Étude comparative de qualité des codecs audio neuronaux – cas de la musique et du contenu mixte parole/musique

Thomas Muller[1, 2], Stéphane Ragot[1], Laetitia Gros[1] et Pascal Scalart[2]

[1]Orange Research, Lannion [2]IRISA – Université de Rennes, Lannion

1. Qualité audio des codecs neuronaux

Contexte et motivations : Le domaine de la compression audio est bouleversé par les réseaux de neurones artificiels. Les codecs audio neuronaux démontrent des performances remarquables sur la parole à très bas débit[1],[2]. La question se pose de la qualité des nouveaux codecs neuronaux qui compressent l'audio (musique, contenu mixte, etc.).

Objectifs de l'étude :
(1) Caractériser la qualité des codecs audio neuronaux sur la musique et le contenu mixte (mélange parole/musique).
(2) Mettre en avant les limites des outils d'évaluation automatique de la qualité audio.

Catégorie	Contenu	Description
1	musique classique	instrumental : orchestre, piano, clavecin, musique médiévale, métallophone
2		vocal : opéra, chorale, voix à capella
3	musique moderne	instrumental : groupe rock, trompette jazz, harmonica, guitare électrique, castagnettes et guitare acoustique, big band
4		vocal : extraits (Jacques Brel, Tracy Chapman, Sarah McLachlan, etc.)
5	contenu mixte	enregistrement naturel : radio, extrait de film, musique d'attente, commentaires sportifs, publicités
6		mélange artificiel parole + musique (rapport signal à bruit de 10 à 25 dB)

2. Test d'écoute subjectif DCR

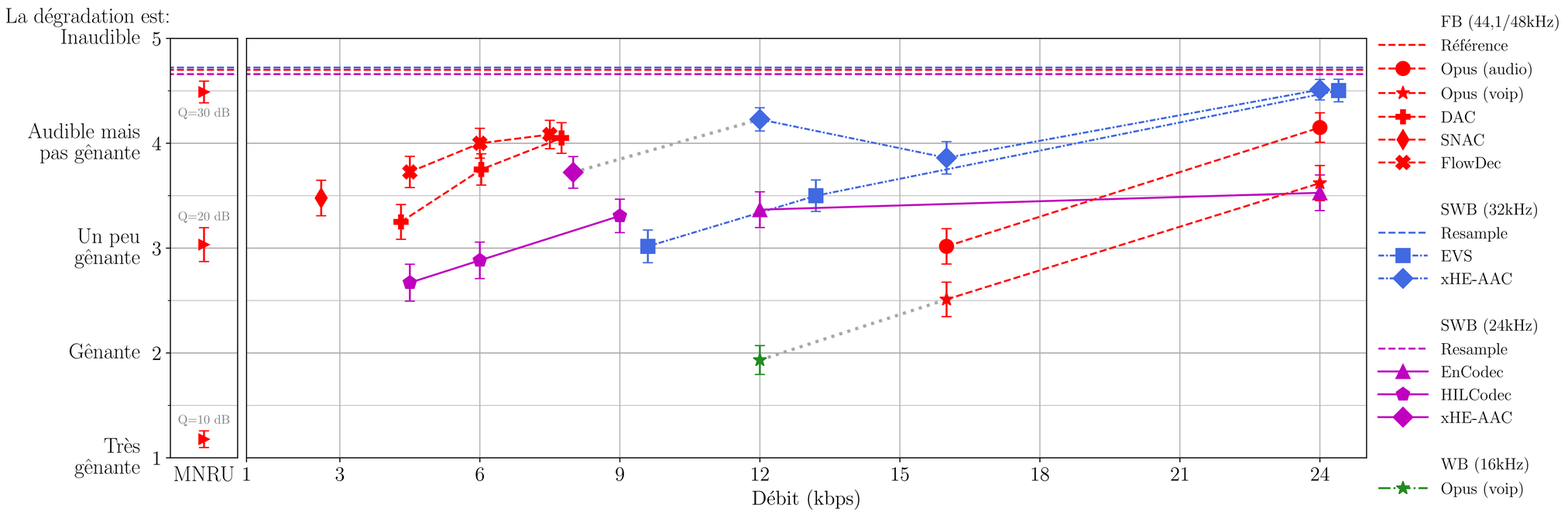
Une évaluation par catégories de dégradation (test P.800 DCR) a été menée pour comparer entre eux cinq codecs neuronaux et trois codecs classiques sur de la musique et du contenu mixte.

