FACULTY OF ENGINEERING
AND ARCHITECTURE

Bitrate-Informed Coded Speech Enhancement Model

Haixin Zhao, Nilesh Madhu

IDLab, Ghent University - imec, Belgium

GHENT
UNIVERSITY

imec

1. Introduction

GHENT UNIVERSITY

imec

PRACTICAL DATA TRANSFER

Bandwidth limitations → lossy codecs (AMR-WB, LC3+, opus, EVS, …)
Received signal distorted → poorer listening experience, fatigue
Multi-channel processing → can distort spatial information

Fig. 1. Lossy codec framework

Aims

- ‘Clean-up’ *received* signal
  - Richer listening experience & lower listening fatigue
  - Improve intelligibility

Enhancement of (De)Coded Speech

DNN-based enhancement

- Data-driven → implicitly learn codec “behaviour”
- End-to-end or gain-based methods
- **BUT:** Enhancement models for different bitrates (generalisability across bitrates)
  - trained on lower bitrate, and apply for all bitrates
  - multiple bitrate-specific models

Solutions

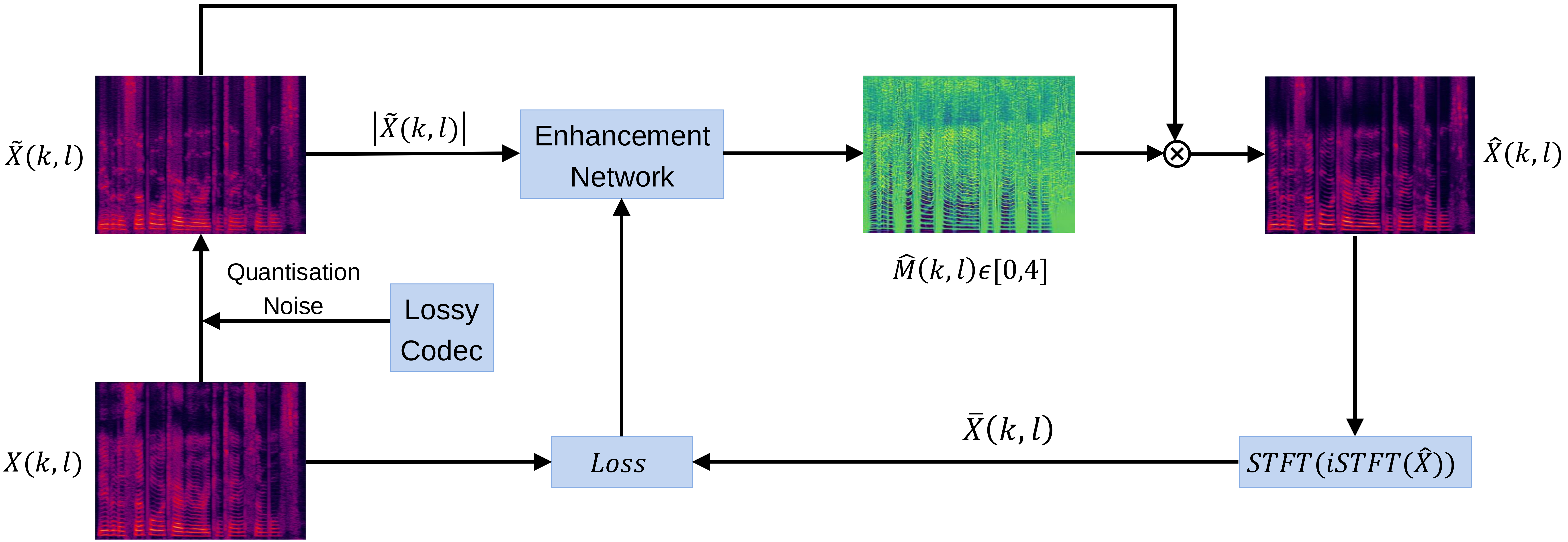
- Multi-condition training (all bitrates shared one enhancement model)
  → Loss bitrate-specific information
  → Can not access best performance for each bitrate
- Utilise bitrate information (utterance level)
  → Integrate bitrate information by fusion methods
  → Make *part* of network bitrate specific

2. Proposed Method

GHENT UNIVERSITY

imec

Generic enhancement framework



$$L = (1-\lambda)\sum_{k,l}(|X(k,l)|^c - |\bar{X}(k,l)|^c)^2 + \lambda \sum_{k,l}\left(|X(k,l)|^c e^{j\phi_X} - |\bar{X}(k,l)|^c e^{j\phi_{\bar{X}}}\right)^2$$

