

# CODAGE AUDIO AMBISONIQUE POUR LES COMMUNICATIONS IMMERSIVES

Pierre Mahé<sup>\*†</sup>

Stéphane Ragot\*

Sylvain Marchand<sup>†</sup>

<sup>†</sup> Université de La Rochelle, L3i, 17000 La Rochelle, France  
<sup>\*</sup> Orange Labs, 2 Avenue Pierre Marzin, 22300 Lannion, France



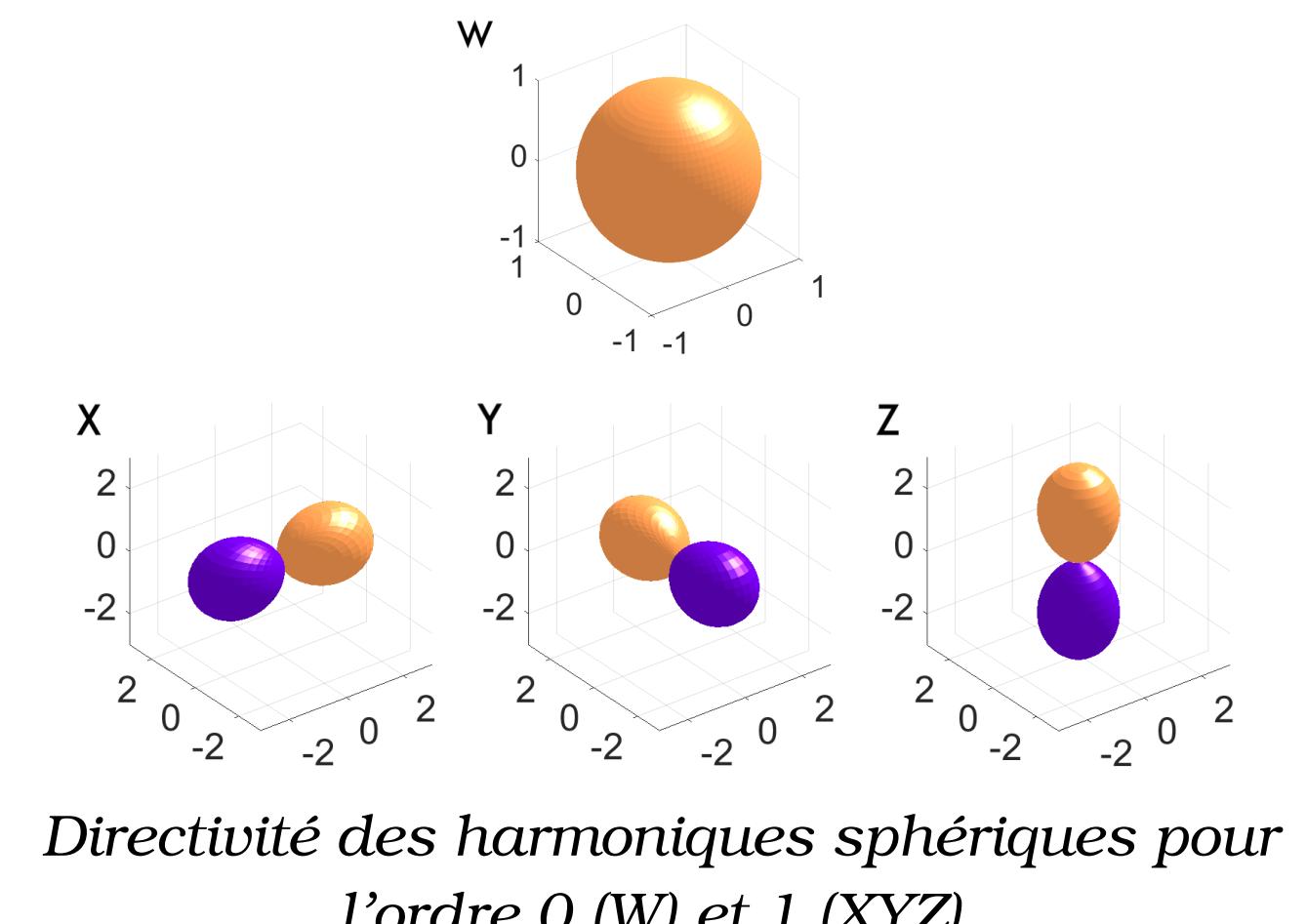
La Rochelle  
Université

## Contexte

Actuellement les codecs téléphoniques sont essentiellement mono, ils sont donc inadaptés pour coder des contenus audio immersifs. Nous proposons d'étudier de nouvelles méthodes de codage permettant d'étendre les codecs téléphoniques existants, tels que *EVS* ou *Opus*, pour permettre de traiter le son spatialisé.