Détails du test (∼1h30 de test) :
- 30 auditeurs "naïfs" (5 groupes de 6 auditeurs)
- Salles de test insonorisées
- Casques Sennheiser HD 380 Pro
- Niveau sonore normalisé (73 dB SPL - écoute diotique)

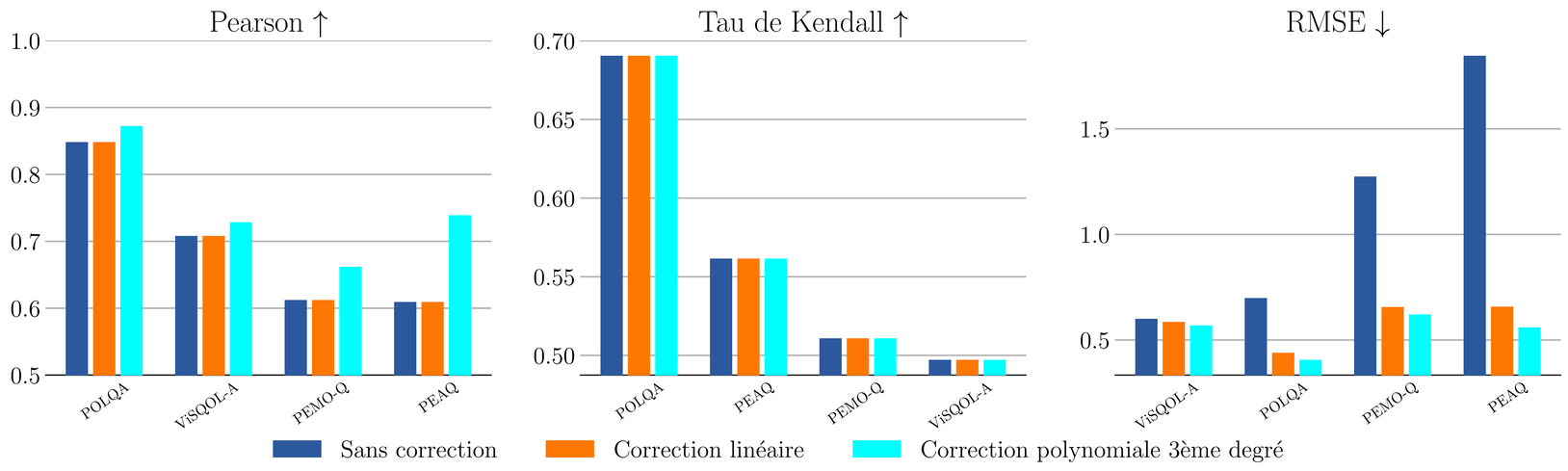
Conditions de test (30 au total) :
- Audio original non codé (Référence) et rééchantillonné à 24 et 32 kHz
- Bruits modulés (P.50 MNRUs) avec $Q = 36$, 23 et 10 dB
- Trois codecs traditionnels et cinq codecs neuronaux

Codec	f_s (kHz)	L (ms)	débit (kbps)
EnCodec	24	13,3	12/24
DAC	44,1	11,6	4,3/6/7,8
HILCodec	24	13,3	4,5/6/9
SNAC	44,1	11,6	2,6
FlowDec	48	13,3	4,5/6/7,5
xHE-AAC	48	16 à 85,3	8/12/16/24
Opus - audio	48	20	16/24
Opus - voip	48	20	12/16/24
EVS	32	20	9,6/13,2/24,4

3. Résultats expérimentaux



Métrique	Contenu	f_s (kHz)
PEAQ	Audio	48
PEMO-Q	Audio	48
ViSQOL-A	Audio	48
POLQA	Parole	48



Conclusion : Les codecs audio neuronaux n'atteignent pas encore une qualité proche de l'audio de référence (non codé), mais proposent une qualité prometteuse à très bas débit comparé aux codecs traditionnels. Les outils d'évaluation automatique de qualité audio testés ne permettent pas une prédiction fiable.

[1]Thomas Muller, Stéphane Ragot, Laetitia Gros, Pierrick Philippe and Pascal Scalart, *Speech quality evaluation of neural audio codecs*, Interspeech 2024
[2]Thomas Muller, Stéphane Ragot, Vincent Barriac and Pascal Scalart, *Evaluation of Objective Quality Models on Neural Audio Codecs*, IWAENC 2024