Opus Speech Coding Enhancement: Bandwidth Extension

draft-ietf-mlcodec-opus-speech-coding-enhancement

Jan Buethe (Meta, jbuethe@meta.com), IETF 123

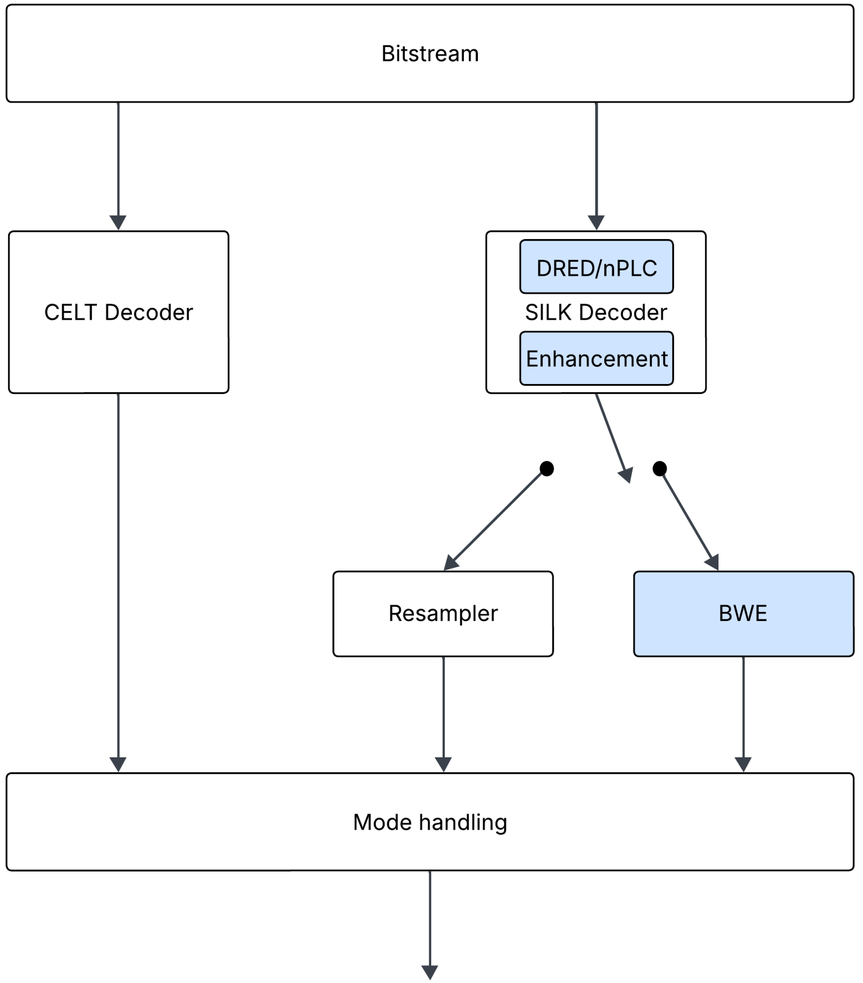
Algorithm Update

BBWENet Integration

- BBWENet: blind (non-guided) wideband to fullband extension algorithm
- Algorithm fully integrated into Opus (main branch, 44f448b2)
- Requires both compile-time (—enable-osce) and run-time (-enable_osce_bwe) activation.
- Once enabled, it is applied if dec_complexity >= 4, output sampling rate = 48 kHz and:
  - mode=MODE_SILK_ONLY, bandwidth=WB, or
  - mode=MODE_HYBRID, no data (PLC/DRED)