Magnitude estimation loss — Phase-aware term

Fig. 2. Generic enhancement framework

Bitrate-informed CRUSE model

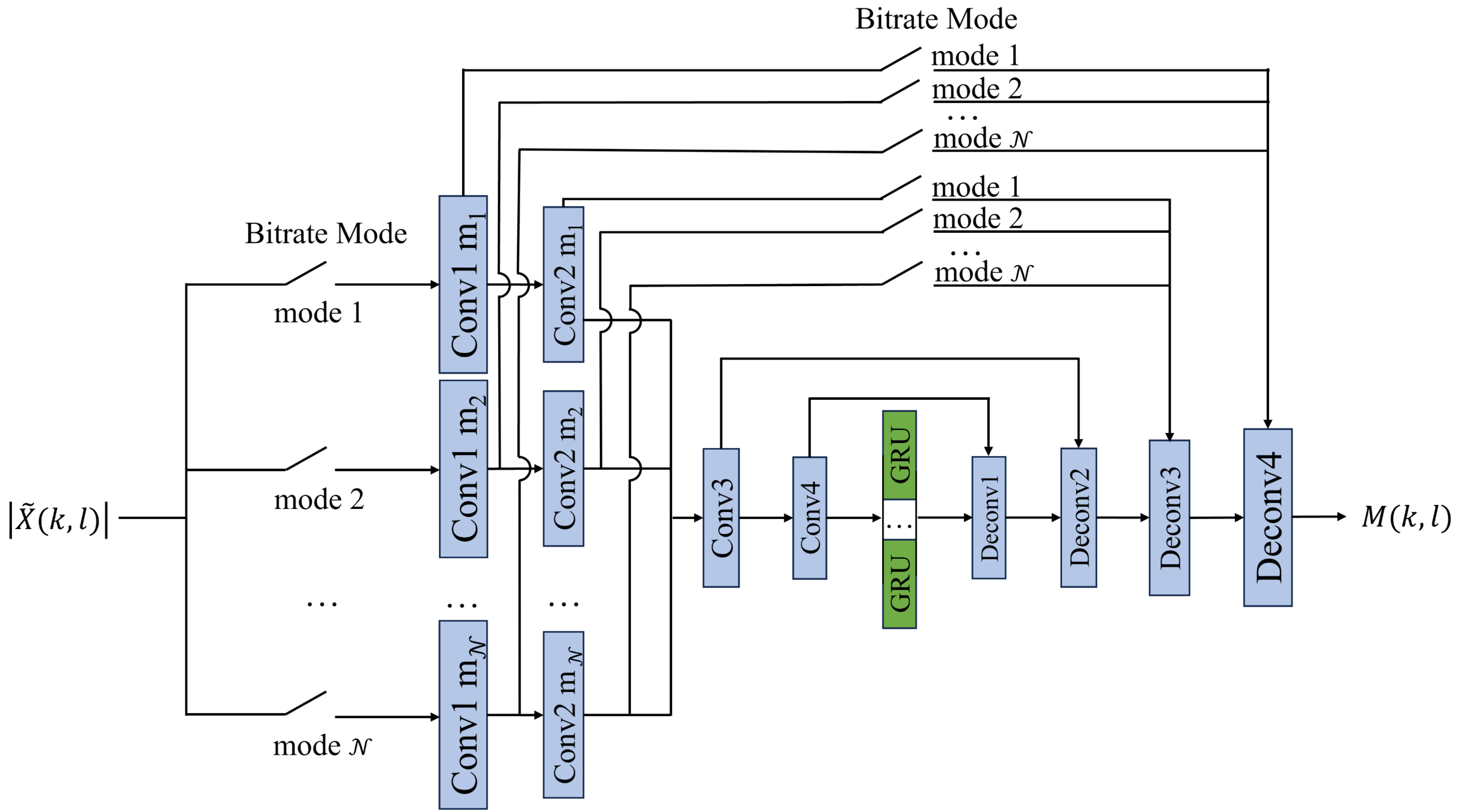


Fig. 3. Bitrate-informed CRUSE model

(first two convolution blocks are with parallel bitrate-dependent layers)

Experiments Setup

Dataset: TIMIT

Codecs

- **LC3+:** 16, 24, 32, 48, 64 kbps
- **AMR-WB:** 6.65, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 kbps

Baselines

- **CED:** the state-of-the-art baseline [1]
- **CRUSE-MSE:** CRUSE model with the MSE loss function for ablation study
- **CRUSE:** CRUSE model with the combined compressed MSE and phase-aware loss function trained by various-bitrate coded speech.

[1] S. Korse, K. Gupta and G. Fuchs, "Enhancement of Coded Speech Using a Mask-Based Post-Filter," ICASSP 2020, pp. 6764-6768.

