



Compromise

Data compression paradigm based on omitting self-evident information

Maribor, 16. 01. 2023

David PODGORELEC & GeMMA Team



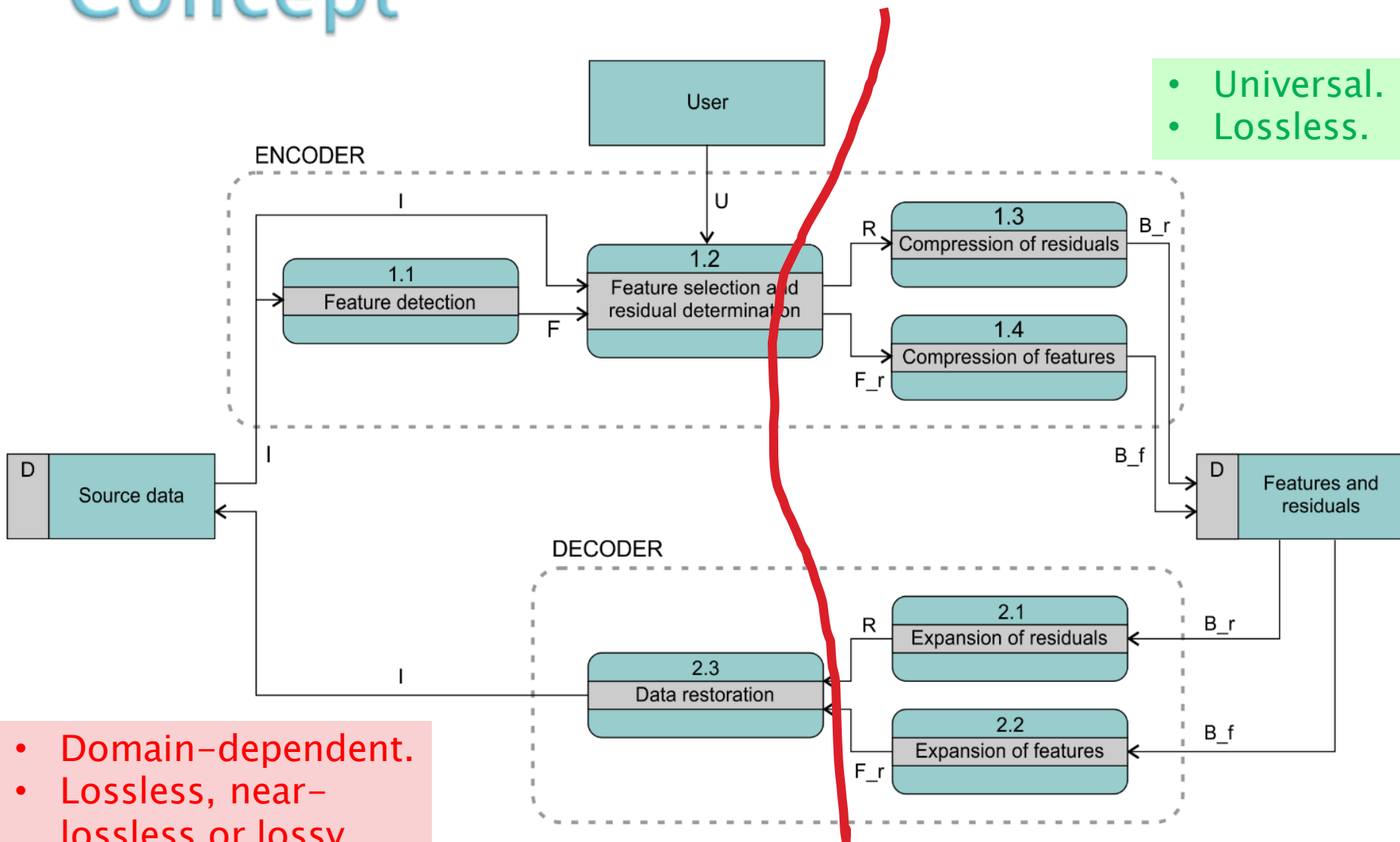
University of Maribor

Faculty of Electrical Engineering
and Computer Science

Institute of Computer Science

Laboratory for Geospatial Modelling, Multimedia and Artificial Intelligence

Concept



- Universal.
- Lossless.

