ITU – Telecommunications Standardization Sector STUDY GROUP 21 Question C/16 **Video Coding Experts Group (VCEG)** 75th Meeting: 2–8 November 2024, Kemer, TR	Document VCEG-BW12

Question: C/16 SG21 (VCEG)

Source: **Christian Helmrich, Heiner Kirchhoffer, Jonathan Pfaff, Sophie Pientka, Christian Rudat, Heiko Schwarz, Detlev Marpe, Thomas Wiegand (Fraunhofer HHI)**

Tel:

Email: firstname.lastname@hhi.fraunhofer.de

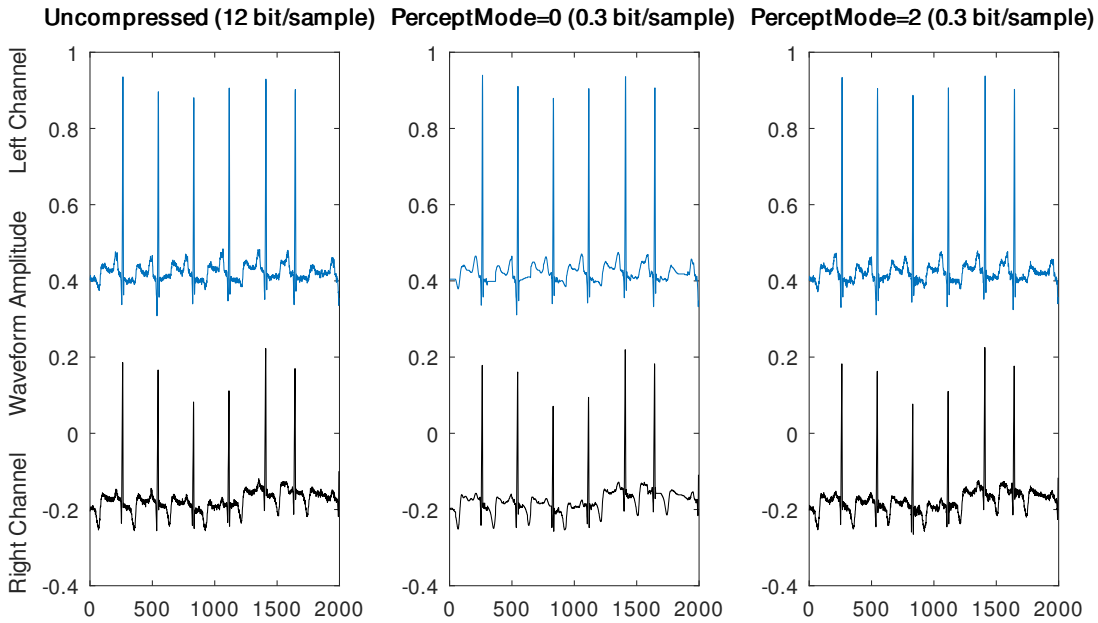
Title: **Optimization of HHI's CfP response for rate constrained or perceptual use cases**

Purpose: Proposal

Abstract

This contribution describes functionality added to Fraunhofer HHI's *objectively* optimized response to the ITU H.BWC Call for Proposals (CfP) on the coding of biomedical waveform data, allowing for lower worst-case bit consumption and higher perceptual quality of the reconstructed waveforms. The resulting *subjectively* optimized configuration has been submitted as a separate CfP response so as to simplify comparisons, but the decoders of both CfP submissions are fully identical; see VCEG-BW02.

The individual codec components related to rate constrained and/or perceptually optimized coding, namely, verbatim residual coefficient coding, guided deblocking pre-/postprocessing, use of a discrete sine transformation (DST), spectrotemporal shaping of the quantization noise, and noise insertion, are introduced, and visualizations are provided to demonstrate their merits. The perceptual configuration is found to be particularly effective in case of low-bitrate coding of single or multi-channel ECG data.



Result of encoding the first seconds of an ECG signal (original: MIT100, left) with Fraunhofer HHI's CfP response and MSE based encoder optimization (PerceptMode = 0, center) or perceptual optimization (PerceptMode = 2, right).

1. Introduction

Fraunhofer HHI's response to the Call for Proposals on biomedical waveform coding [1] consists of two bitstream and decoder submissions: one *objectively* optimized submission, minimizing 'metrical' compression distortions based on frame-wise MSE calculations, such as PRD and CPRD [2], and one s*ubjectively* optimized submission, minimizing perceptual compression distortions when viewing the reconstructed waveforms. Since both submissions employ the same decoding software and, therefore, a fully compatible overall codec architecture and underlying coding/decoding toolsets, this document focuses only on those aspects not discussed in [2], namely, provisions to reduce worst-case data rate of encoded bitstreams (Sec. 2) and perceptually relevant coding artifacts in the reconstructed waveforms (Sec. 3). All technology mentioned herein is well-known and used in several other waveform codecs.

2. Maximum-Rate Constraint

When optimizing waveform codecs for lossy operation, applying (re)quantization of residual block samples or transform coefficients, there may be a risk of the encoded bitstream consuming more rate than the PCM input waveform data when the quantization step-size approaches 1 (i.e. the compression turns lossless). A conventional feature to address this issue, utilized in legacy lossless codecs such as FLAC [3], is the block-wise provision of a residual entropy coding method bypassing variable-length and/or context adaptive extensions. Such 'verbatim' residual coefficient coding, using a fixed number of $N = IBD - ZLSB$ bits for each quantized residual value (where *IBD* is the input bit-depth in bits per sample and *ZLSB* holds the number of zeroed LSBs in the block input), is applied in Fraunhofer HHI's CfP submissions when, for a given block, both the block-wise and sample-wise predictors are **off** and trigonometric transformation is **not** employed. Thus, no flag is needed to signal the 'verbatim' mode.

To assess whether this mode succeeds in limiting the worst-case bit consumption, several waveform recordings with high bits/sample outcomes during lossless coding were studied, and a full-scale white noise PCM signal generated in Octave (at different *IBD* values) was identified as the worst case. When encoding this noise signal at $IBD = 16$ without support for the 'verbatim' mode, the resulting bitstream consumes about 16.51 bits/sample on average. The proposed mode reduces this to 16.**0**11 bits/sample.

3. Perceptual Optimization

Since subjective optimizations tend to reduce the objective performance of waveform codecs, usage of perceptual optimization is signalled, by way of an Intra period-wise 2-byte element, in Fraunhofer HHI's CfP response. In the *objectively* optimized submission, this *if_percept_mode* element equals 0 in every bitstream to indicate that all perceptual tools discussed in the following are bypassed, while in the *subjectively* optimized submission, *if_percept_mode* equals 2 to indicate that all perceptual tools are enabled. Note that an element value of 1 indicates that only deblocking pre/postprocessing is being employed in the Intra period associated therewith and that a value of 3 is reserved and not to be used.