Evaluation on LC3+ & AMR-WB

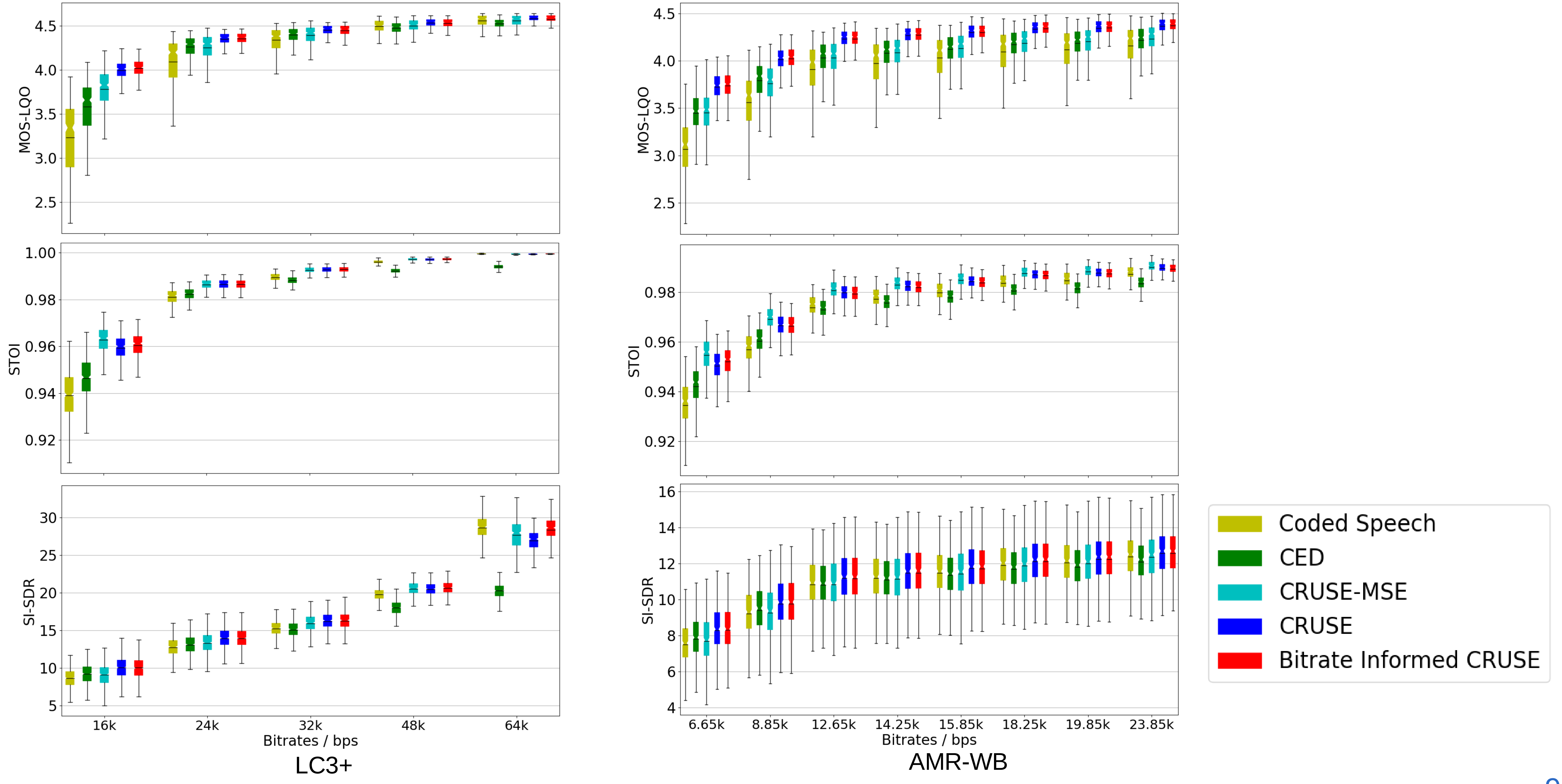


Fig. 4. Evaluation results on LC3+ & AMR-WB codecs in WB-PESQ and STOI

Bitrate (Utterance-Level) Information Utilisation

- **Parallel Bitrate-Dependent Block (PBD)**
  - Common dependency extraction across bitrates
  - Extent of bitrate information utilisation

- **Bitrate Gating Block (BG)**

- **1-Hot Vector Block (1-H)** with FiLM (Feature-wise Linear Modulation)

Bitrate Gating Block

- Continue to use parallel bitrate-dependent layers structure – inter-bitrate dependency
- Remain the primary encoder layer (shared for all bitrates modes) – common dependency
- Combine them with soft gating with bias
- Make scaling & bias values from local T-F bins, not sharing for all T-F bins