## Ambisonique d'ordre 1

**L'ambisonique** [1] est une représentation d'une scène audio 360° basée sur une décomposition du champ sonore selon une base d'harmoniques sphériques.



Cette décomposition peut être apparentée à une transformée de Fourier sur la sphère (dans le domaine spatial). Il est possible de tronquer cette décomposition pour ne garder que le 1<sup>er</sup> ordre (*First-Order Ambisonic* ou *FOA*).

## Évaluer la qualité audio

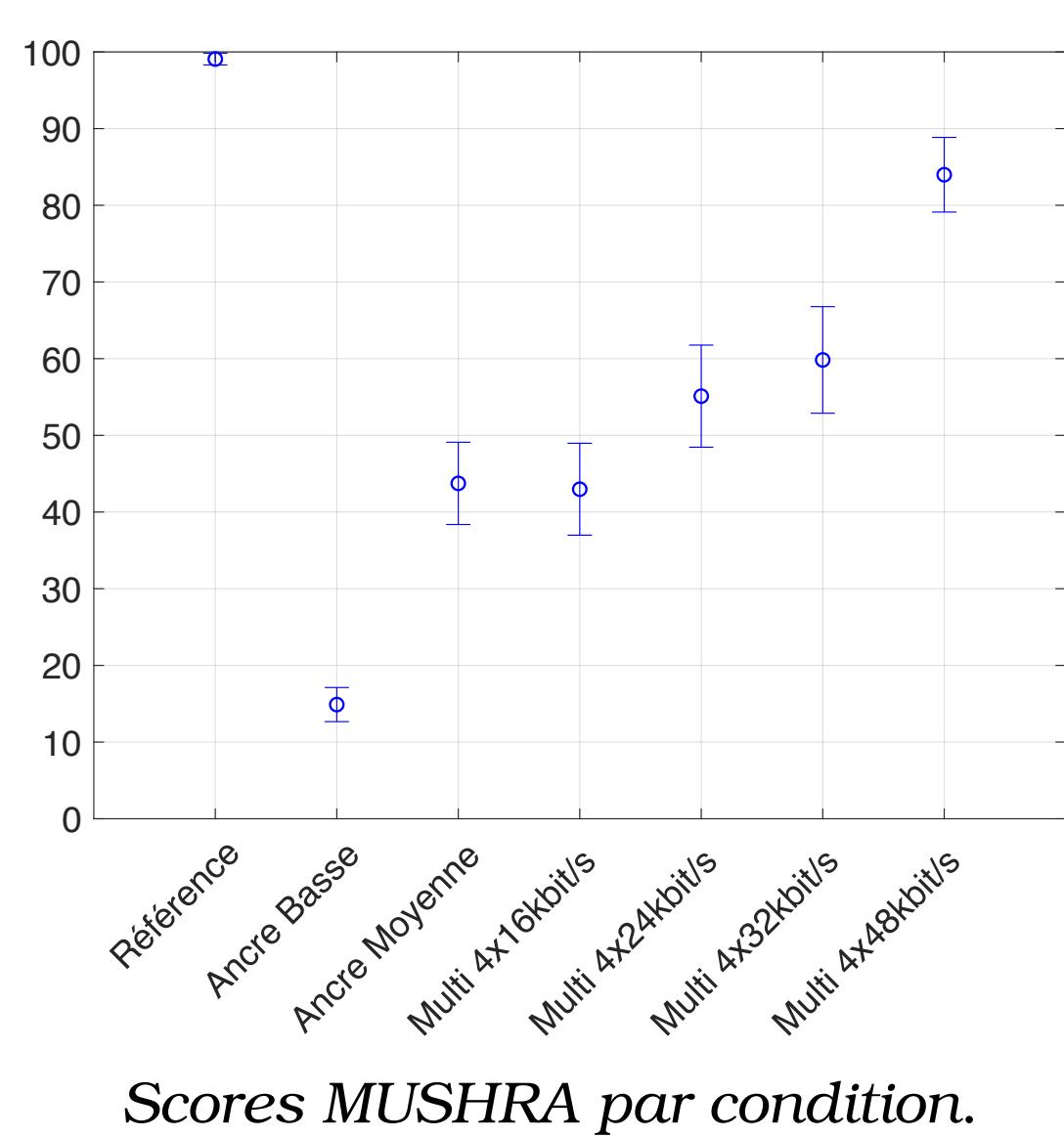
Pour les codecs avec perte basés sur des critères perceptuels, il n'est pas possible d'utiliser une mesure de qualité objective (type erreur quadratique moyenne). Pour comparer les signaux codés il faut donc mettre en place des tests perceptifs.

