

Real-Time Perceptually Motivated Neural Network for Echo and Noise Reduction

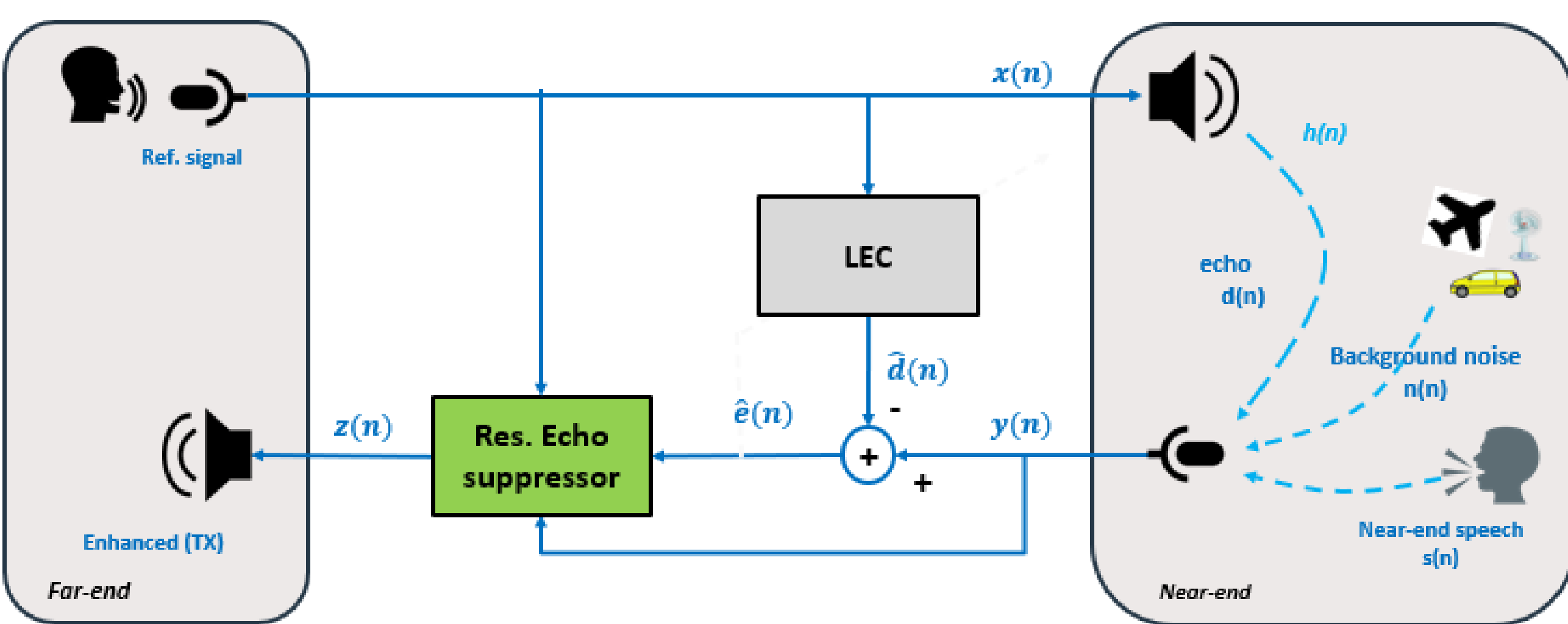
Pejman Mowlae, Robert James,



Motivation and Introduction

- Echo leak and noise are major obstacles in speech communication.
- Reliable and efficient denoising and echo control solutions are required to enable a natural conversation between near-end and far-end and to avoid participant hearing noise and delayed echo of their own voice.