3.1. Guided Deblocking Pre- and Postprocessing

The use of trigonometric transforms without frame overlap (specifically, the type-II discrete cosine transform, DCT-II, in this CfP response) may cause blocking artifacts in the reconstructed waveforms at low bitrates, similar to those which can be observed in e.g. JPEG image coding. To counteract such waveform discontinuity related artifacts, anchor values are being determined, for each channel, on the block and sample prediction residual block signal at the left and right block boundaries before forward transformation of the residual block signal. Using these anchor values, a corrective curve signal is then constructed and subtracted from the residual block signal, with the objective of smoothly tapering the corrected residual block signal towards zero at both boundaries. With a proper choice of the trigonometric transform (see Sec. 3.2), this approach increases the signal compaction in the spectral domain.

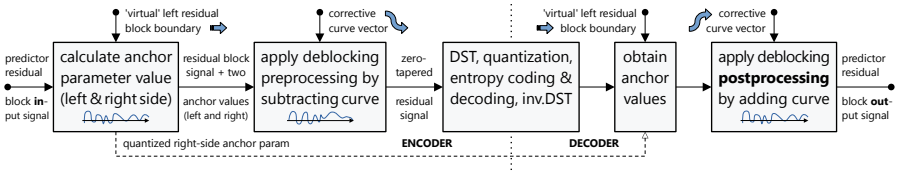


Figure 1*. Schematic illustration of parameter guided deblocking pre-/postprocessing at encoder and decoder.*

At the decoder side, the same corrective curve signal is constructed and added back to the inversely transformed residual block signal before applying the sample and/or block predictors. Figure **1** depicts this approach schematically. In order to make this possible, the right-side anchor value for a block b is quantized, using step-size$_b$ = QP$_b$ to allow for comparatively higher value resolution toward high QPs, and transmitted as side information after the quantized residual transform coefficients for that block, using the same entropy coding method. Since the left-side anchor value can be determined equally on both encoder and decoder side (by taking past reconstructed sample data, from which a 'virtual' past residual signal segment can be calculated as an extension of the current block's residual block signal, see the appendix), it does not need to be signalled in the bitstream. This reduces the deblocking related side information overhead without, according to the authors' observations, perceptual disadvantages.

3.2. Use of Discrete Sine Transform (DST-II)

As noted in Sec. 3.1, for best deblocking performance, a careful choice of the trigonometric transform applied on the corrected residual block signals is important. Support for the type-II discrete sine transform (DST-II) was, therefore, added to Fraunhofer HHI's CfP response, since all basis functions of this transform variant cross the zero-line at both block boundaries. Given the observation that, with some input, use of the DST-II leads to objective performance improvements even when not using the guided deblocking method, support for a block-wise selection between DCT-II and DST-II, on top of a 'transform skip' mode to bypass trigonometric transformation, was included in both the subjectively and objectively optimized submission. For both submissions, the encoder was configured to select the trigonometric transform producing the lower distortion (MSE) of the reconstructed block waveform.

3.3. Spectral and Temporal Noise Shaping

The most fundamental component of perceptual optimization in lossy coding of digital waveforms is the support for encoder control over the spectral and temporal shape of the coding distortion, usually referred to as *quantization noise*. Aside from functionality also used for objective optimization, such as signal adaptive frame partitioning (i.e., block size variation) or step-size selection (QP variation), a support for pre- and postfiltering of the time-domain and/or spectral-domain block signals around the quantizer is integrated into Fraunhofer HHI's CfP response. Spectral noise shaping (SNS) by means of time-domain encoder-side analysis filtering before, and decoder-side synthesis filtering after, quantization of the residual block samples, is a well known approach first applied in the 1970s [4]. Its support in Fraunhofer HHI's CfP response is realized *implicitly* by allowing the sample-wise prediction filter to be operated in an open-loop, instead of closed-loop, structure (which is simply a change of encoder behavior; the decoder does not need to be changed). In this way, the quantization noise can be shaped according to the prediction filter's transfer function. Temporal noise shaping (TNS), representing the frequency-domain counterpart (or duality) to SNS, pursues a similar approach but on the DCT or DST transformed residual block samples (i.e. transform coefficients), by conducting a similar encoder-side analysis filtering before, and corresponding decoder-side synthesis filtering after, the quantization of the transform coefficients [5]. In this way, the quantization noise can be shaped temporally according to the TNS filter's transfer function, *within* a block. Cascaded execution of the SNS predictor (time-domain outer filter) and TNS 'predictor' (frequency-domain inner filter around the quantizer) enables support for spectrotemporal quantization noise shaping in a straightforward and well-known fashion.

Unlike the implicit SNS support in HHI's CfP response, support for TNS is signalled *explicitly*, by transmitting for each DCT or DST coded block, when *if_percept_mode* > 1, a 1-bit flag indicating that either TNS filter (value 1) or noise insertion (value 0) data are present in the bitstream for that block. When the presence of TNS data is indicated, the filter range (fractions of the spectrum), order (2 or 4), and weights (quantized ParCor coefficients) are written to, and read from, the bitstream. TNS is being enabled in a transform coded block when the TNS filter's prediction gain exceeds a given threshold, and the TNS filter coefficients are quantized and entropy coded the same way as the LPC coefficients. Otherwise, noise insertion parameters are determined and transmitted, as outlined in the next section, and (as in case of blocks not subjected to trigonometric transformation) no TNS data are transmitted.

3.4. Insertion of Pseudorandom Noise Values

At low bitrates (i.e., high QPs), transform-domain quantization causes the decoded spectral block residuals to become very sparse, thus leading to so-called 'spectral holes' and a pronounced denoising effect upon waveform reconstruction. To ameliorate such frequently visible artifacts, a noise insertion (NI) process is integrated into Fraunhofer HHI's perceptually optimized CfP response, enabled in all transform coded blocks when *if_percept_mode* > 1 and TNS filtering is not used (flag value 0). It was found that NI is unnecessary, or even undesirable, in the vast majority of blocks where TNS in being used, hence the mutual exclusion of TNS and NI processing. NI acts like a dithering method, but only on those residual transform coefficients quantized to zero. To avoid degrading prediction performance of the overall codec architecture, the inserted, scaled pseudorandom noise values are inversely transformed separately and added only to the reconstruction output, not to the prediction buffer memory, as in, e.g., film grain synthesis in video coding [6]. In addition to a noise scaling factor (or level), a filter order (also 2 or 4) and weights (also quantized ParCor coefficients) are transmitted, as for TNS, with delta-time coding of the weights allowed. The NI filter is used for spectral shaping of the noise data.

4. Illustration of Merits of Perceptual Tools

Figures **2** – **4** illustrate the deblocking and spectro-temporal noise shaping capability of Fraunhofer HHI's CfP response, while figure **5** depicts the effect of noise insertion on the spectrogram of an ECG signal. Note how TNS acts as a *deringing* filter around signal spikes, complementing the deblocking.

