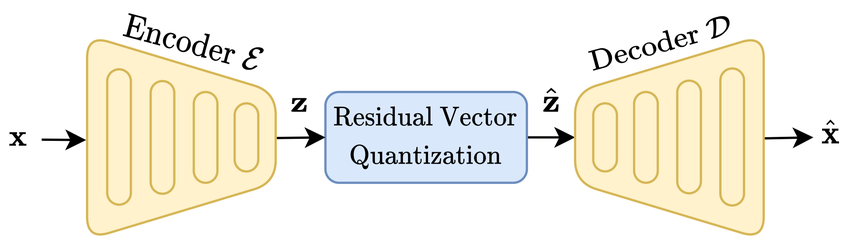
Speech Quality Evaluation of Neural Audio Codecs

Thomas Muller[1, 2], Stéphane Ragot[1], Laetitia Gros[1], Pierrick Philippe[1] and Pascal Scalart[2]

[1]Orange Innovation, France [2]IRISA – University of Rennes, France

1. Neural Audio Coding

Context: A new generation of audio codecs has emerged using artificial neural networks. Neural audio codecs demonstrate promising audio quality at low bitrates at the cost of higher computational complexity compared to traditional audio codecs.



Goal: Characterize the performance of recent neural audio codecs on clean speech. Focus on publicly available implementations.

Codec	f_s (kHz)	L (ms)	bitrate (kbps)
LPCNet	16	10	1.6
Lyra V2	16	20	3.2, 6, 9.2
EnCodec	24	13.3	1.5, 3, 6, 12, 24
AudioCraft	24	13.3	1.5, 3, 6
AudioDec	24	12.5	6.4
DAC	44.1	11.6	1.7, 2.6, 5.2, 7.8
AudioDec	48	6.25	12.8
Opus	48	20	12, 16, 24
EVS-WB	16	20	7.2, 8
EVS-SWB	32	20	9.6, 13.2, 24.4

2. Test methodology and test setup

Choice of test methodology: The aim of the test is to compare several codecs operating with different coded bandwidths and a large range of bitrates (1.5 - 24.4 kbps). Degradation Category Rating (DCR) was chosen for its capability to test a wider range of source material which are multi-bandwidth.

Details on test plan:

Audio corpus:

- Clean speech (French)
- 3 male and 3 female talkers
- 8s phonetically balanced sentence pairs

Preprocessing:

- 20-20,000 Hz bandpass FIR filtering
- Level normalization to -26 dB LKFS

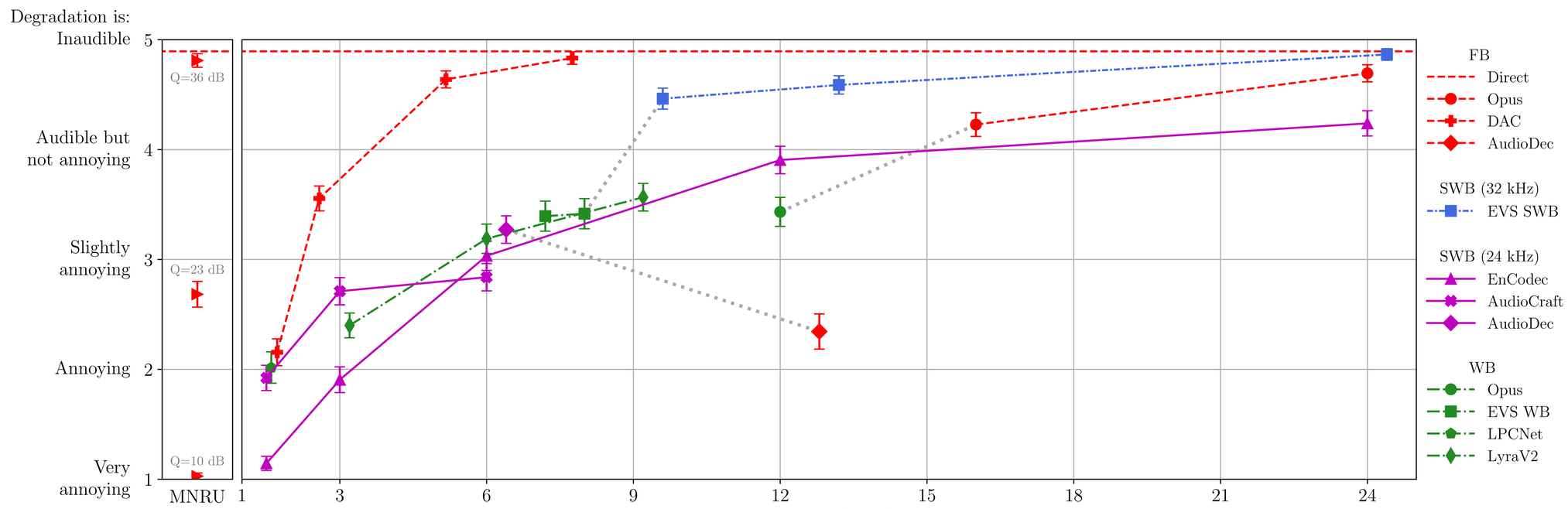
Test setup (∼1.5h of testing for each listener):