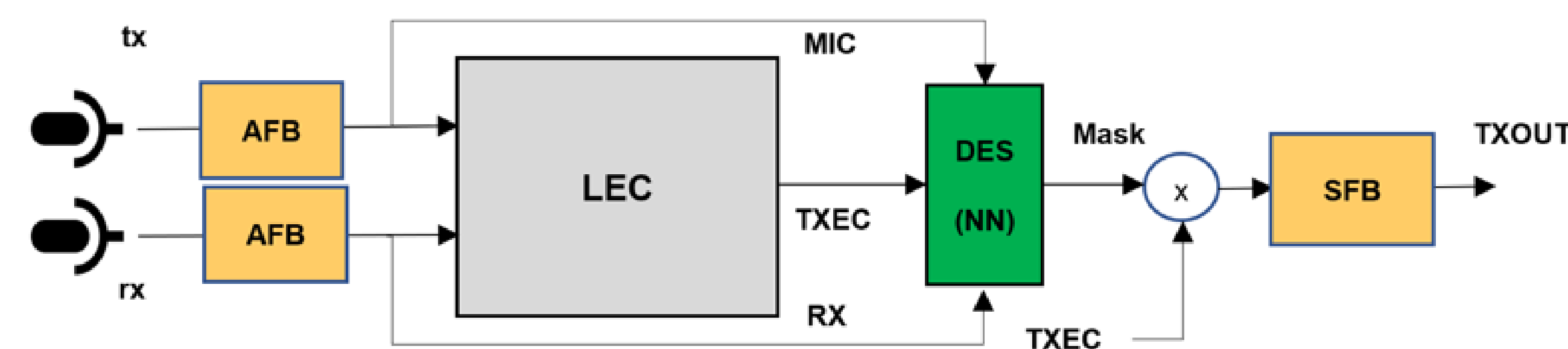
Presented: **Real-time AI echo control and noise reduction**
 Allowing experience own voice echo-annoyance vs. near-end preservation



Proposed Design: DSP + AI

Our Proposal: Hybrid design composed of two stages

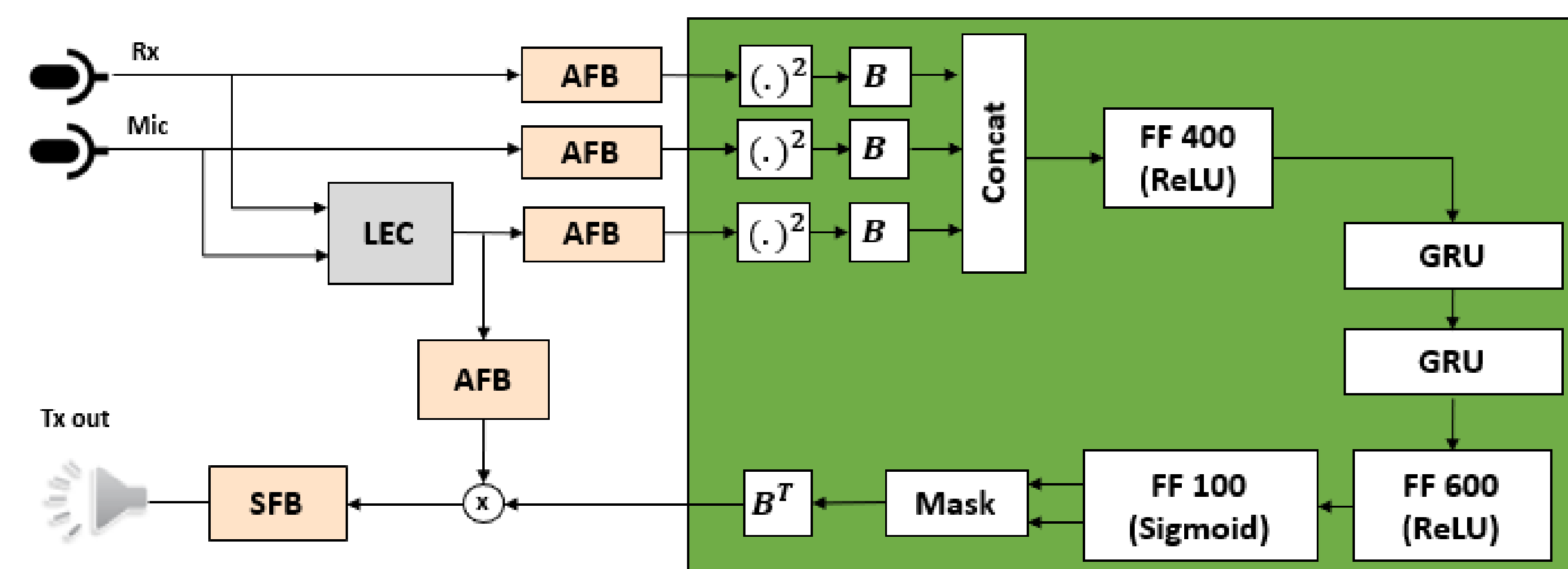
- DSP:** Linear echo canceller (LEC)
- AI:** Neural network (NN) based postfilter a.k.a. deep echo suppression