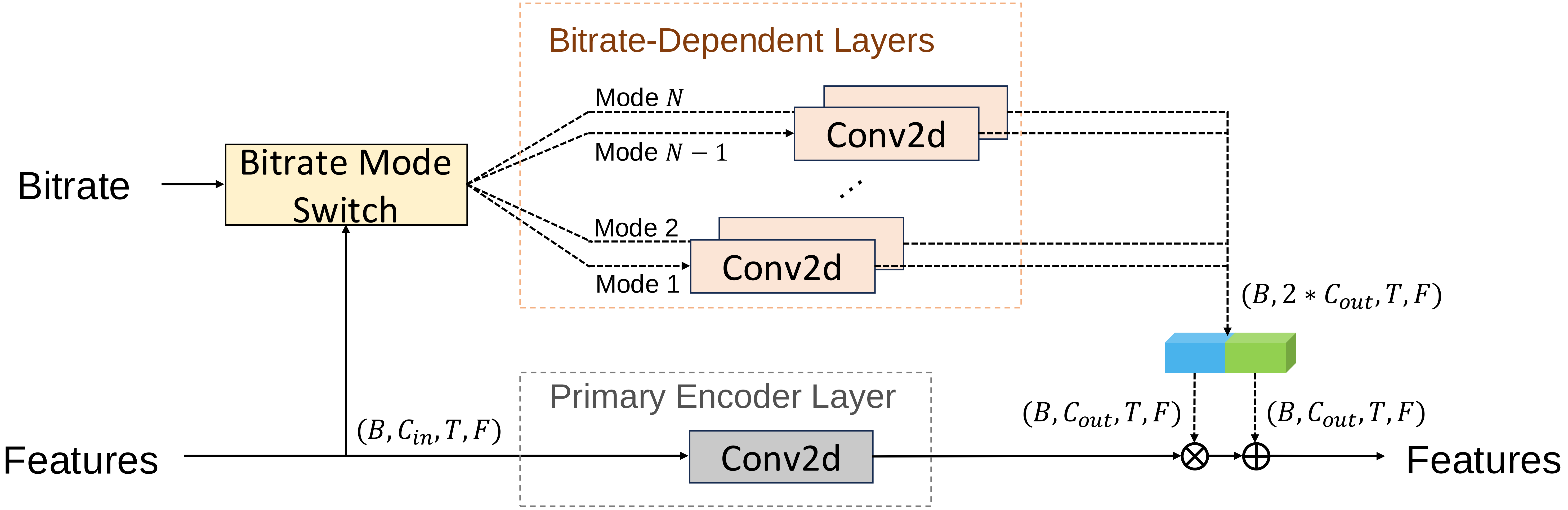


Fig. 5. Structure of Bitrate-Gating convolution block

1-Hot Vector Block with FiLM

- Fusion method, bitrate information is used as one of the inputs
- 1-hot vector – utterance level information
- Share the same information for each T-F bin
- Feature-wise linear modulation

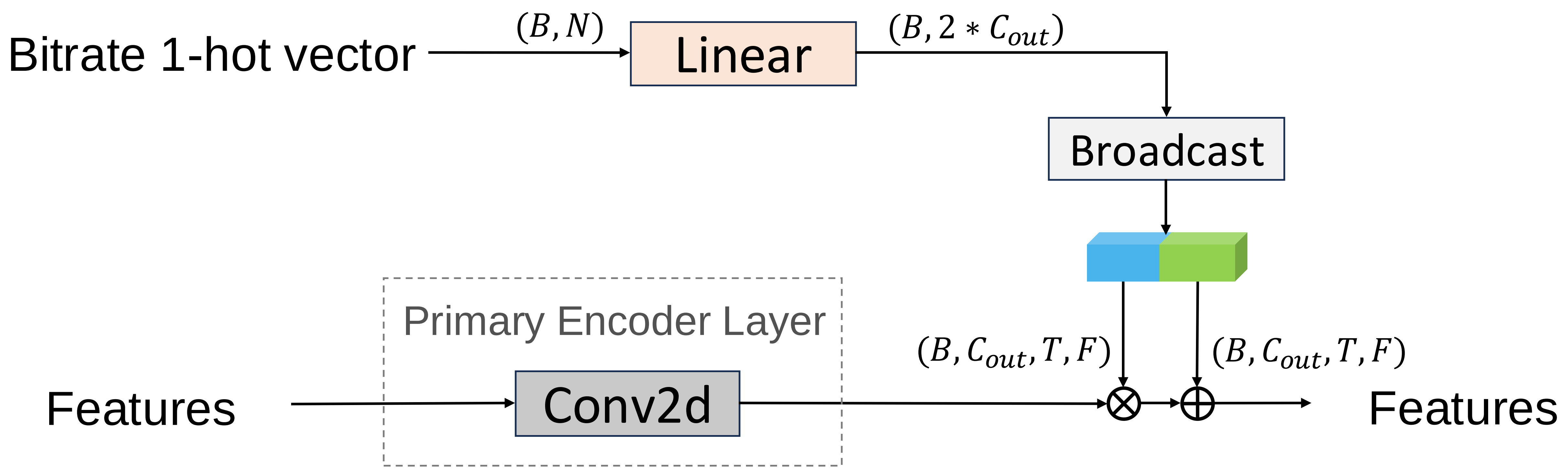


Fig. 6. Structure of 1-hot vector block with FiLM

3. Experimental Results

GHENT UNIVERSITY

imec

Experiments in Bitrate-Informed methods

Dataset: DNS3 challenge (140 hours, wideband)