### Méthodologie MUSHRA :

- Pour chaque échantillon, les sujets notent la qualité des conditions codées par rapport à un original servant de référence. La note reflète la qualité globale ce qui englobe à la fois la qualité du timbre et la fidélité spatiale.
- Les notes vont de 0 à 100, avec un découpage en 5 sections allant de mauvais (0-20) à excellent (80-100).
- Pour chaque échantillon sonore, il existe 3 conditions particulières pour calibrer le test : la référence cachée et deux ancrés (original limité à 3.5 et 7 kHz avec une déformation spatiale).

## Une approche naïve n'est-elle pas suffisante ?

Un test MUSHRA préliminaire a été mené pour étudier la qualité audio d'un codage naïf multi-mono, où chaque composante ambisonique est codée indépendamment.



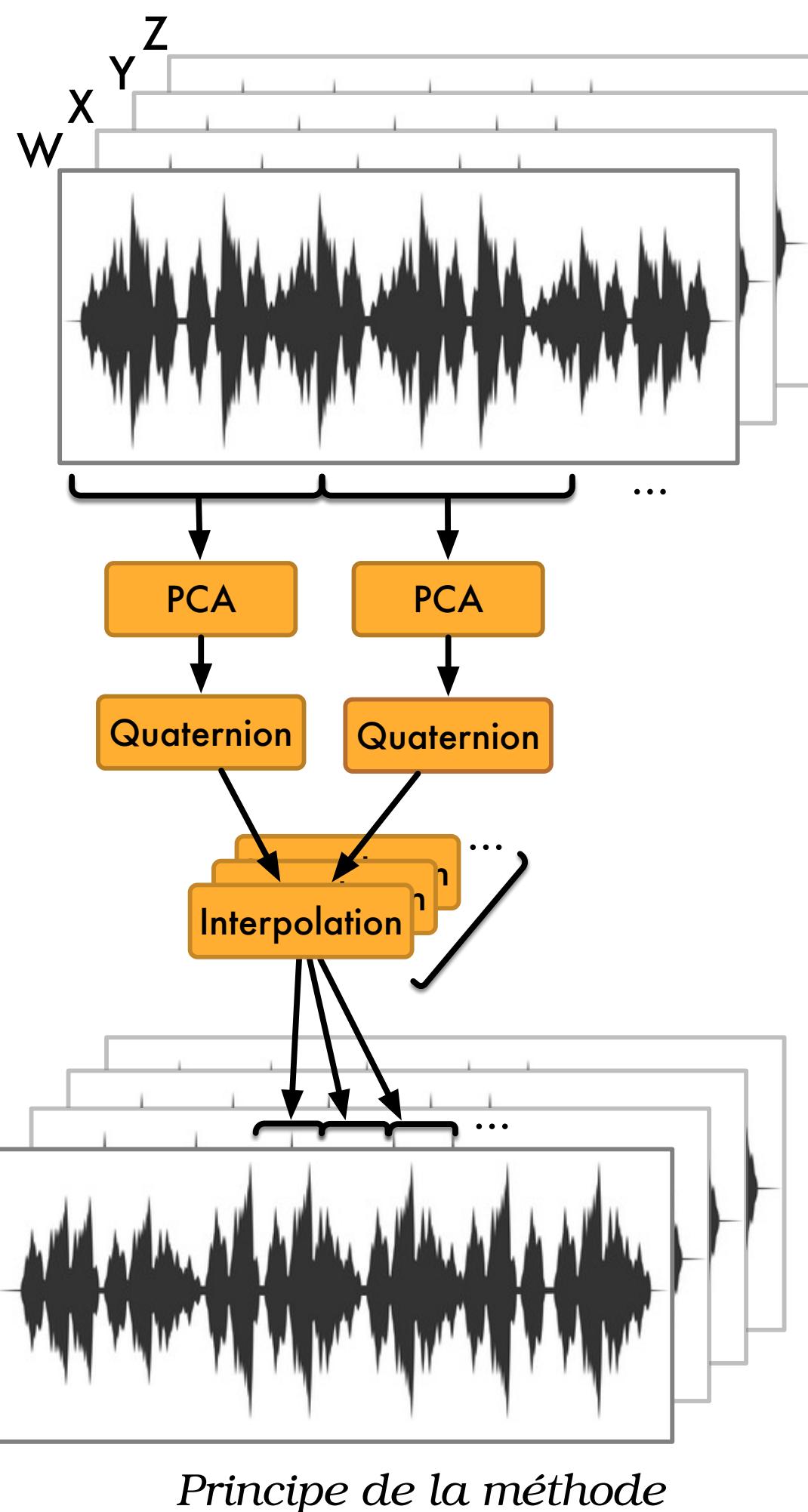
Les signaux FOA (4 canaux) sont codés séparément par le codec EVS à un débit de 4x 13.2, 16.4, 24.4 ou 32 kbit/s.

### Conclusion :

- Qualité acceptable (score >80) uniquement pour un débit élevé (192kbit/s).
- Apparition d'artefacts spatiaux : bruits diffus, recentrage des sources, sources fantômes.