AFB/SFB: Analysis/Synthesis filter bank

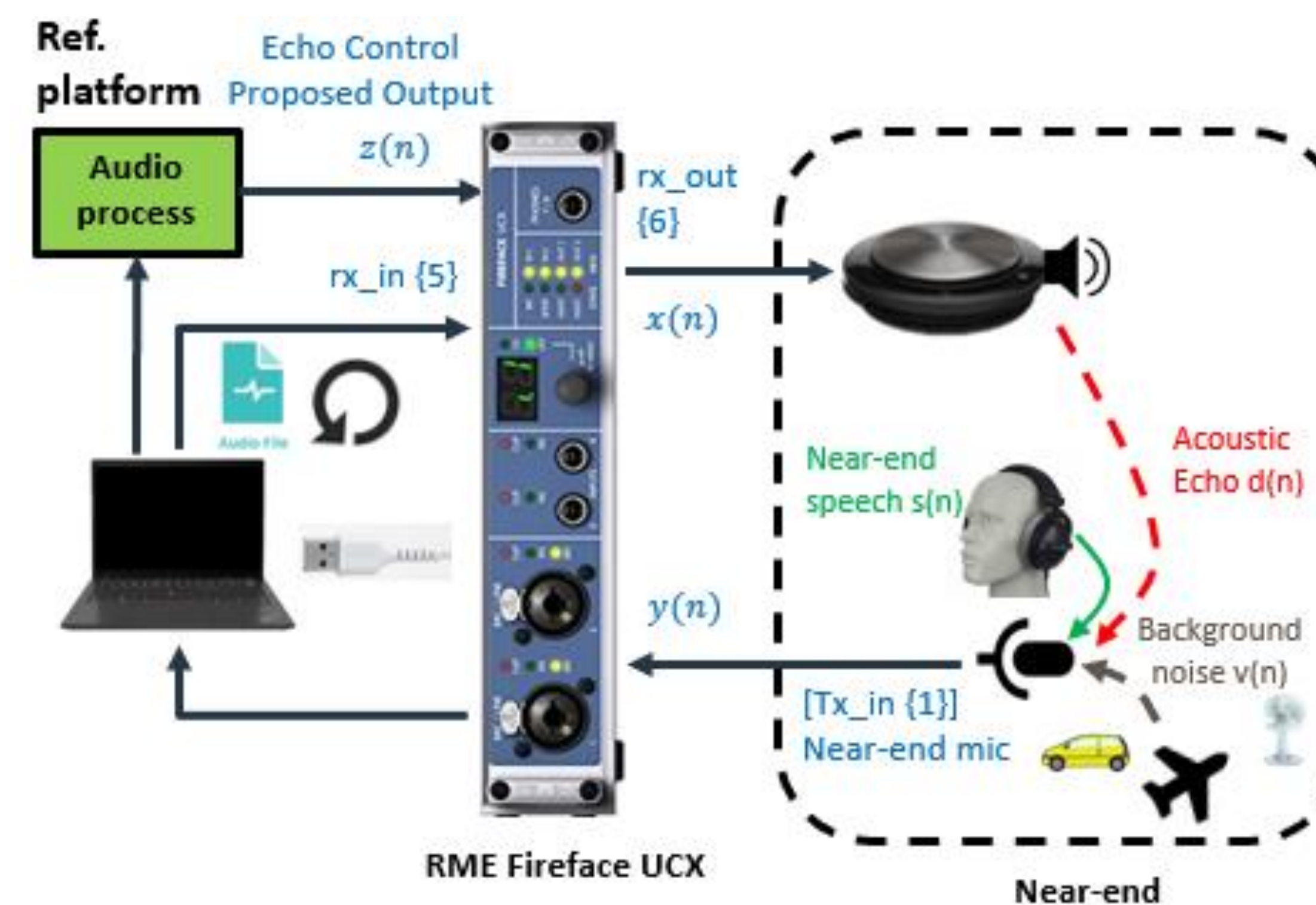
Neural Network Residual Echo Suppressor

- NN Topology:** 2 x GRU layers + 3 x FF layers
- Train loss:** Complex compressed MSE (NN output and NE ref.)
- Frame setup:** block length = 32 ms, overlap = 75%
- B:** mel transform, #mel = 100
- Design Spec:** MACs: 235 (M), Flops: 413 (M), #Param = 1.6 M



