Description of Fraunhofer HHI's response to the Call for Evidence on the compression of biomedical waveform data
Document VCEG - BU03

Jonathan Pfaff, Heiner Kirchhoffer, Christian Rudat, Sophie Pientka, Christian Helmrich, Heiko Schwarz, Detlev Marpe, Thomas Wiegand

Fraunhofer Heinrich Hertz Institute, Berlin

Introduction

Liaison statement from DICOM-WG32:

- Sent to Question 6 of ITU-T Study Group 16 (VCEG).
- Points out absence of well accepted codec for compression of biomedical waveform data.
- Asks for assessment of existing technology and possible development of a new codec.
- DICOM-WG32-LS20221107, SG16-TD103/Ge.

Identification of a benchmark set: (VCEG-BT05, VCEG-BU01):

- Extended HE-AAC as state of the art audio codec.
- Design of an MSE/PRD-optimized encoder for Extended HE-AAC.
- Discussion and agreement on methodology with MPEG Audio coding group.

Call for Evidence (CfE)

Document VCEG-BT07 of Hannover meeting:

- Asks for compression methods for biomedical waveform data.
- Goal: Find out if technology with significantly better compression than benchmark codec exists.

Datasets to compress provided by DICOM for three scenarios:

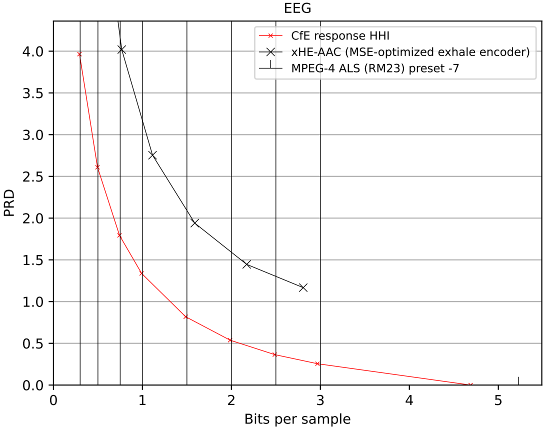
- Electroencephalography (EEG)
- Electromyography (EMG)
- Electrocardiography (ECG)

Reporting of results:

- Distortion measure:
  - MSE-based.
  - Scaled version of square-root of sum of squared errors (PRD).
- Curves for rate (bits per sample, BPS) versus PRD.

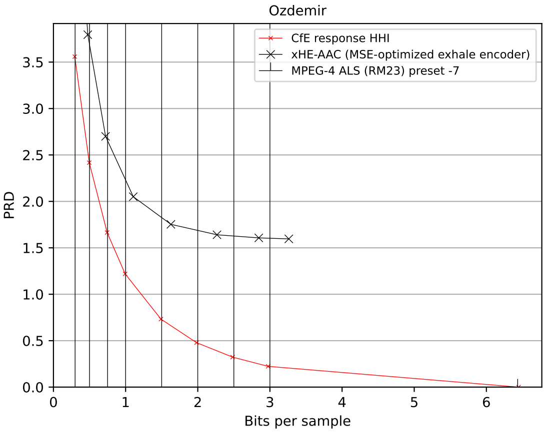
Experimental results of CfE response

Results for EEG Data



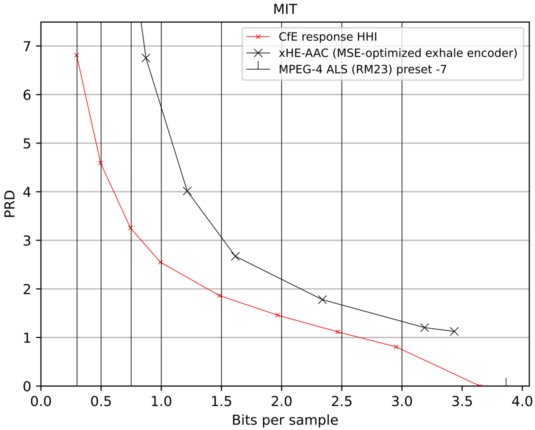
BPS versus PRD curve for the EEG dataset

Results for EMG Data



BPS versus PRD curve for the EMG dataset

Results for ECG Data



BPS versus PRD curve for the ECG dataset

Technical description

Overview of structure

Block-based, hybrid architecture:

- Partitioning into blocks.
- Prediction generation per block.
- Transform of prediction residuals.
- Quantization of tranform coefficients.
- Entropy coding of transform coefficient levels and side information.

Partitioning and processing order

Partitioning

Biomedical waveform signals:

- Comprised of M channels, and N samples per channel.
- 16-bit input sample values $x[i][j]$, $0 \leq i \leq M-1$, $0 \leq j \leq N-1$.

Partitioning:

- Partitioning into sequence of B blocks $b_0, \ldots, b_{B-1}$.
- Block b_k of length l_k and starting position s_k:

$$b_k = \{x[i][j] \colon 0 \leq i \leq M-1 \colon s_k \leq j < s_k + l_k\}.$$

- Length l_k integral power of 2.
- Consecutive blocks: $s_0 = 0$ and $s_{k+1} = s_k + l_k$.
- Partitioning of b_k into channel-wise subblocks:

$$b_{k,m} = \{x[m][j] \colon s_k \leq j < s_k + l_k\}.$$

Processing order

Sequential coding $\mathbf{b}_0 \to \mathbf{b}_1 \to \cdots \to \mathbf{b}_{B-1}$

- Start with b_0.
- Until $k = B - 1$: Code b_{k+1} after having coded b_k.