Integration with Opus Decoder

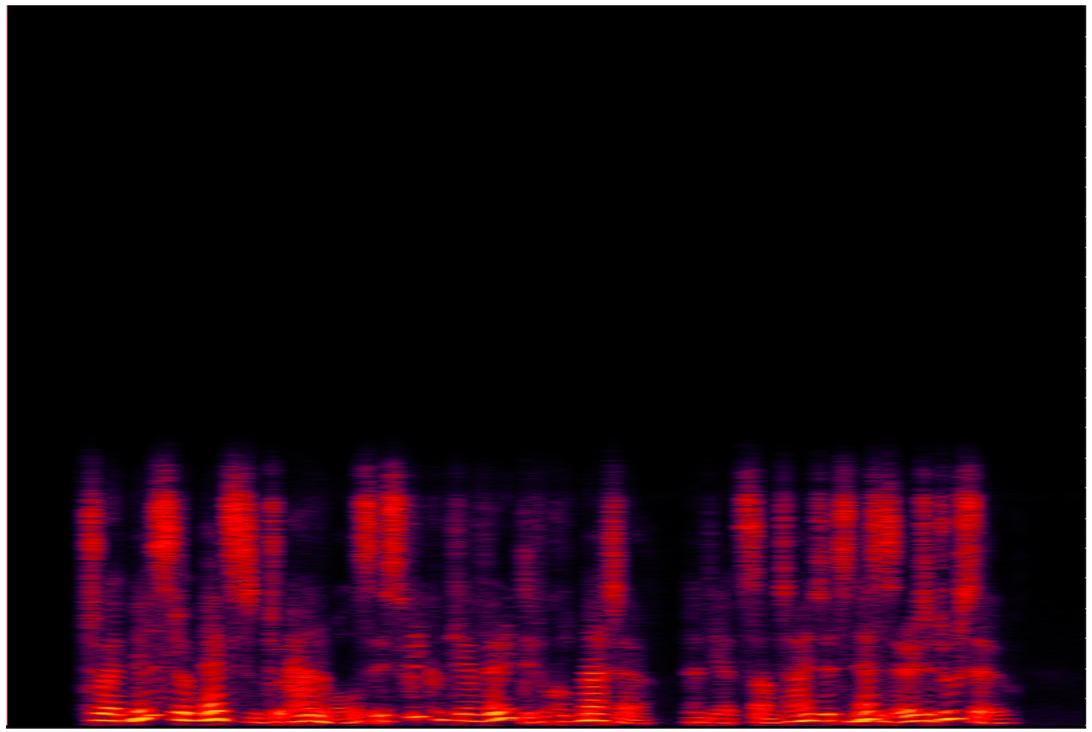
- BWE algorithm, if active, replaces 16->48 kHz resampling
- Adds no delay
- Can work on top of
  - Legacy SILK WB output
  - Enhanced SILK WB output
  - Classic / neural PLC
  - DRED output