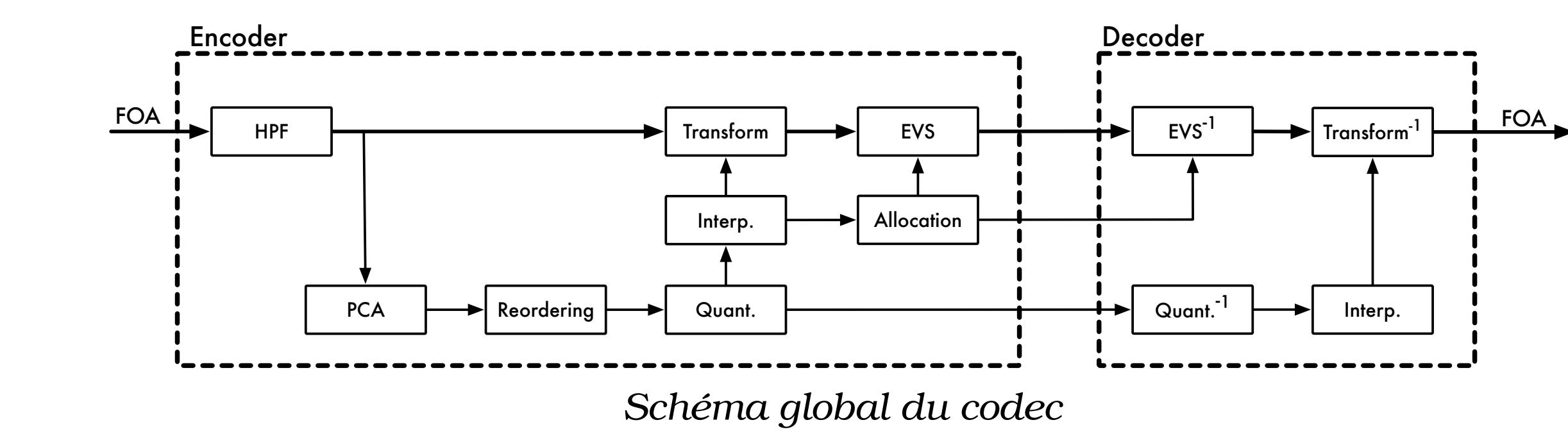
## Méthode proposée : approche par PCA

Notre méthode s'inspire des travaux [2] sur le rematriage des composantes par décomposition en valeurs singulières dans le domaine fréquentiel. Nous utilisons une approche temporelle pour minimiser le retard de codage.



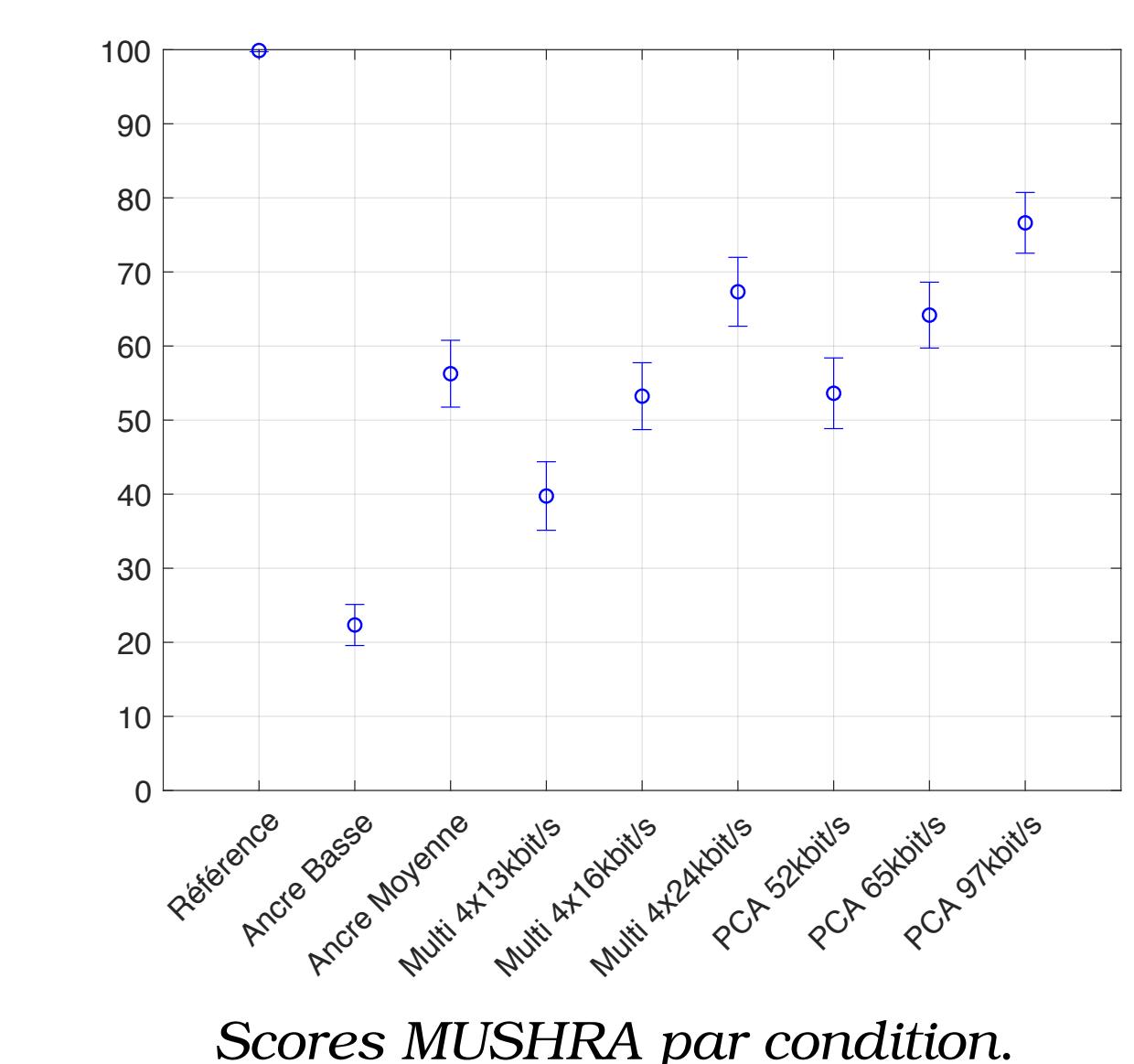
- Pour chaque trame, la matrice de vecteurs propres est calculée par PCA. Cette matrice permet de projeter les composantes ambisoniques dans un domaine où elles sont décorrélatées.
- Pour éviter les discontinuités du signal entre trames, deux mécanismes sont mis en place :
  - Un ré-alignement des vecteurs propres entre trames pour éviter les permutations de canaux et les inversions de signe.
  - Une interpolation des matrices de rotation est faite dans le domaine des quaternions. Chaque matrice de rotation 4D est décomposée en 2 quaternions par factorisation de Cayley [3], puis les quaternions sont interpolés deux à deux.