For each $\mathbf{b}_k$: Sequential coding $\mathbf{b}_{k,0} \to \mathbf{b}_{k,1} \to \cdots \to \mathbf{b}_{k,M-1}$

- Start with $b_{k,0}$.
- Until $m = M - 1$: Code $b_{k,m+1}$ after having coded $b_{k,m}$.

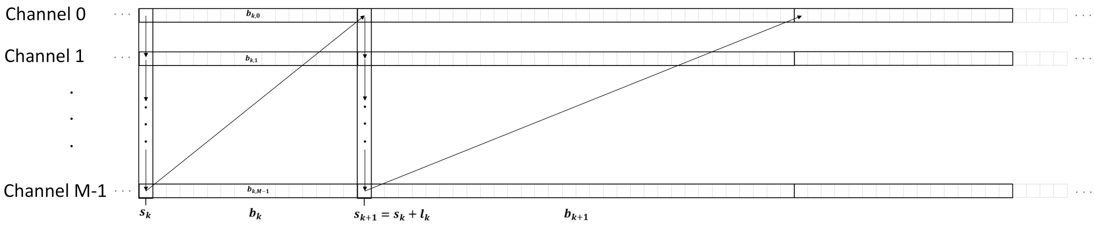


Illustration of partitioning and processing order

Prediction

Prediction

Prediction signal per block $b_{k,m}$

- Five different prediction modes.
- Zero prediction also supported.
- Prediction mode transmitted.

Prediction from samples of same channel:

- Input are reconstructed samples

$$\{y[m][p]\colon p < s_k\}.$$

- DC-, Half-Slope-, Quarter-Slope-Prediction
- Block-Copy Prediction

Prediction from samples of previous channels:

- Input are reconstructed samples

$$\{y[r][p]\colon r < m\colon p < s_k + l_k\}.$$

- Inter-Channel prediction

DC-, Half-Slope-, Quarter-Slope-Prediction

DC-prediction:

- Mean value on four preceding samples:

$$\text{dcVal} = \left(\sum_{p=0}^{3} y[m][s_k - 4 + p] + 2\right) >> 2, \quad \text{pred}[j] = \text{dcVal}, \quad 0 \le j < l_k.$$

Half- and Quarter-Slope-Prediction

- Straight line from preceding reconstructed sample with slope μ:

$$\text{pred}[j] = y[m][s_k - 1] + \mu \cdot (j + 1), \quad 0 \le j < l_k.$$

- Slope determined on two adjacent reconstructed samples:

$$\mu = \begin{cases} (y[m][s_k - 1] - y[m][s_k - 2] + 1) >> 1 & \text{for Half-Slope Prediction} \\ (y[m][s_k - 1] - y[m][s_k - 2] + 2) >> 2 & \text{for Quarter-Slope Prediction} \end{cases}.$$

Inter-Channel Prediction

- Linear prediction from collocated reconstructed samples of channel m_{ref}:

$$pred[j] = (\alpha \cdot y[m_{ref}][s_k + j] + \beta + ro) >> w, : 0 \leq j < l_k.$$

- Channel index $m_{ref} < m$ transmitted.
- Model parameters α and β derived at the decoder.

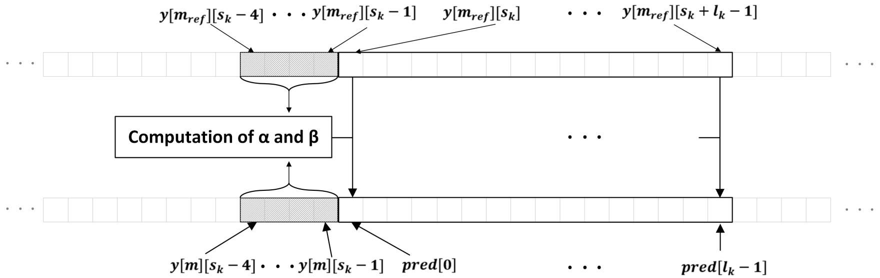


Illustration of Inter-Channel Prediction

Block-Copy Prediction

- Copy already reconstructed sample values of same channel:

$$\text{pred}[j] = y[m][s_k - l_k - t_r + j], \quad 0 \leq j < l_k.$$

- Location t_r of reference-block is transmitted.
- Half-sample accurate prediction supported.

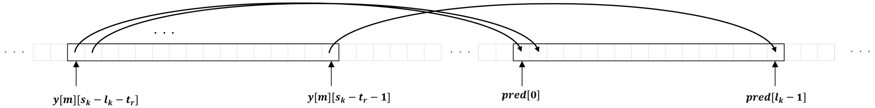


Illustration of Block-Copy Prediction

Transforms and quantization

Block transforms

Transform coding of prediction residuals

- Trigonometric transforms or identity transform can be used per block $b_{k,m}$.
- Basis functions of inverse transforms always supported on $b_{k,m}$.

