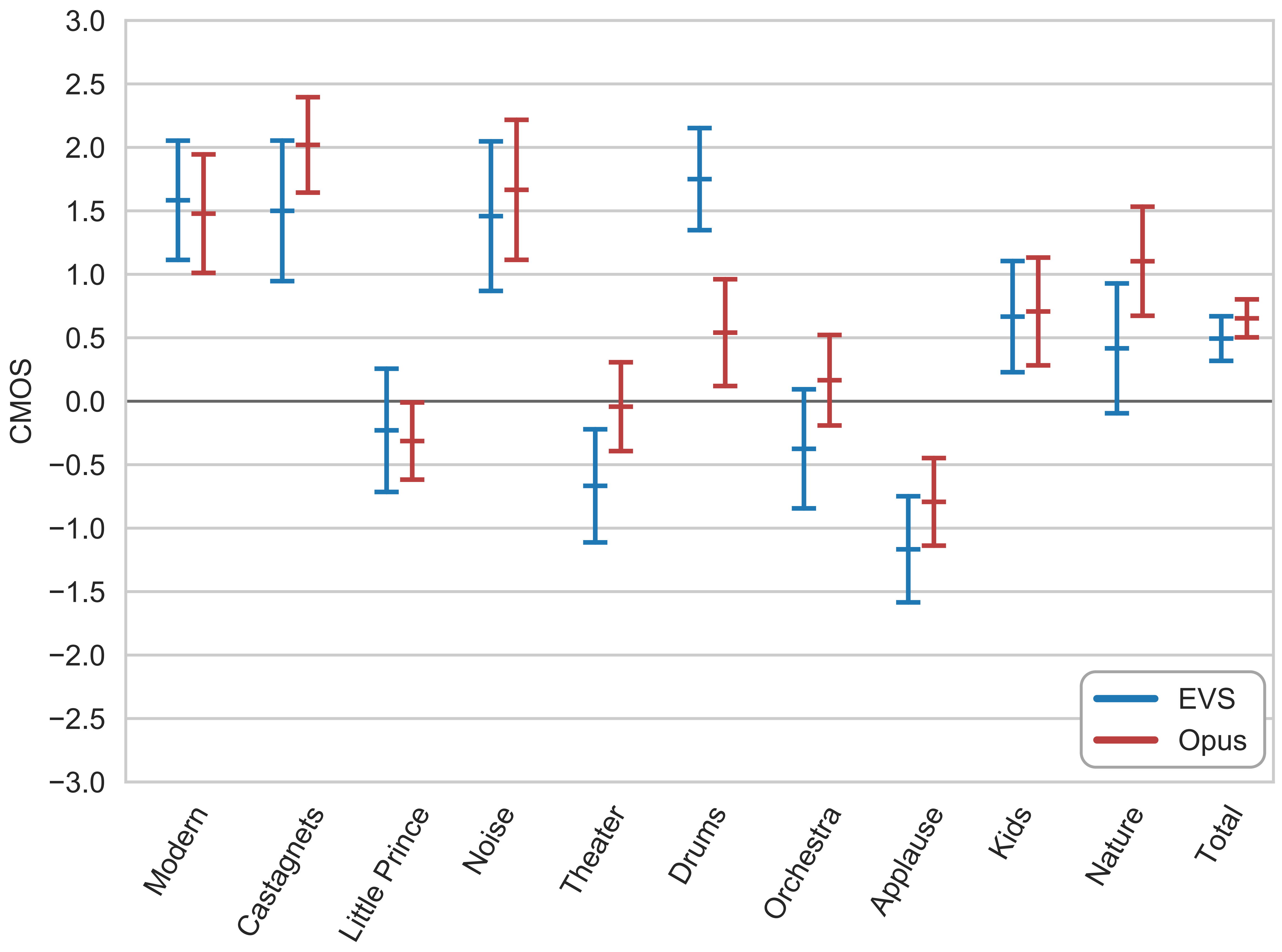


Subjective test and results

Test conditions

- Ref AB test
- Comparison with and without the post-processing
- 2 tested core codecs: EVS and OPUS
- 2 subject panels: naive and expert
- 2 tested bitrates: 97.6 kbit/s and 128.0 kbit/s