## Architecture du codec



## Comparaison avec le codage multimono

Nous avons comparé notre méthode avec la méthode multi-mono à 3 débits différents : 52.8, 65.6 et 97.6 kbit/s.



- L'approche par PCA améliore significativement la qualité audio par rapport à l'approche multi-mono pour cette gamme de débits.
- Les artefacts spatiaux sont en grande partie corrigés dans l'approche par PCA.

Pour étudier ces déformations spatiales, une métrique objective permettant de quantifier ces dégradations est en cours d'exploration, l'idée étant de pouvoir déterminer pour diverses conditions de codage la fidélité de l'image spatiale.

## Références et Publications

### Références

- B. RAFAELY, *Fundamentals of spherical array processing*. Springer, 2015, chap. 8.
- S. ZAMANI, T. NANJUNDASWAMY et K. ROSE, "Frequency domain singular value decomposition for efficient spatial audio coding", in *Proc. WASPAA*, 2017, p. 126-130.

- A. PEREZ-GRACIA et F. THOMAS, "On Cayley's factorization of 4D rotations and applications", *Advances in Applied Clifford Algebras*, t. 27, n° 1, p. 523-538, 2017.

### Publications

- P. MAHÉ, S. RAGOT et S. MARCHAND, "First-Order Ambisonic Coding with Quaternion-Based Interpolation of PCA Rotation Matrices", *EAA SASP*, 2019, Forthcoming
- P. MAHÉ, S. RAGOT et S. MARCHAND, "First-Order Ambisonic Coding with PCA Matrixing and Quaternion-Based Interpolation", *Digital Audio Effects - DAFX*, 2019, Forthcoming



### Plus d'infos et contact :

- mail : pierre.mahe@orange.com
- site : pmea.github.io