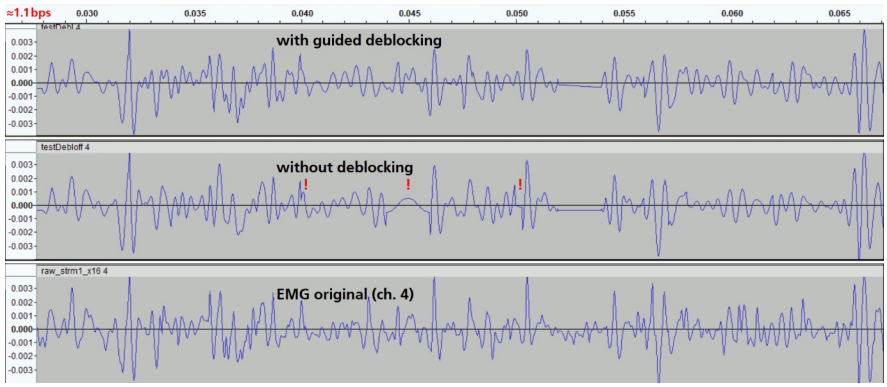


***Figure 2**. Effect of deblocking + DST-II on an EMG signal encoded with HHI's perceptually optimized response.*

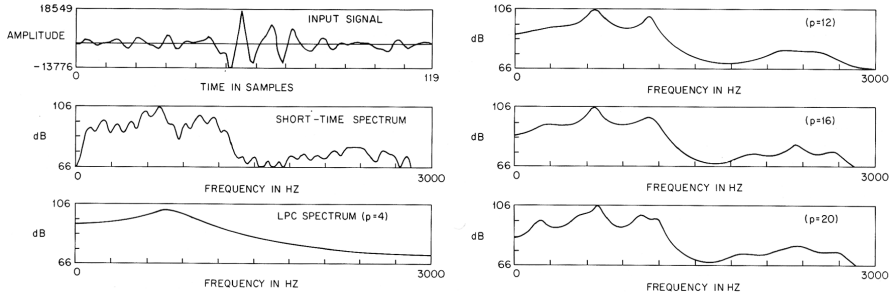


***Figure 3**. Effect of spectral noise shaping pre-/postfiltering on coding error spectrum of waveform signal [4].*

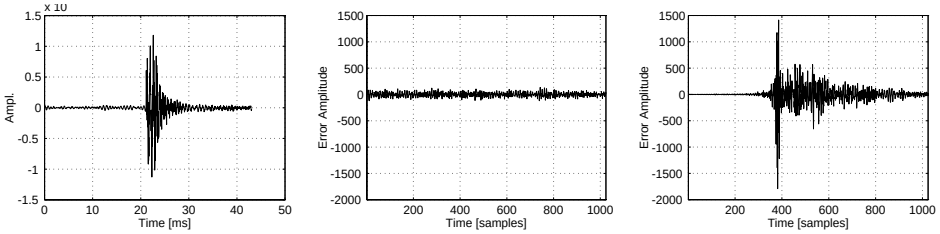


***Figure 4**. Effect of temporal noise shaping pre-/postfiltering on coding error of time-domain PCM signal [5].*

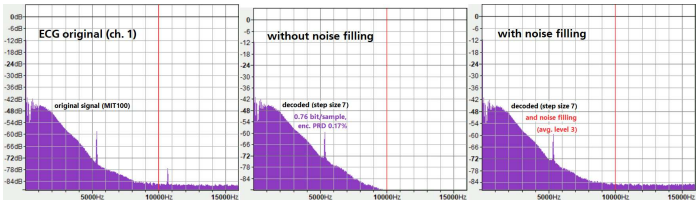


***Figure 5**. Effect of decoder-side noise insertion on a time-average short-term spectrogram of an ECG signal.*

5. Experimental Results

Since an objective evaluation of a subjectively optimized codec, e.g. by means of PRD, CPRD, or a form of PSNR assessment, is only barely meaningful (especially when perceptually motivated noise components are added to the decoded waveform signals), it is only noted here that, when deactivating the above-described noise insertion process on the decoder side, Fraunhofer HHI's perceptually optimized CfP submission is, in terms of measurable signal distortions, only slightly less efficient than its MSE optimized counterpart. However, detailed visualizations for the datasets listed in the H.BWC CfP document [1] may be provided upon request and/or as supplementary material at the VCEG meeting.

6. Patent Rights Declaration

Fraunhofer HHI may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).

7. References

[1] VCEG, "Call for Proposals for a new ITU-T Recommendation H.BWC on the coding of biomedical waveform data," *VCEG document SG16-TD286-R1*, Apr. 2024.

[2] J. Pfaff, H. Schwarz, *et al.*, "Fraunhofer HHI's Response to the Call for Proposals on the Coding of biomedical waveform data," *VCEG document BW02 v1*, Nov. 2024.

[3] J. Coalson *et al.*, "Free Lossless Audio Codec," FLAC version 0.5, Jan. 2001. Available online: https://xiph.org/flac/changelog.html (format version history).

[4] B. S. Atal and M. R. Schroeder, "Predictive Coding of Speech Signals and Subjective Error Criteria," in *Proc. IEEE ICASSP*, Tulsa, OK, USA, Apr. 1978.

[5] J. Herre and J. D. Johnston, "Enhancing The Performance of Perceptual Audio Coders By Using Temporal Noise Shaping," in *Proc. AES 101st Conv.*, Los Angeles, Nov. 1996. Available online: https://www.ee.columbia.edu/~dpwe/papers/HerreJ96-noisesh.pdf.

[6] D. Grois, Y. He, *et al.*, "Film grain synthesis technology for video applications," *JVET document AG2020*, 2024. Online: https://jvet-experts.org/doc_end_user/current_document.php?id=13913.

A. Appendix

The following pages contain more detailed information on the decoding process of all perceptual components of Fraunhofer HHI's CfP response introduced earlier in this document, as well as visual comparisons of the three configurations of Fraunhofer HHI's response for all objective metrics in [1]