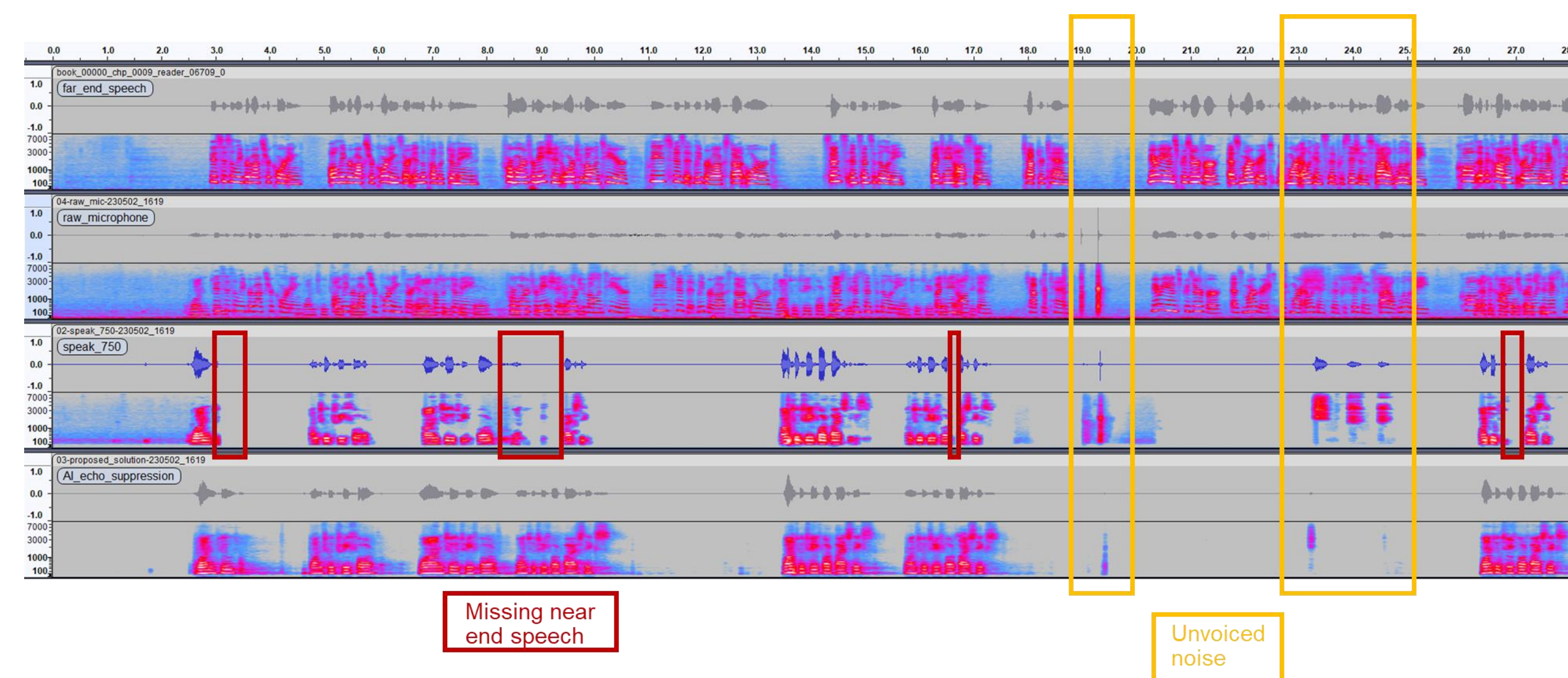
Real-time demo setup

- Modes:** 1) Bypass, 2) Speak750 NREC, 3) Proposed
- Setup:** Speak 750 + PCB amplifier + RME soundcard
- Latency:** processing update every 16 ms
- Directly listening to the Tx out (processed audio by the proposed method) via headset connected to sound card.



Real-time demo on reference platform

Scenarios covered: Single-talk near-end/far-end (EQUEST), double-talk nom./max., hovering on device, speaker movement, echo path change, ...