Codec:

- **Opus:** 6, 9, 12, 16 kbps

Bitrate-informed methods:

- Parallel Bitrate-Dependent Block **(PBD)**
- Bitrate Gating Block **(BG)**
- 1-Hot Vector Block **(1-H)** with FiLM (Feature-wise Linear Modulation)

Baselines:

- Multi-Conditional Training **(MCT)**
- Bitrate-Specific Training **(BST)**

Validation Loss

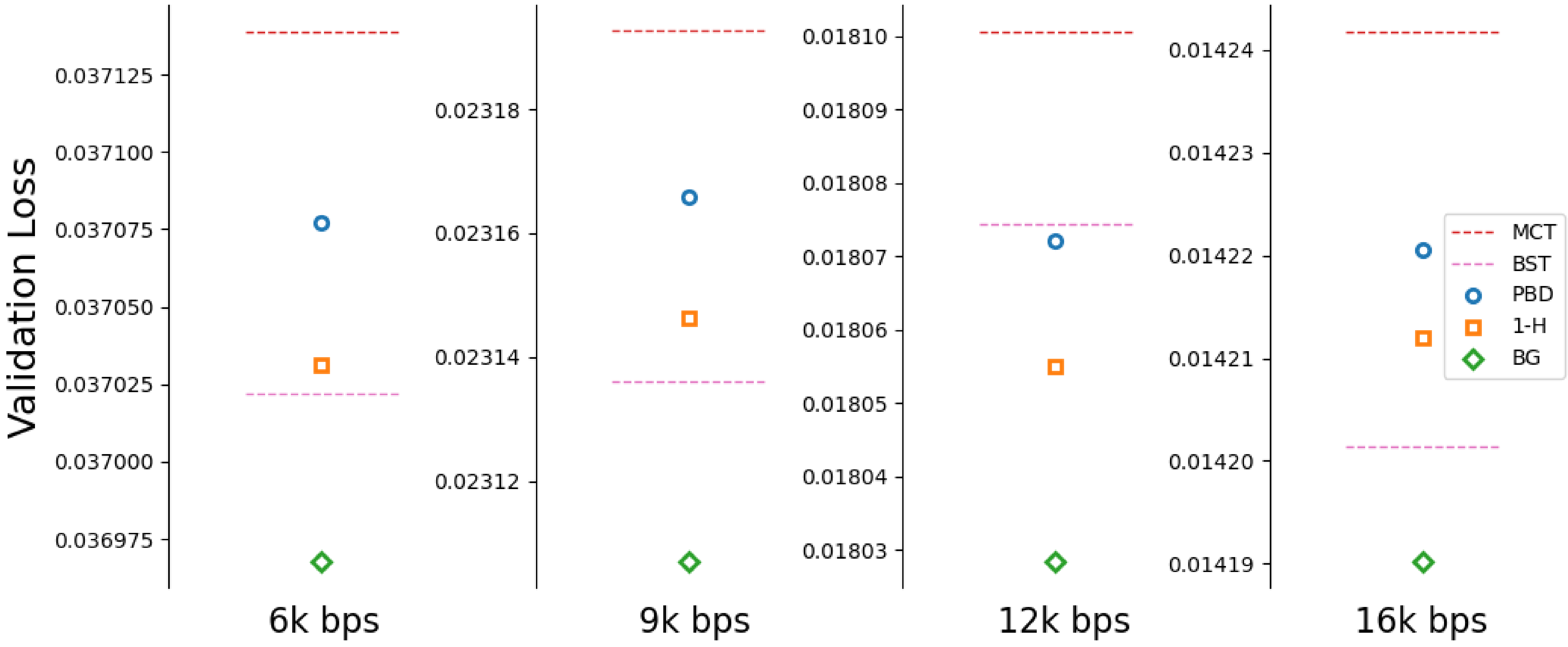


Fig. 7. Validation loss of bitrate-informed methods and baselines after convergence