A.1. Guided Deblocking Pre- and Postprocessing

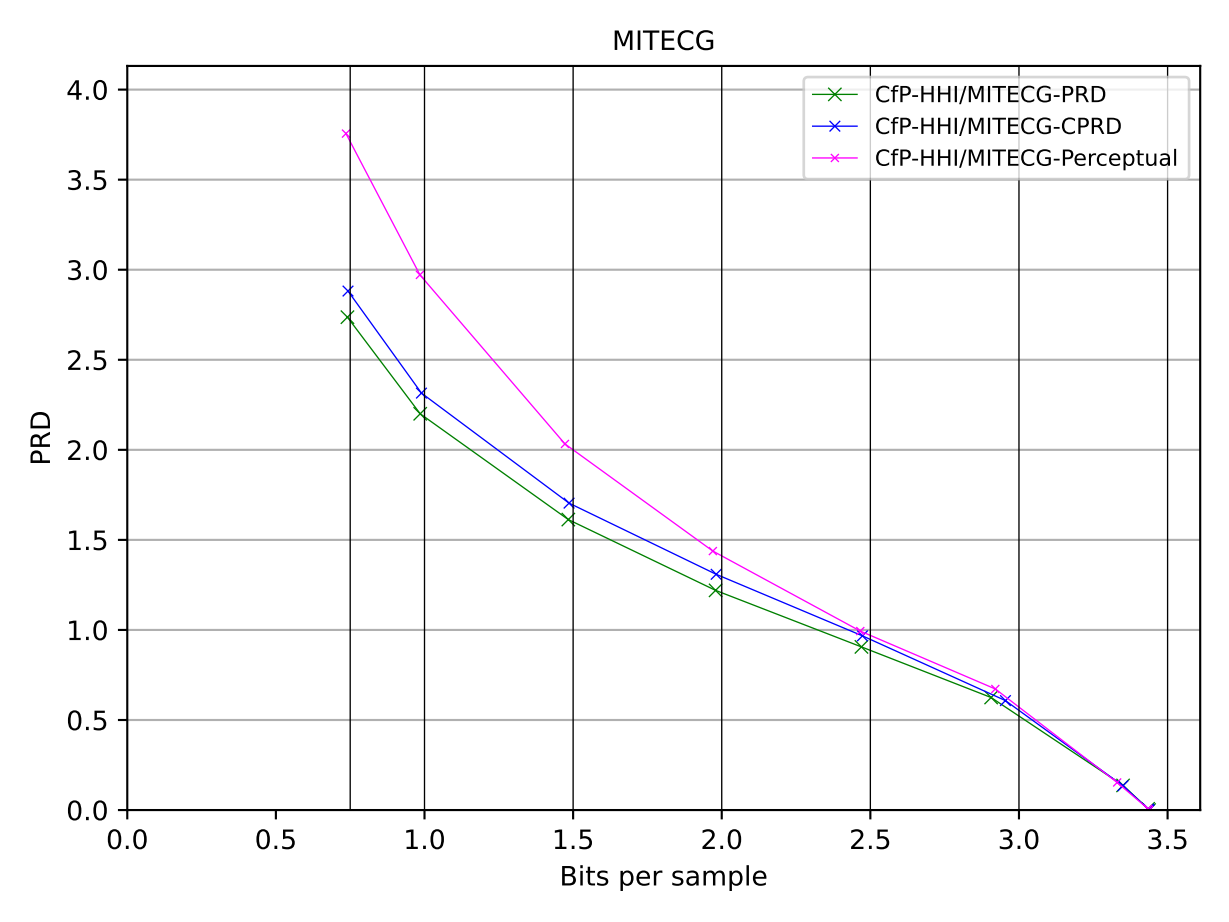
to be provided in a later version of this document

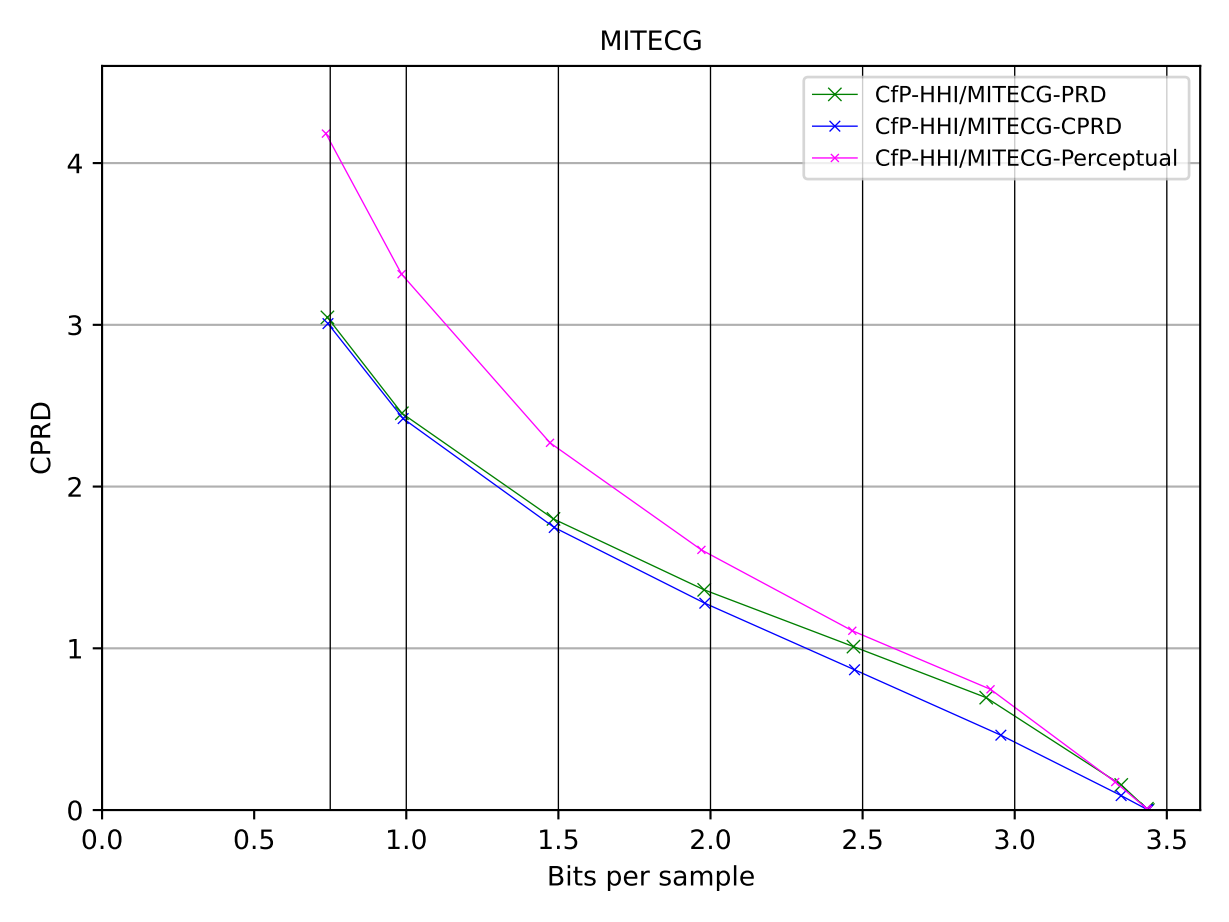
A.2. Spectral and Temporal Noise Shaping