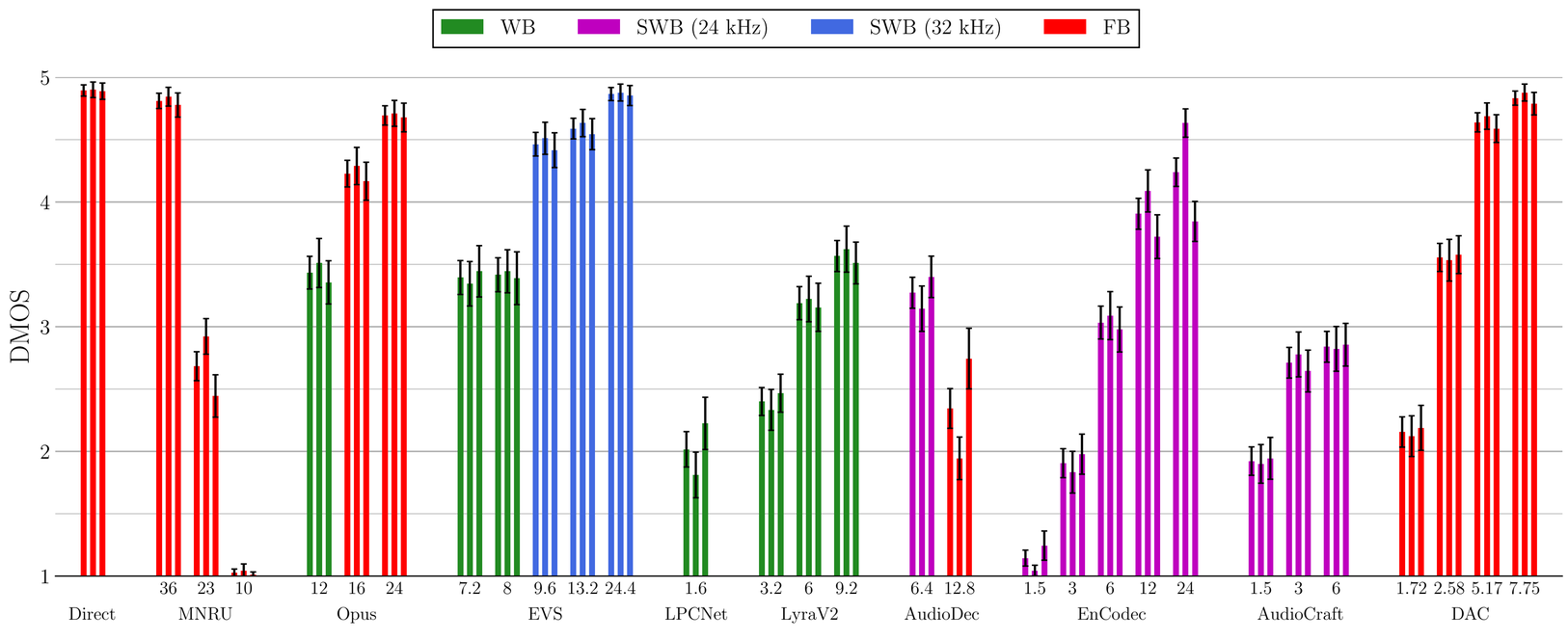
- 30 naive listeners (5 panels of 6 listeners)
- Dedicated soundproof test rooms
- Sennheiser HD 380 Pro headphones
- Diotic listening level of 73 dB SPL

Test conditions (30 conditions overall):

- Uncoded original audio (”Direct”)
- P.50 Modulated Noise Reference Units (MNRUs) with $Q = 36$, 23 and 10 dB
- Two traditional audio codecs: Opus and EVS
- Six neural audio codecs: LPCNet, Lyra V2, EnCodec, AudioCraft, AudioDec and DAC

3. Experimental results





Real-Time Factor (RTF)

Codec	Enc	Dec	Enc+Dec
EVS-WB	99.6	207.8	67.5
EVS-SWB	57.1	135.2	44.0
Opus	32.5	506.8	30.7
Lyra V2	86.1	108.9	48.2
LPCNet	111.1	3.8	3.7
EnCodec	8.9	10.6	5.1
AudioDec 24k	6.04	7.66	3.38
AudioDec 48k	2.80	2.58	1.34
DAC	1.23	0.73	0.46
AudioCraft	8.90	0.04	0.04

Interspeech 2024, Kos Island, Greece