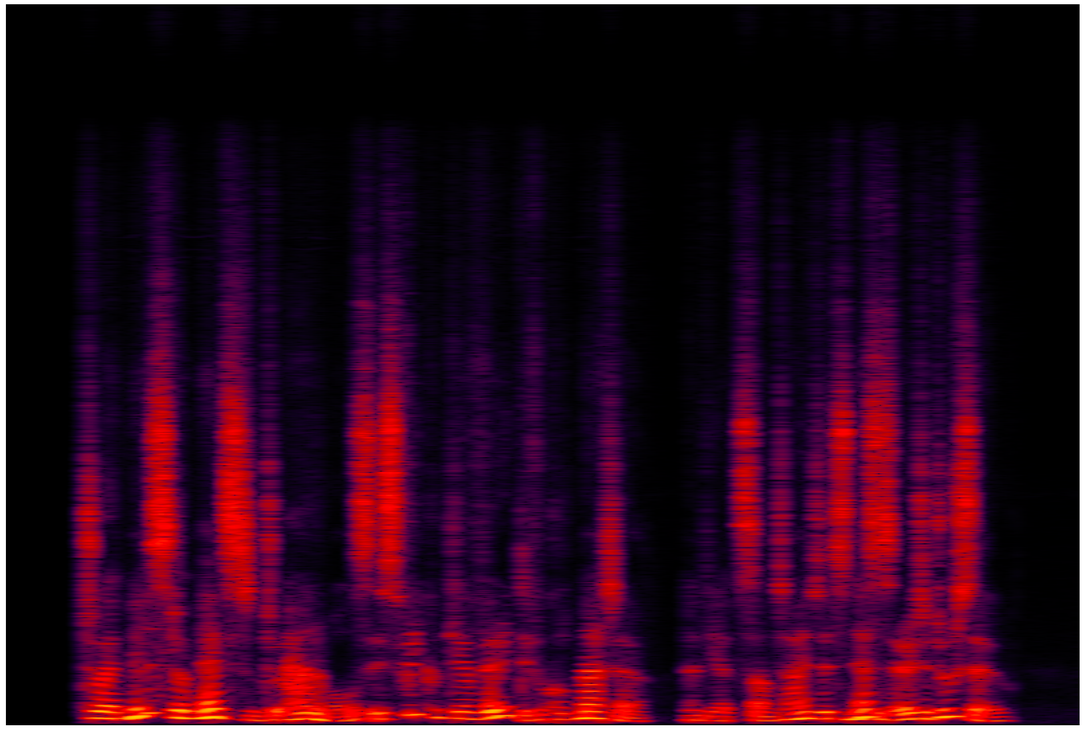
Example 1

Decoding with and without -enable_osce_bwe

Legacy Opus



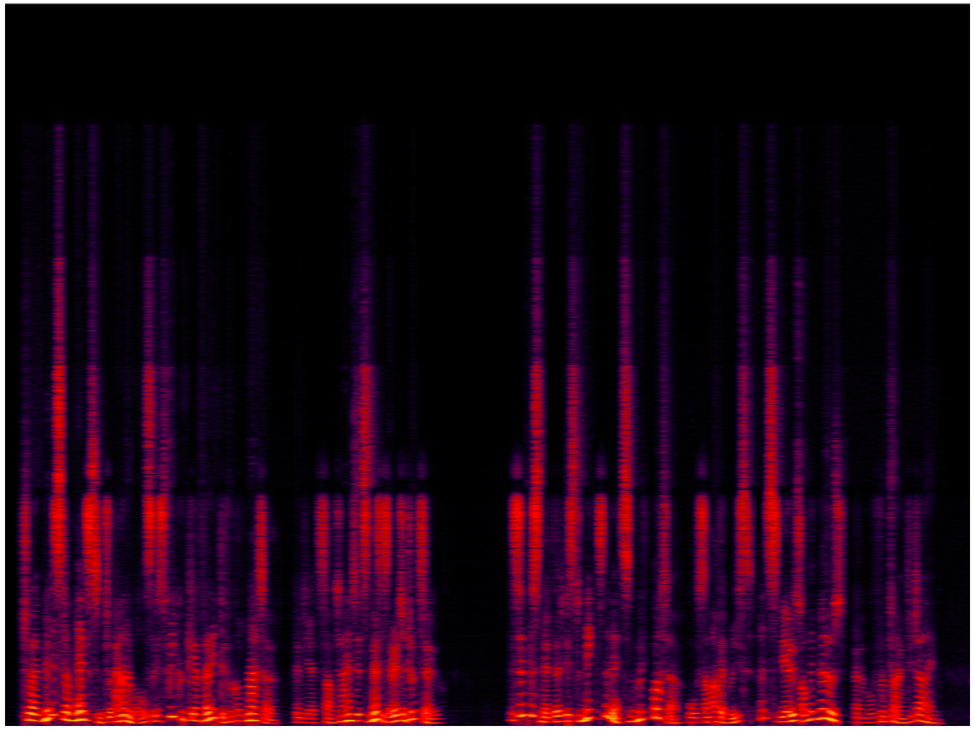
Opus with BWE



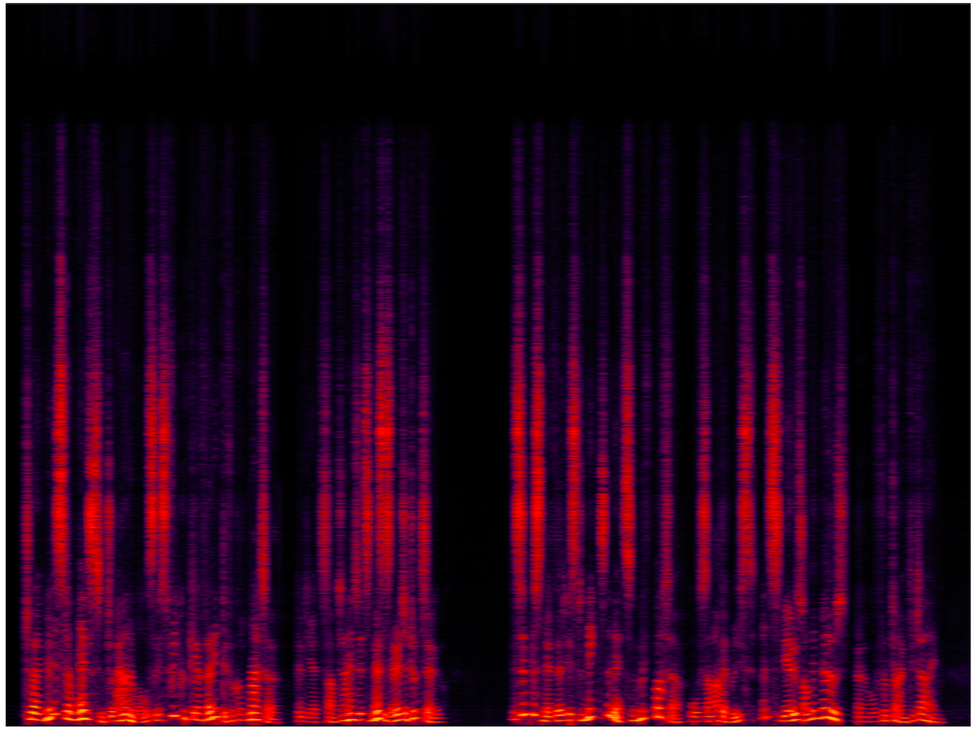
Example 2

Fullband + DRED with and without -enable_osce_bwe (90% loss)