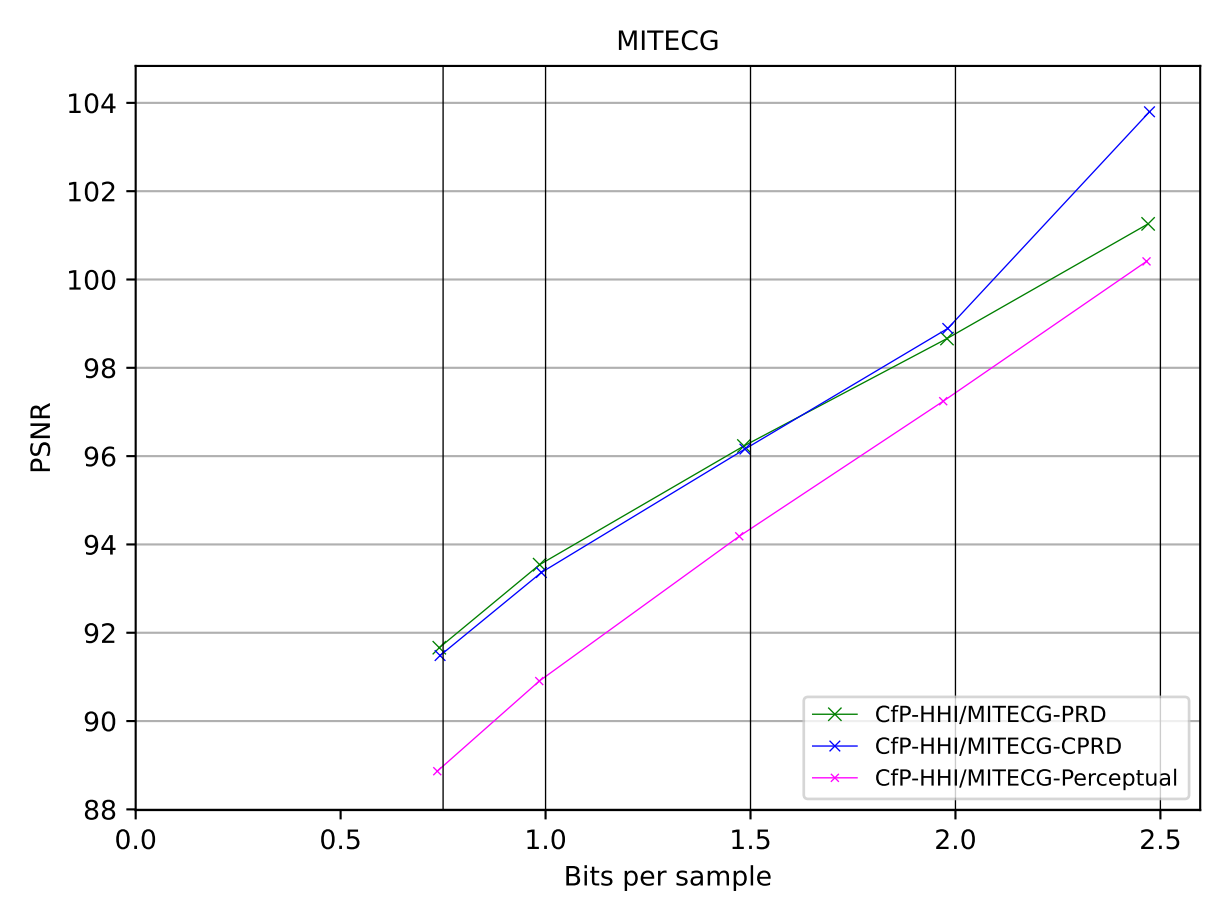
Spectral noise shaping (SNS), as mentioned earlier, is realized by invoking a *sample*-wise linear predictive filtering process on the output of a transform-domain (DCT-II or DST-II) coded residual block signal, prior to invoking the *block*-wise prediction and reconstruction process. To this end, up to 16 boundary "virtual residual" samples are determined at the left side of the current block, directly before the start of the current block and using the respective reconstructed samples for that channel and an extension (towards the past) of the block-wise prediction signal for the current block. These "virtual residual" samples are then utilized as initialization data for the sample-wise linear predictive filter, such that the filtering can start at the first sample in the current block, with past data available.

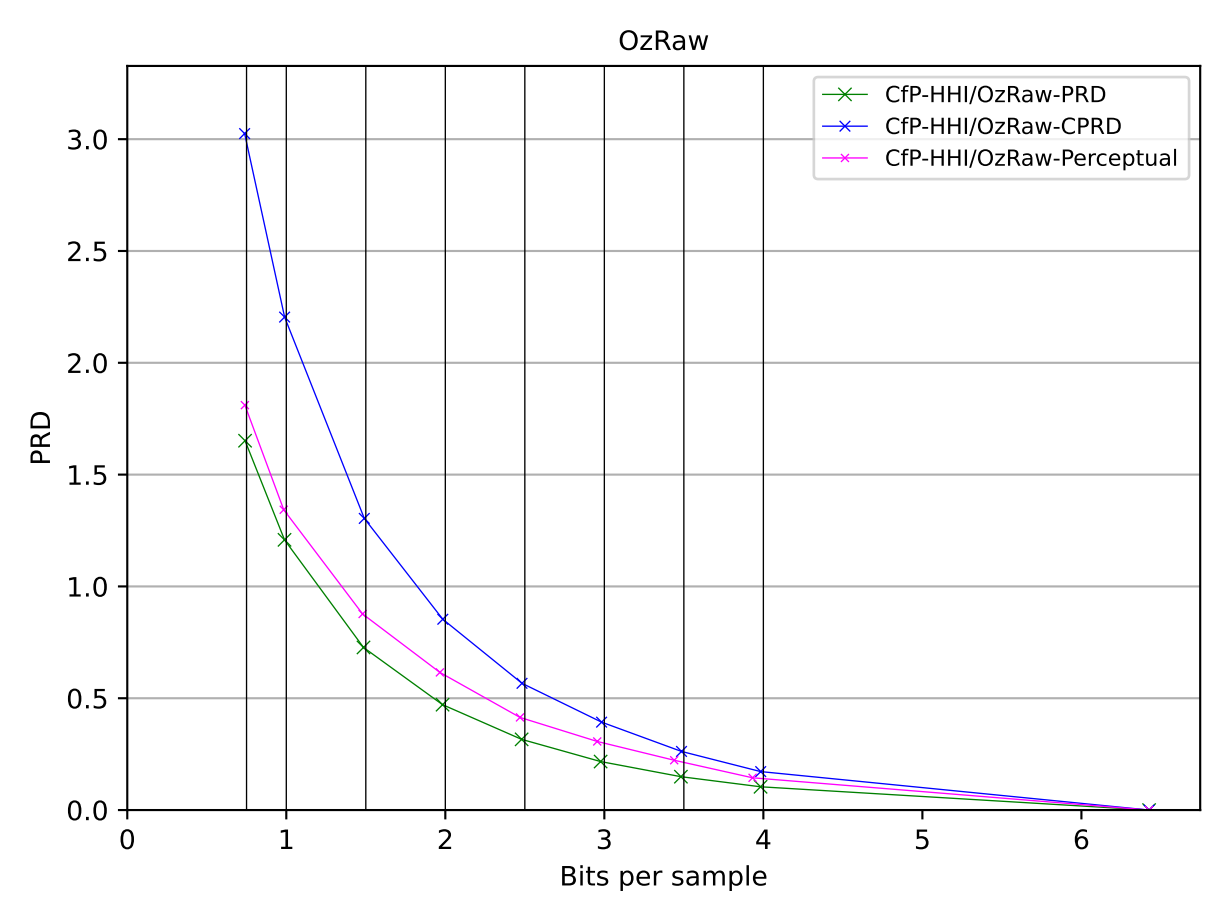
Temporal noise shaping (TNS) is realized almost identically as in prior codecs applying this tool. When, for a block, the single-bit noise parameter flag indicates usage of TNS, an additional three bit are read from the bitstream, indicating the spectral range (in $1/9^{th}$ of the full range) to which to apply frequency-domain linear predictive filtering identical to that for SNS. The spectral start index of the TNS filtering process equals $1/8^{th}$ of the spectral length, or the filter order if that is larger (to have all prior data), and the filtering direction is from low to high frequencies. Unlike in most prior codecs, the filtering process is integer-only, like the time-domain sample-wise linear predictive filtering discussed above (i.e., using only integer multiplies-accumulations, additions, and binary right-shifting).