Trigonometric transforms

- DCT-II supported if $l_k \leq 1024$.
- If $l_k < 256$: DST-VII also supported. Transform type signaled.
- Implemented in fixed point arithmetic with full matrix-vector multiplication.

Identity transform

- Can be combined with sample-wise prediction on block-prediction residuals.
- Applicable for residuals of Inter-Channel-, Block-Copy-, Zero-Prediction.
- Determined by weights $\alpha_1, \ldots, \alpha_K$, $1 \leq K \leq 20$.
- Weights either from fixed set or transmitted per block.

Quantization

- Scalar quantization, uniform reconstruction quantizer.
- Same stepsize for all transform coefficients.

Sample-wise residual prediction

Decoder perspective

- First: Decoding and reconstruction of intermediate residual $\hat{s}[0], \ldots, \hat{s}[l_k - 1]$.
- Then: Generation of final residual values:

$$\hat{u}[j] = \begin{cases} \hat{s}[j], & j \leq K \\ \hat{s}[j] + \sum_{p=1}^{K} \alpha_p \cdot \hat{u}[j-p], & j > K \end{cases}.$$

➡ Final reconstruction $\hat{y}[m][s_k + j] = pred[j] + \hat{u}[j]$.

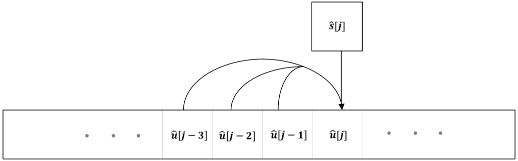


Illustration of the sample-wise residual prediction

Entropy coding

Entropy coding: Setup

Arithmetic coding

- Context based adaptive binary arithmetic coding (CABAC) used.
- Probability estimator and arithmetic coding engine based on MPEG NNC standard.

Context models

- Separate context models per channel.
- Separate context models for trigonometric and identity transforms.

Coding of trigonometric transform coefficients

- Let $c[0], \dots, c[l_k - 1]$ transform coefficients on block $b_{k,m}$ of length l_k
- Last position lastPos signaled first:

$$c[j] = 0 \quad \forall j: \ \text{lastPos} < j < l_k - 1 \ \&\& \ c[\text{lastPos}] \neq 0 \ \text{if } lastPos \neq 0.$$

- Then: Single pass coding in backwards-scan:

$$c[\text{lastPos}] \rightarrow c[\text{lastPos} - 1] \rightarrow \cdots \rightarrow c[0].$$

Coding of c[j] with lastPos $\geq$ j $\geq$ 0

- If $j \neq$ lastPos or lastPos $= 0$: Code significance flag,

$$\text{sig}[j] = \begin{cases} 1, & \text{if } c[j] \neq 0 \\ 0, & \text{else} \end{cases}.$$

- If sig$[j] \neq 0$: Code sign of $c[j]$, then absolute value:

$$|c[j]| = 1 + u[j] + v[j], \quad 0 \leq u[j] \leq u_{max} = 19, \quad 0 \leq v[j].$$

- Truncated unary coding of $u[j]$.
- If $u[j] = u_{max}$: Exponential Golomb coding of $v[j]$.

Coding of identity transform coefficients

Similar to trigonometric transforms. Changes:

- No coding of last position lastPos. Coding

$$c[l_k - 1] \rightarrow c[l_k - 2] \rightarrow \cdots \rightarrow c[0]$$

  in backwards scan and single past.
- First: Coding of significance flag sig[j].
- If sig[j] $\neq 0$: Coding of sign and of absolute value $|c(j)|$.
- For absolute value write:

$$|c[j]| = 1 + u[j] + r_1[j] + r_2\,[j]\,, \quad 0 \leq u[j] \leq u_{max} = 5, \quad 0 \leq r_1[j] \leq r_{max} = 40, \quad 0 \leq r_2[j].$$

- Truncated unary coding of $u[j]$.
- If $u[j] = u_{max}$: Rice coding of $r_1[j]$; adaptive Rice parameter selection, based on

$$\xi = \sum_{p=j+1}^{l_k - 1} |c[j]|\,.$$

- If $r_1[j] = r_{max}$: Exponential Golomb coding of $r_2[j]$.

Context modeling

Trigonometric transform:

- Context coding of significance flags sig[j].
  - 45 context models per channel.
  - Context-model index $cidx_{\text{sig}} \in \{0, \ldots, 44\}$ determined by j and by

$$\kappa = \sum_{p=0}^{min(2, l_k - j - 2)} |c[j + 1 + p]|$$

- Context coding of sign, one context model.
- Context coding of remainder $u[j]$:
  - 15 context models per channel.
  - For a fixed j, single context model for all bins of $u[j]$.
  - Context-model index $cidx_{gtr} \in \{0, \ldots, 14\}$ determined by position j.

Identity transform:

- Context coding of significance flags sig[j] and signs with single context models.
- Context coding of truncated unary parts for $u[j]$ and $r_1[j]$.
- 10 context models per channel in total.