Legacy DRED



DRED with BWE



Draft Update

draft-ietf-mlcodec-opus-speech-coding-enhancement

- Only minor updates to draft text: replaced ‘recognizable reproduction’ to ‘human-recognizable reproduction’ as suggested by Mo
- Draft already distinguishes between extending and non-extending SILK enhancement algorithms
- Requirement for extending SILK enhancement is TBD
- BBWENet integration provides first example of an extending enhancement method

Requirement Proposal for extending SILK Enhancement Algorithms

- Challenge: blind bandwidth extension increases quality but target is not well-defined
- Every extension (e.g. 8 -> 9kHz, 8 ->12 kHz) is an improvement
- => The most we can demand from extension (highband): should not be worse than doing nothing
- Proposal:
  - Use non-extending enhancement requirements for 16 kHz for lowband (run_osce_test.py updated to support 48kHz decoding)
  - Add dedicated highband evaluation implementing a ‘not-worse-than-doing-nothing’ test

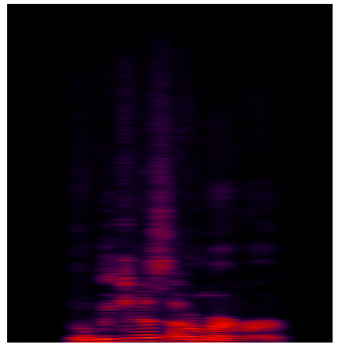
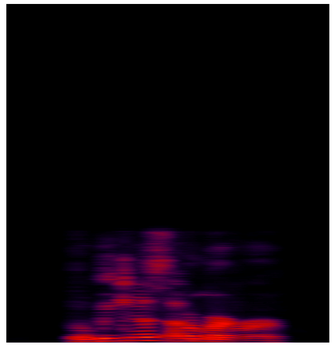
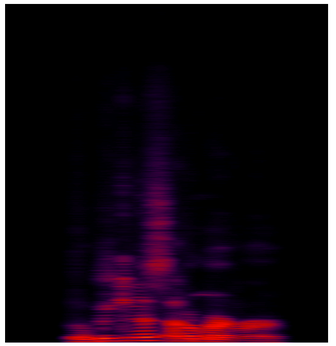
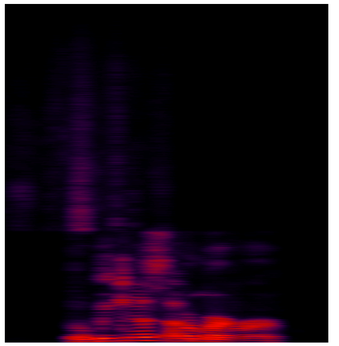
Highband Evaluation

Band-wise distortion

- Assumption:
  - For good BWE algorithm, highband energy will correlate with reference signal
  - For a faulty BWE algorithm, correlation will be weak
- => Distortion measured on power spectra will be
  - lower between good BWE and Reference compared to unextended and Reference
  - higher between faulty BWE and Reference compared to unextended and Reference
- Implementation in https://gitlab.xiph.org/xiph/opus/-/blob/main/dnn/torch/osce/stndrd/evaluation/highband_eval.py on bark-scale subbands

Highband Evaluation Example

Tested on subset of EARS dataset

		Original	8 kHz lowpass	BWE	Temp. Highband Flip
Spectrogram					
Distortion	16.6 - 20 kHz	0.0	0.082	0.07	0.102
	12 - 16.6 kHz	0.0	0.115	0.082	0.134
	9.6 - 12 kHz	0.0	0.144	0.088	0.166
	8.0 - 9.6 kHz	0.0	0.131	0.057	0.165

Expressive **A**nechoic **R**ecordings of **S**peech (EARS): https://github.com/facebookresearch/ears_dataset

Next steps

- Algorithm:
  - Quantize model (requires retraining / fine-tuning)
- Standardization:
  - Collect feedback on highband evaluation
  - Add PLC masking to run_osce_test.py (carry-over)
  - Select fullband test material for highband evaluation (CommonVoice is SWB and EARS is CC-NC)

Thank You!

Backup

BBWENet MOS Test