Model Optimization and Exploration

Studied Scenarios

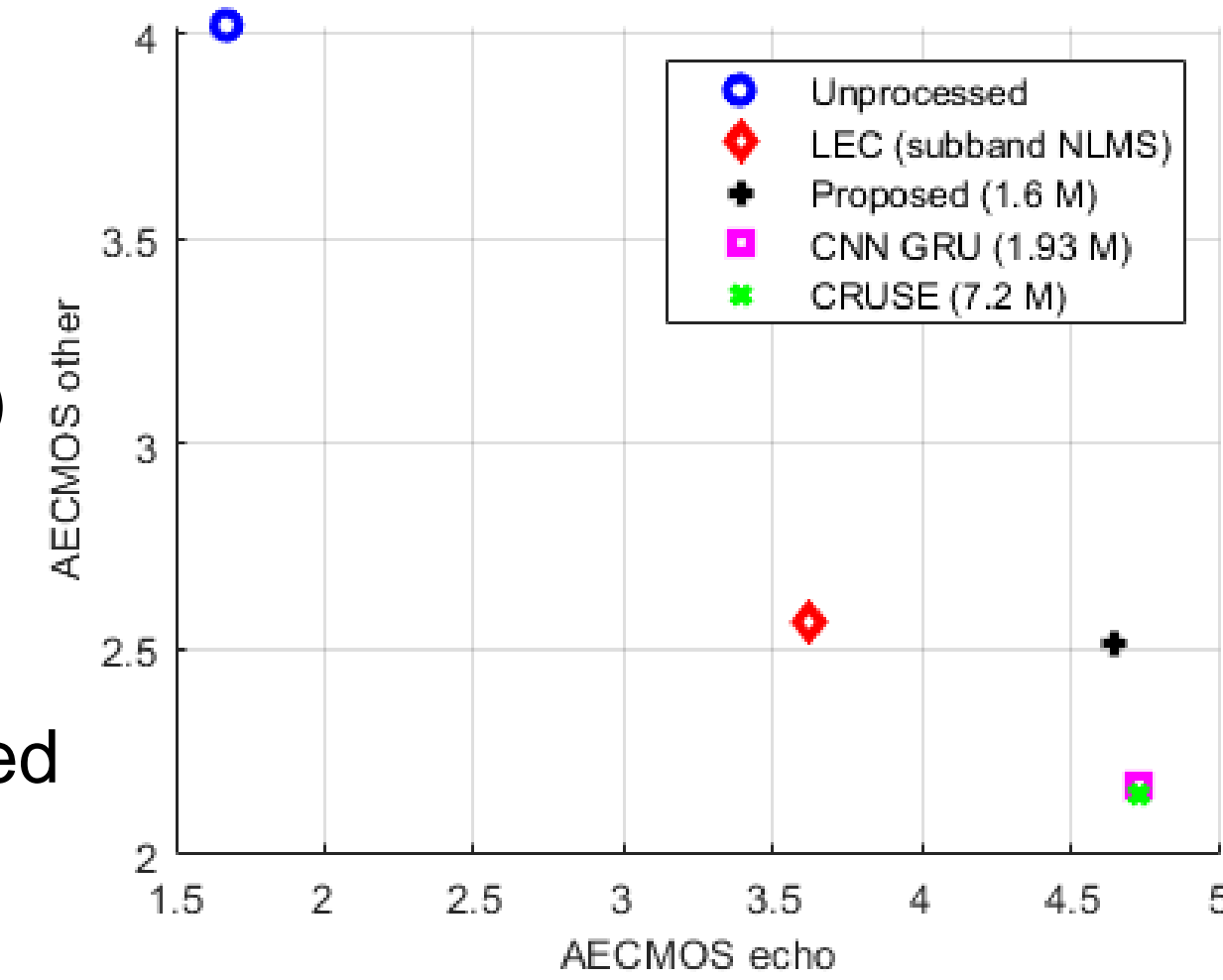
- Dtalk nom./max., circle walk, hand move, rocking, static

AEC-MOS (non-intrusive quality predictor)

- Developed by Microsoft Research
- Echo leak (DT-Echo), NE att. (DT-other)

Methods: LEC, CNN-GRU, CRUSE, Proposed

- Consistently outperformed LEC
- Trade-off echo leak vs. NE attenuation



Performance Evaluation

- Scenario example:** ETSI double talk Rx max @ 87dB
- Plots:** Spectrograms, level vs. time (LVT), RMS EC depth vs. time
- Scenarios:** **DT** Echo leak, NE attenuation [ITU-T-REC-P.502], **STNE** WB-PESQ, **STFE** EC-depth eval, ERLE