Evaluation in instrumental metrics

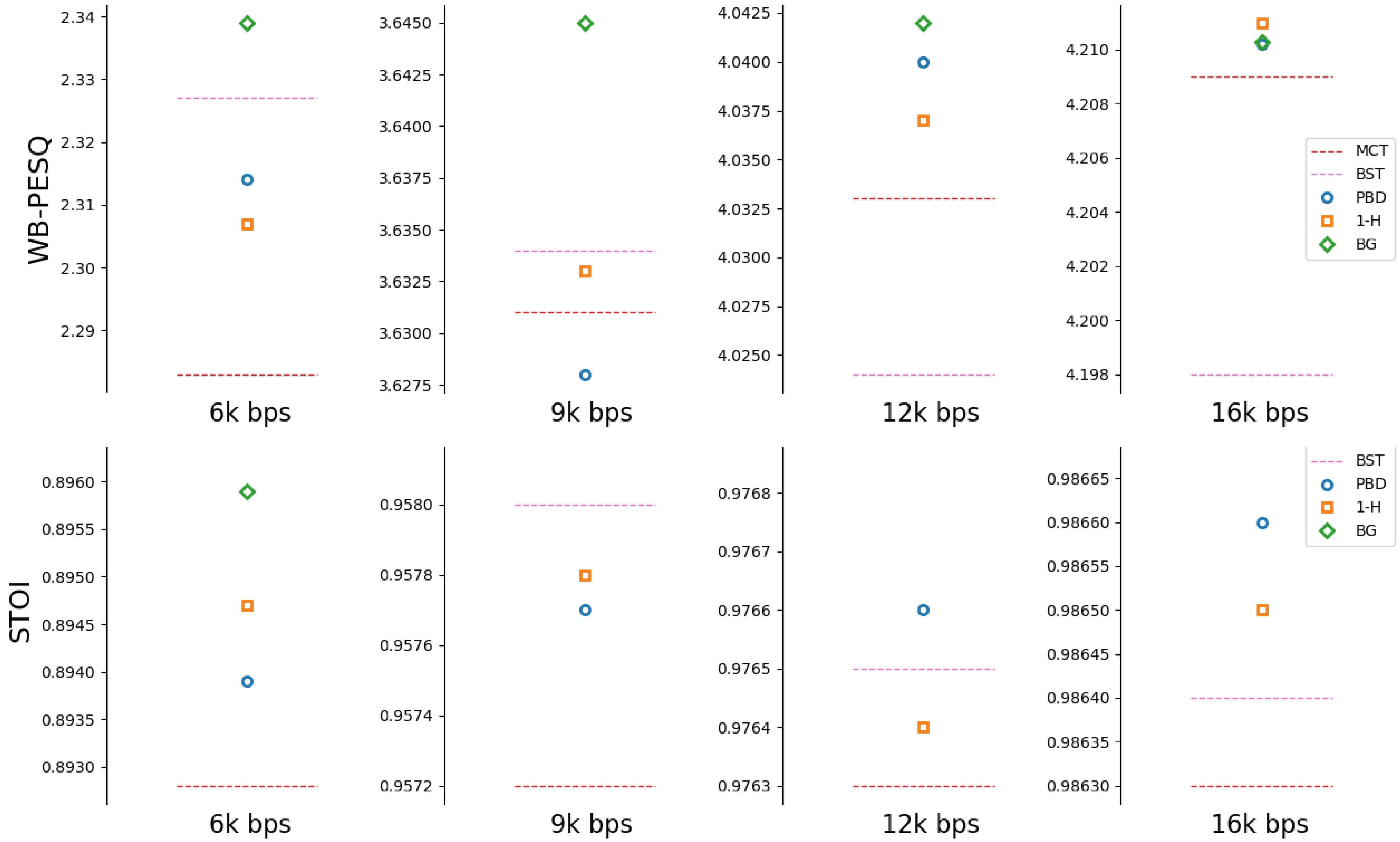
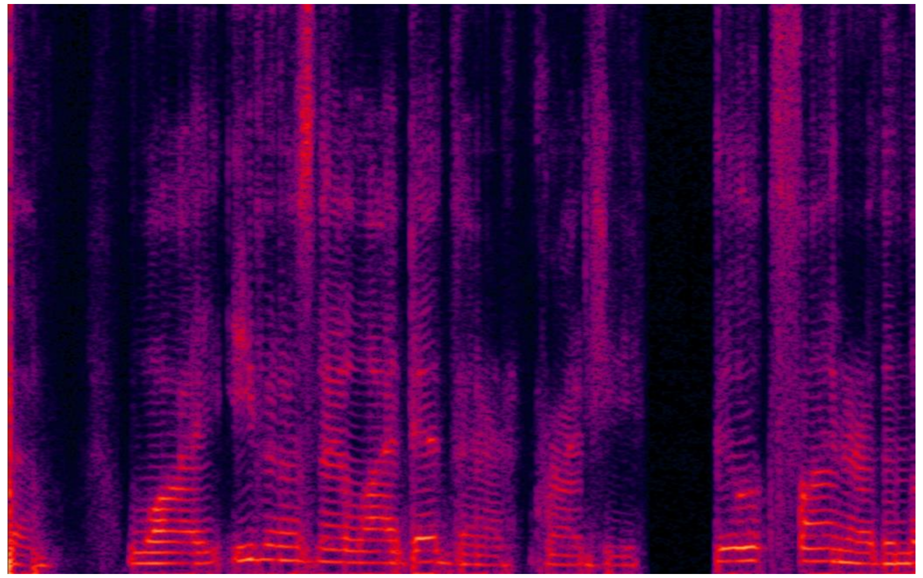