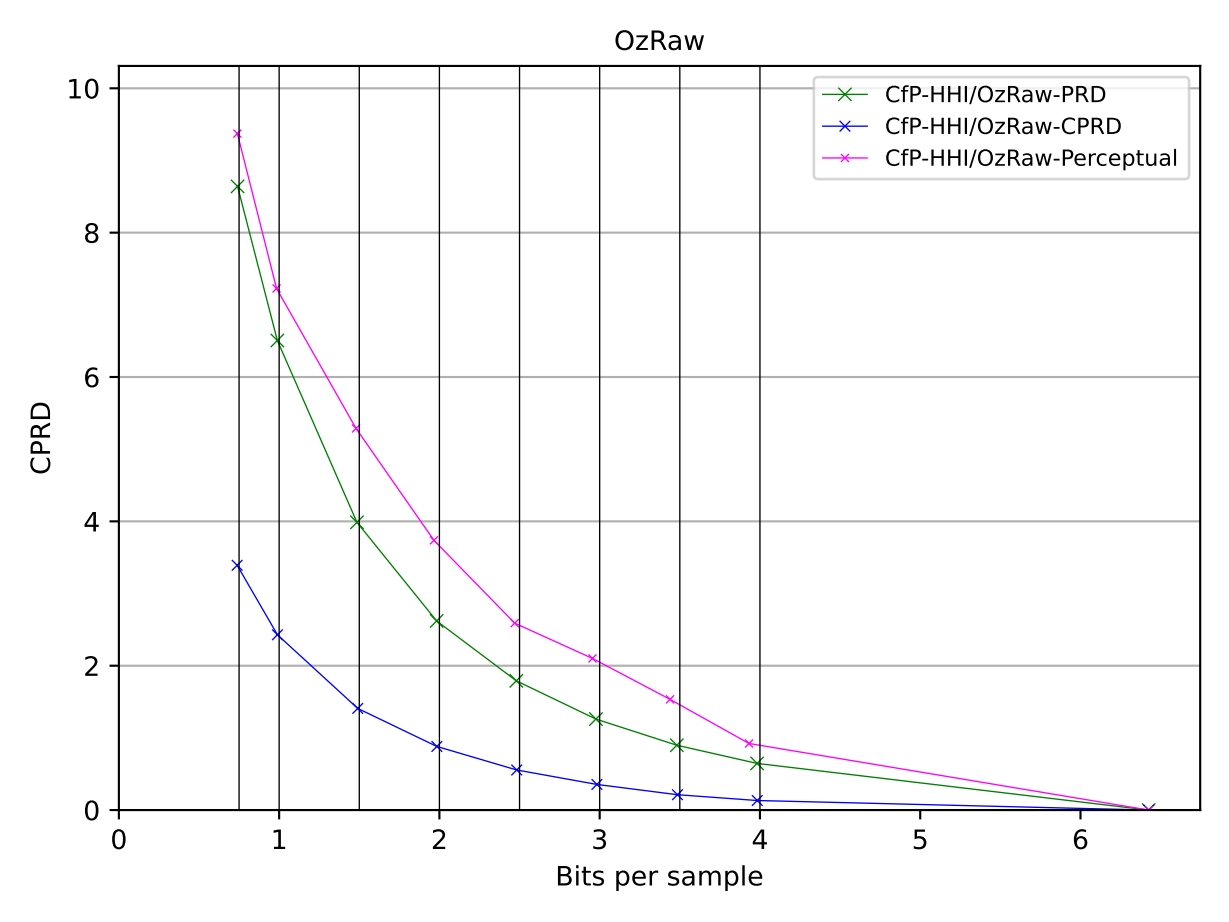
The TNS filter coefficients are quantized and entropy coded identically to the filter coefficients in use for the time-domain sample-wise predictive filtering (companded ParCor domain), with the only exception being that quantized TNS coefficients are clipped, after entropy decoding, to value range $-1 < \text{value} < 1$, so as to guarantee filter stability (the sample-wise predictive filter allows ± 1 values).