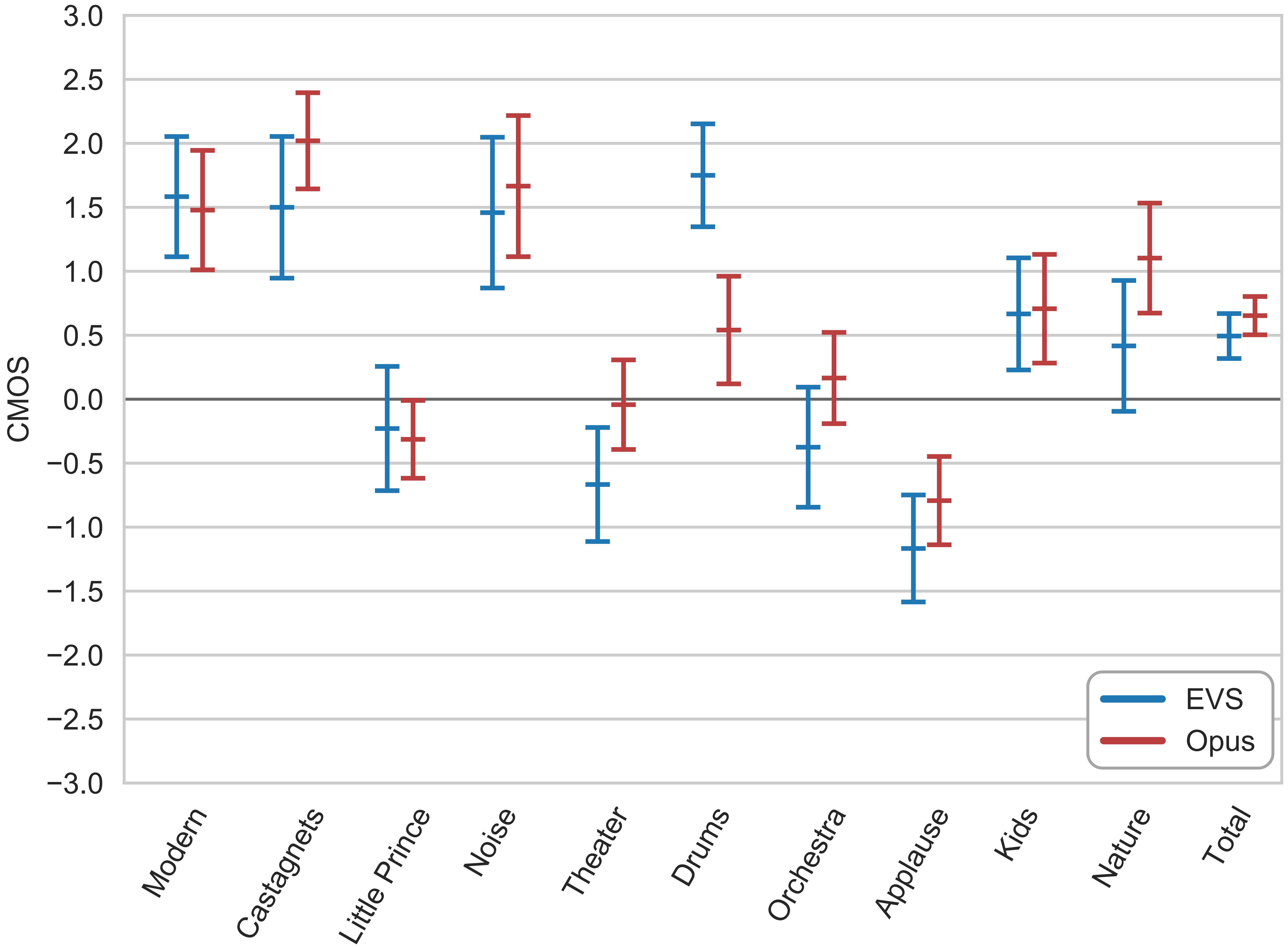
- On average, audio quality is improved



Subjective test and results

Results

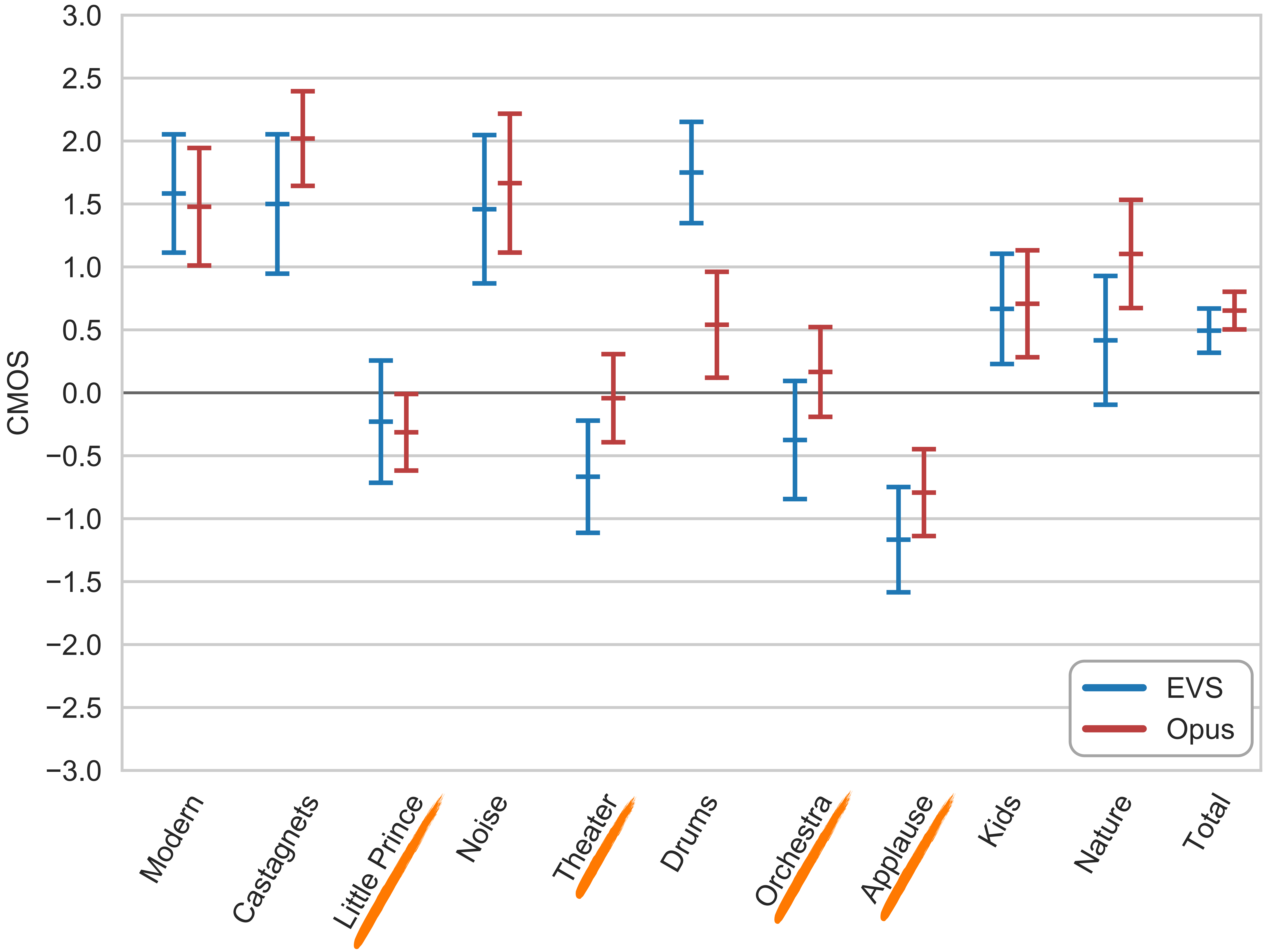
- On average, audio quality is improved
- ANOVA shows that results are similar for both codecs and both bitrates



Subjective test and results

Results

- On average, audio quality is improved
- ANOVA shows that results are similar for both codecs and both bitrates



Conclusion

- We proposed a post-processing for multi-mono coding
- The purpose is to correct spatial artifacts and distortions
- The post-processing is independent from the core codec
- Subjective test showed improvement for different core codecs and different bitrates

Thanks for your attention

Any questions ?

Pierre MAHE - Orange Labs and L3i, University of La Rochelle, France
pierre.mahe@orange.com

EUSIPCO2020
28th European Signal Processing Conference,
August 24 - 28, 2020
Amsterdam