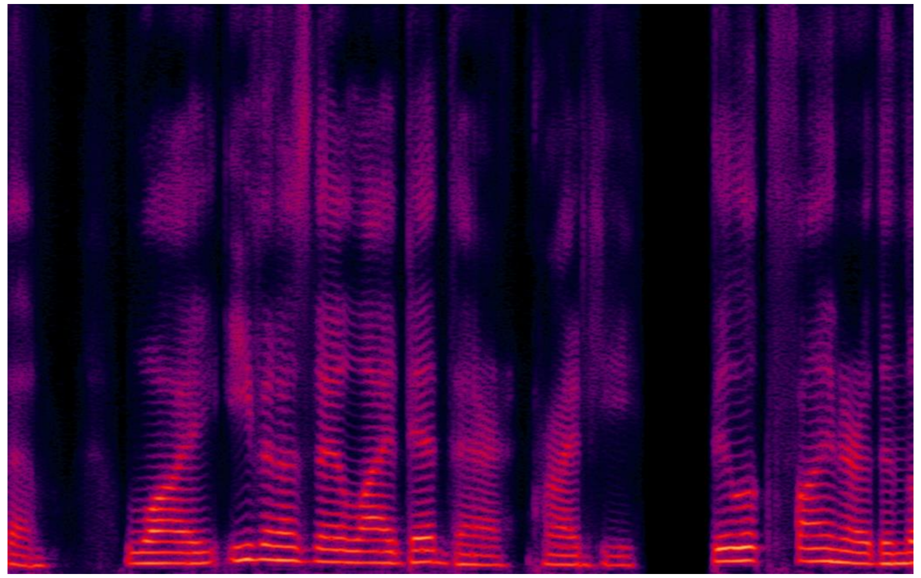
Fig. 8. Evaluation results of bitrate-informed methods and baselines in WB-PESQ & STOI

Audio Samples

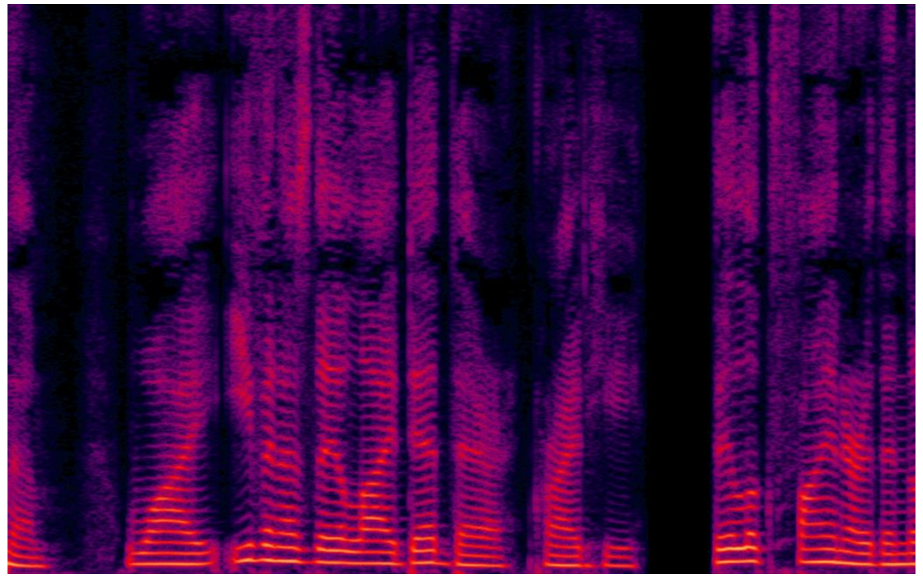
Sample 1
(6 kbps)