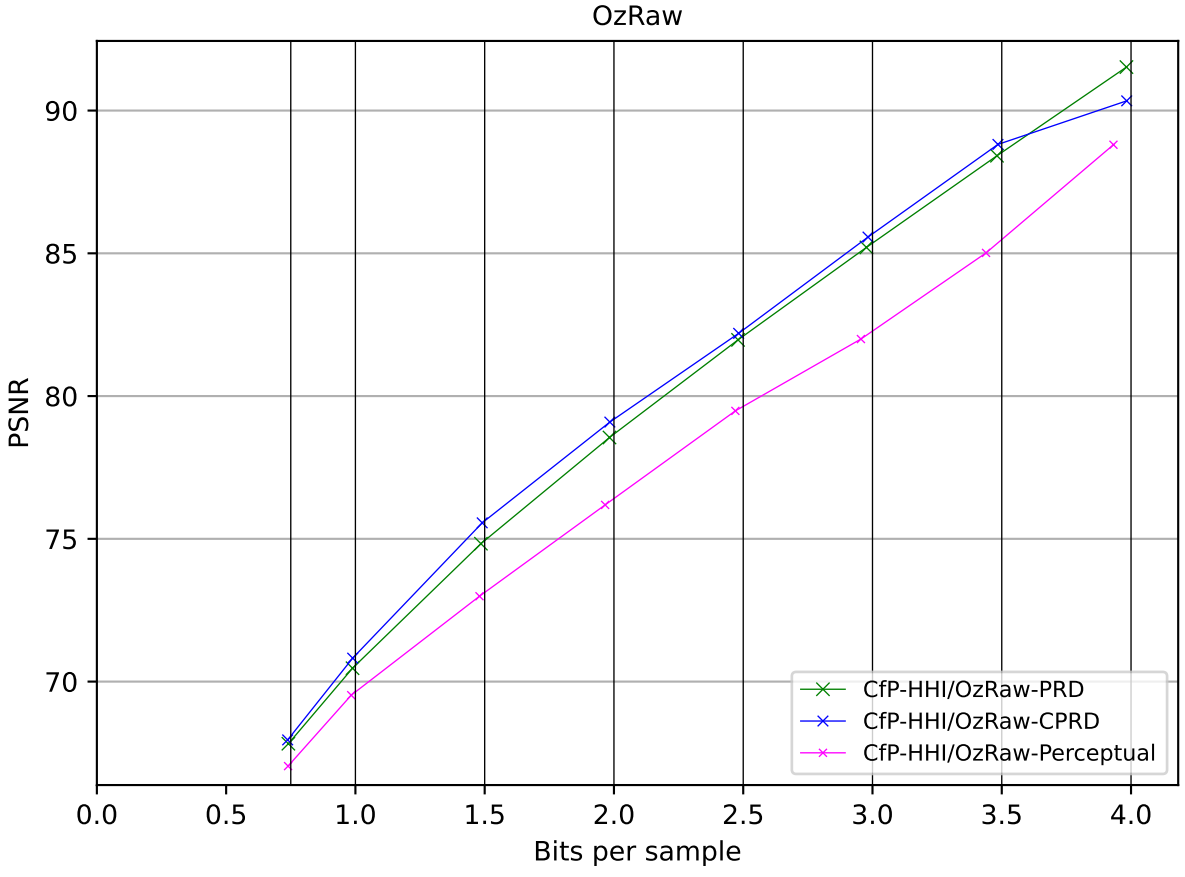

MITECG
CfP-HHI/MITECG-PRD
CfP-HHI/MITECG-CPRD
CfP-HHI/MITECG-Perceptual
PRD
Bits per sample
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

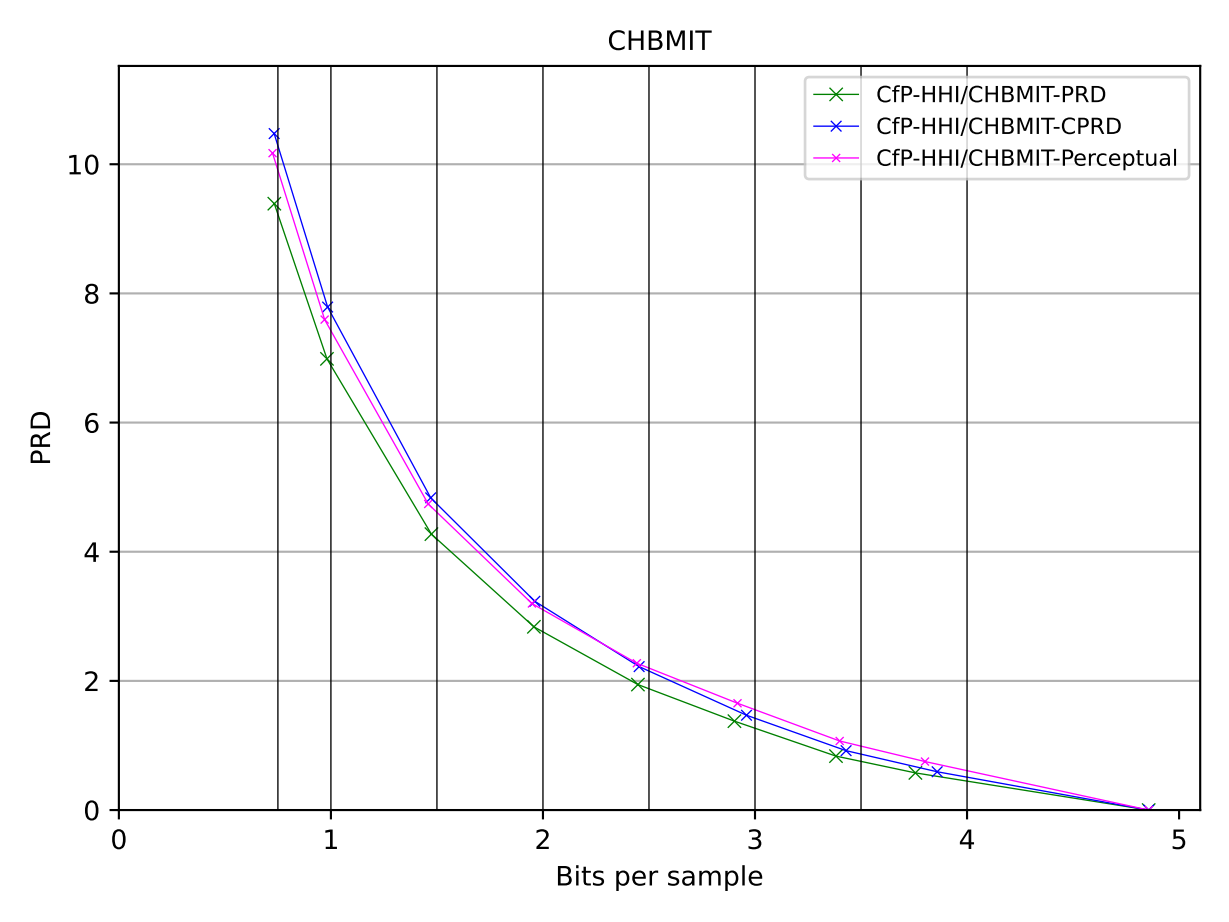

MITECG
CfP-HHI/MITECG-PRD
CfP-HHI/MITECG-CPRD
CfP-HHI/MITECG-Perceptual
CPRD
Bits per sample
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