Coded speech

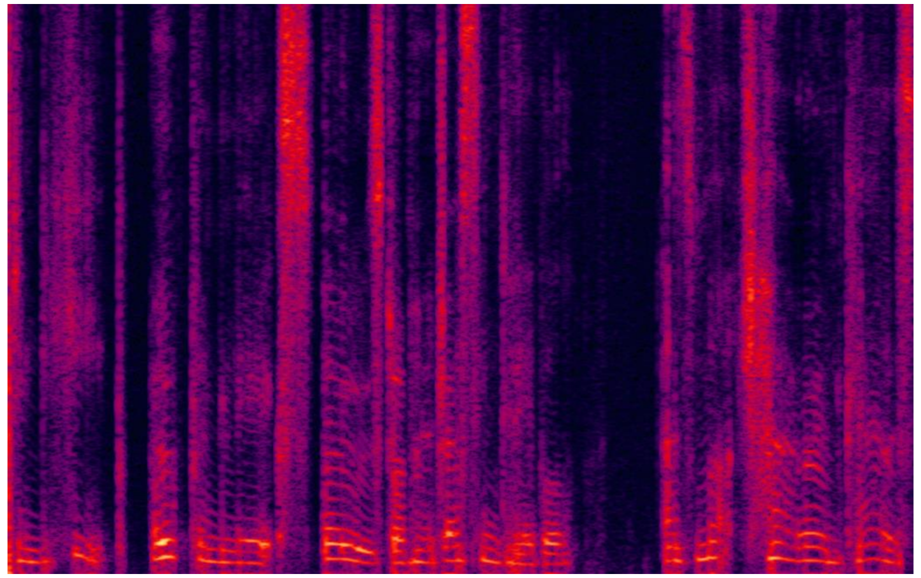


Enhanced speech

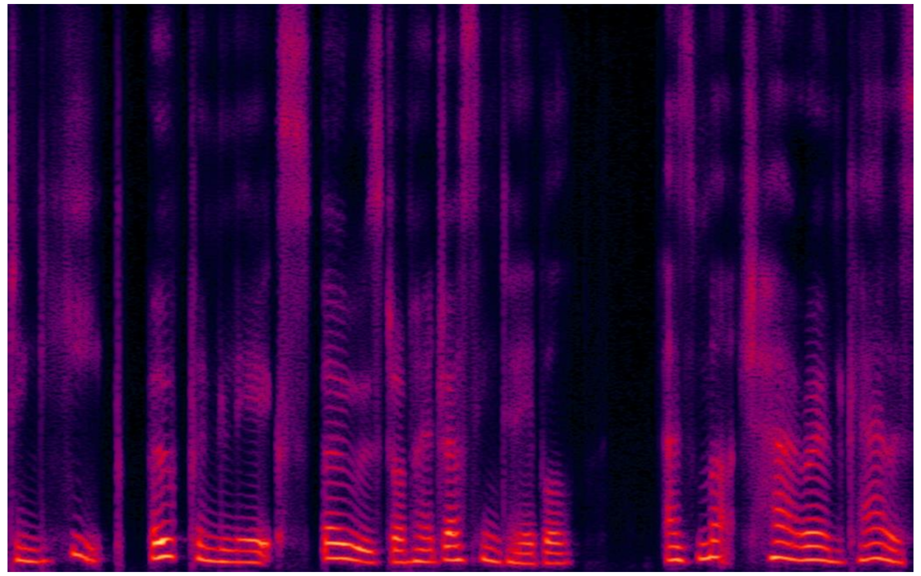


Clean speech

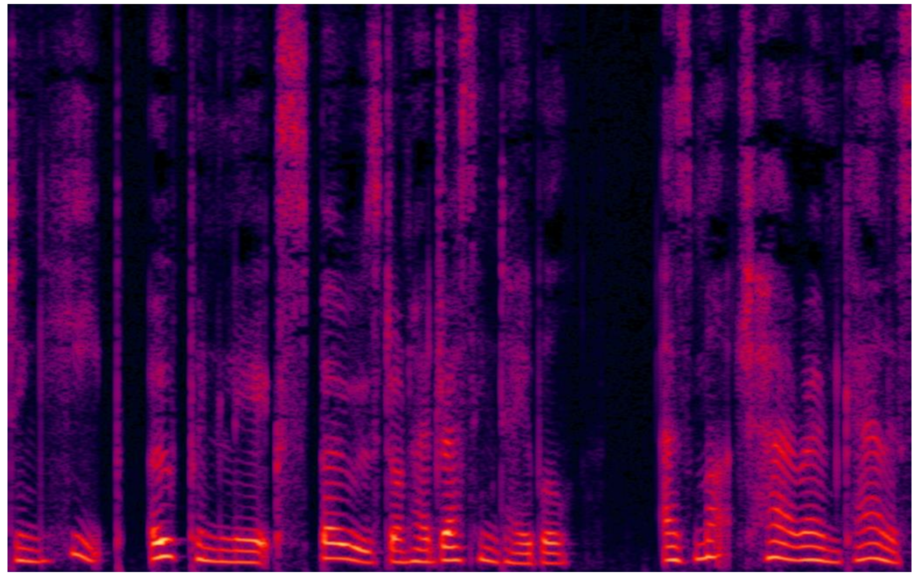
Sample 2
(6 kbps)



Coded speech



Enhanced speech



Clean speech

4. Conclusions

GHENT UNIVERSITY

imec

Conclusions

- Bitrate-informed methods improved performance on instrumental metrics: WB-PESQ, STOI, with little parameter and complexity and increase (1% model footprint, and 2% in MACs)
- Introduced joint exploitation mechanism within the enhancement network to capture both Inter-bitrate & common dependency, showing better information exploitation and generalisability.
- Compared with other bitrate-informed methods, the Bitrate-Gating block maximises the utilisation of the bitrate information.
- Bitrate-gating can even outperform the bitrate-specific trained models in validation loss and instrumental metrics – showing the largest improvement extent from bitrate information.

FACULTY OF ENGINEERING
AND ARCHITECTURE

Haixin Zhao and Prof. Dr.-Ing. Nilesh Madhu

ASPIRE, IDLab

haixin.zhao@ugent.be
nilesh.madhu@ugent.be

https://aspire.ugent.be/

Universiteit Gent
@ugent
@ugent
Ghent University

GHENT
UNIVERSITY

imec