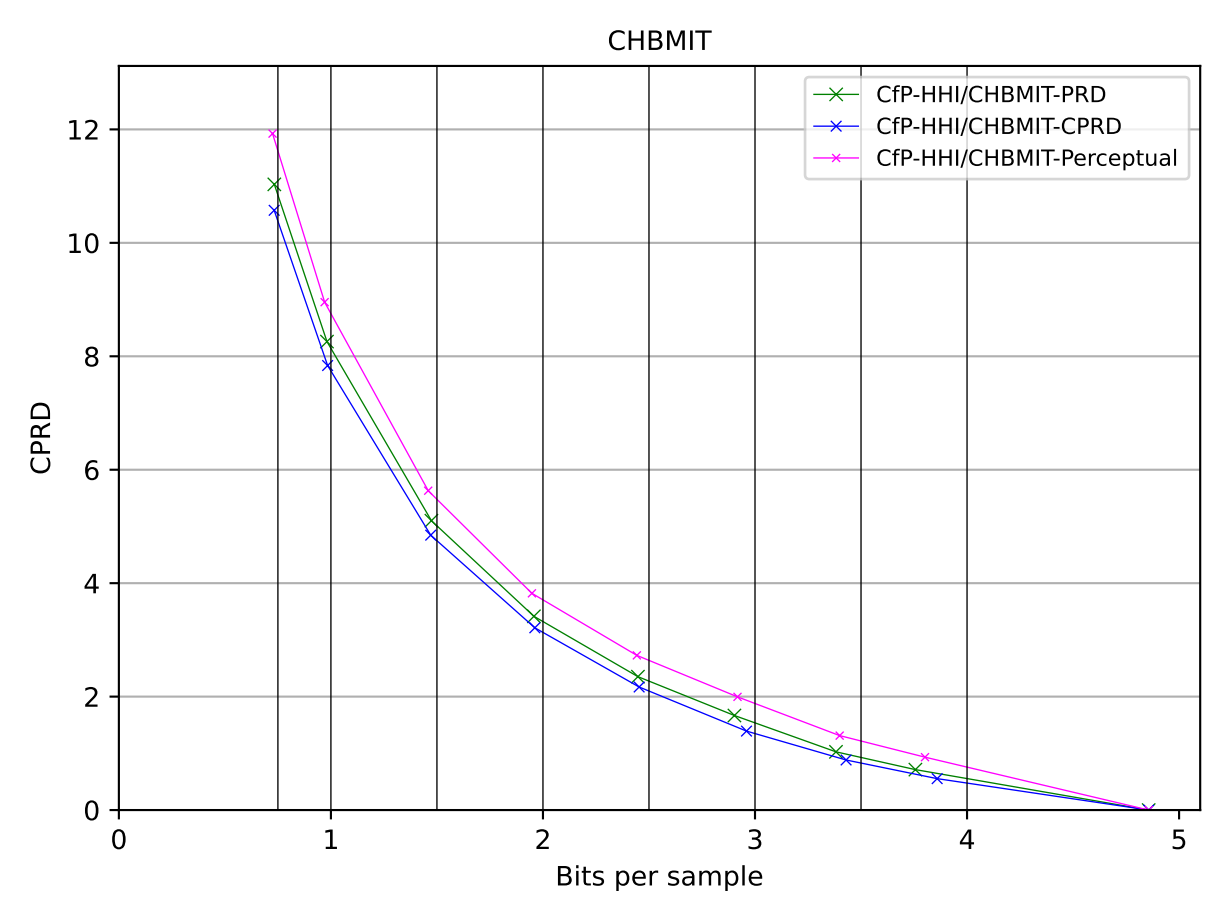

MITECG
PSNR
Bits per sample
104
102
100
98
96
94
92
90
88
0.0
0.5
1.0
1.5
2.0
2.5
CfP-HHI/MITECG-PRD
CfP-HHI/MITECG-CPRD
CfP-HHI/MITECG-Perceptual

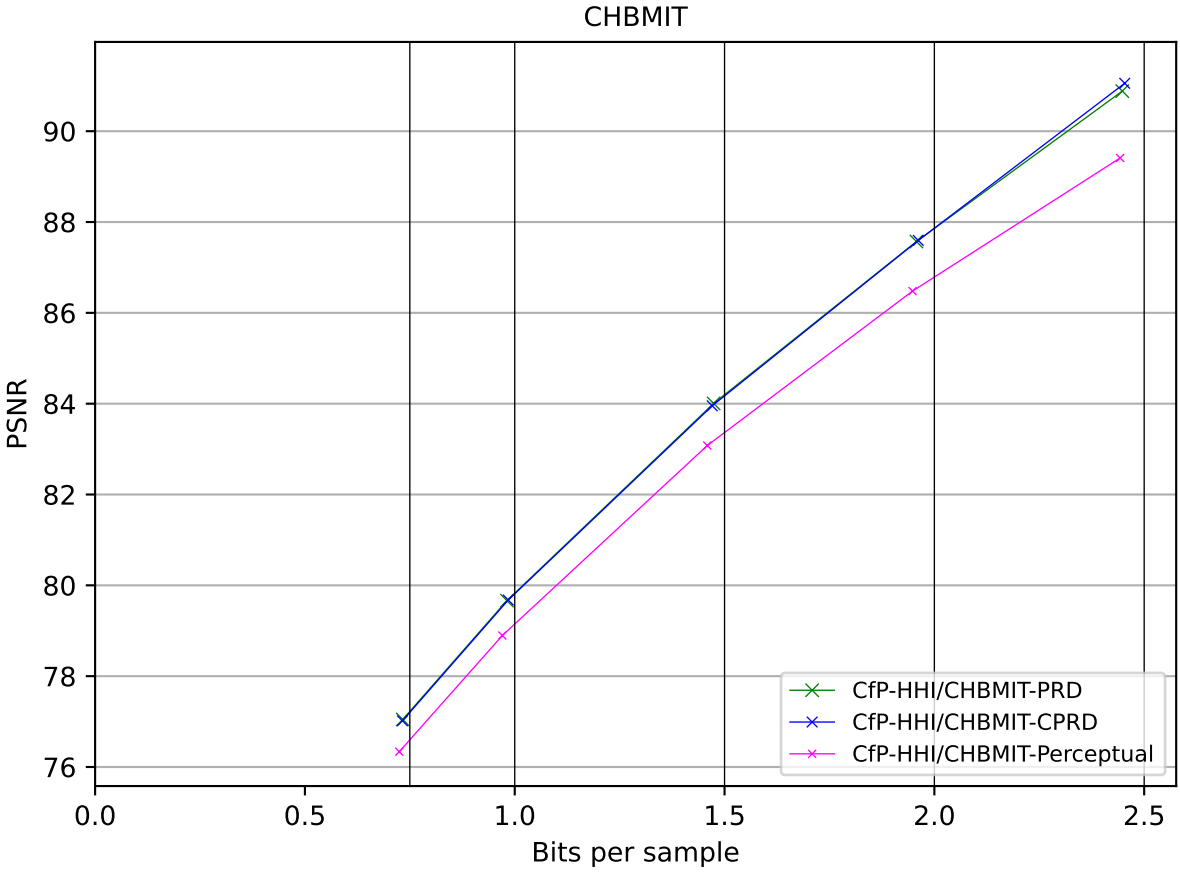

OzRaw
CfP-HHI/OzRaw-PRD
CfP-HHI/OzRaw-CPRD
CfP-HHI/OzRaw-Perceptual
PRD
Bits per sample
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
6


OzRaw
CfP-HHI/OzRaw-PRD
CfP-HHI/OzRaw-CPRD
CfP-HHI/OzRaw-Perceptual
CPRD
Bits per sample
0
2
4
6
8
10
0
1
2
3
4
5
6


OzRaw
PSNR
Bits per sample
90
85
80
75
70
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
CfP-HHI/OzRaw-PRD
CfP-HHI/OzRaw-CPRD
CfP-HHI/OzRaw-Perceptual


CHBMIT
CfP-HHI/CHBMIT-PRD
CfP-HHI/CHBMIT-CPRD
CfP-HHI/CHBMIT-Perceptual
PRD
Bits per sample
0
2
4
6
8
10
0
1
2
3
4
5


CHBMIT
CfP-HHI/CHBMIT-PRD
CfP-HHI/CHBMIT-CPRD
CfP-HHI/CHBMIT-Perceptual
CPRD
Bits per sample
0
2
4
6
8
10
12
0
1
2
3
4
5


CHBMIT
PSNR
Bits per sample
90
88
86
84
82
80
78
76
0.0
0.5
1.0
1.5
2.0
2.5
CfP-HHI/CHBMIT-PRD
CfP-HHI/CHBMIT-CPRD
CfP-HHI/CHBMIT-Perceptual