- Domain-dependent.
- Lossless, near-lossless or lossy.

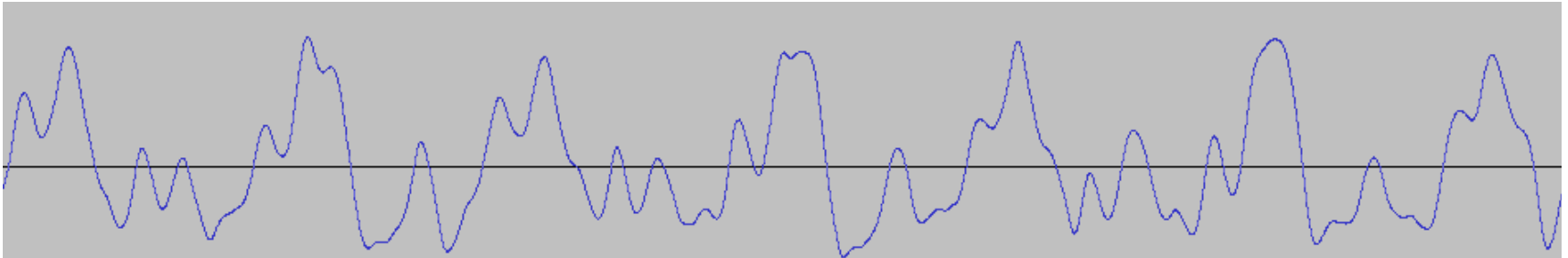
How to start

- ▶ **STEP 1:** Choose your data domain.
 1. Raster images.
 2. Digital audio.
 3. Biomedical signals (promised to be covered by UWB).
 4. Sparse voxel grids.
 5. Other ideas also acceptable to prove the concept.
- Within the chosen domain(s), think about reference state-of-the-art results (e.g. PNG, JPEG-LS, JP2 for lossless images, FLAC, MPEG-4 ALS for lossless digital audio,...).
- Near lossless extension also needed. Also ability for lossy mode, but we do not compete with lossy methods results!
- Near lossless = lossy with locally controlled errors.

How to start

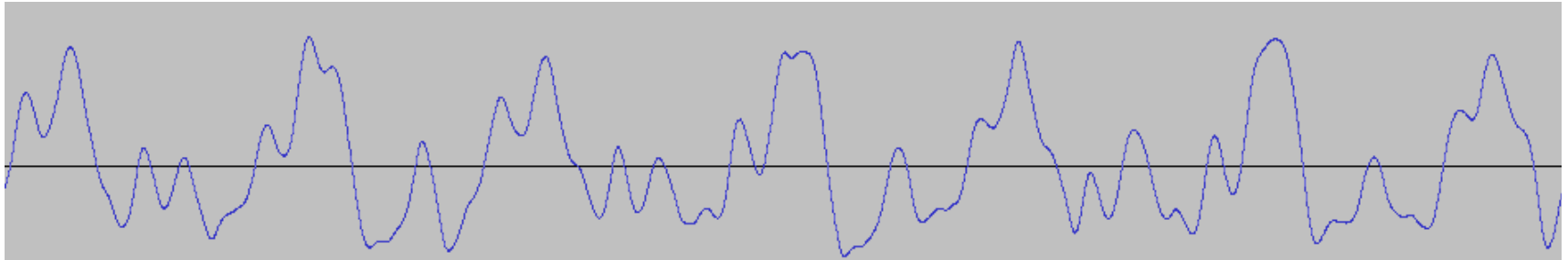
- ▶ **STEP 2:** Think about features to be extracted from your input data streams.
 - Domain-dependent feature repertoires till **M3**.
 - Universal feature taxonomy till **M6**.
 - Think about universal feature classes while designing domain-dependent features already. *Does my domain-dependent feature make sense in some other pilot domain as well?*
 - Examples of features that could possibly be generalized : extreme, sequence, border, pattern, region of interest, rhythm (repetition, trend, ...), symmetry, self-similarity... Use your imagination here!

STEP 3: feature detection

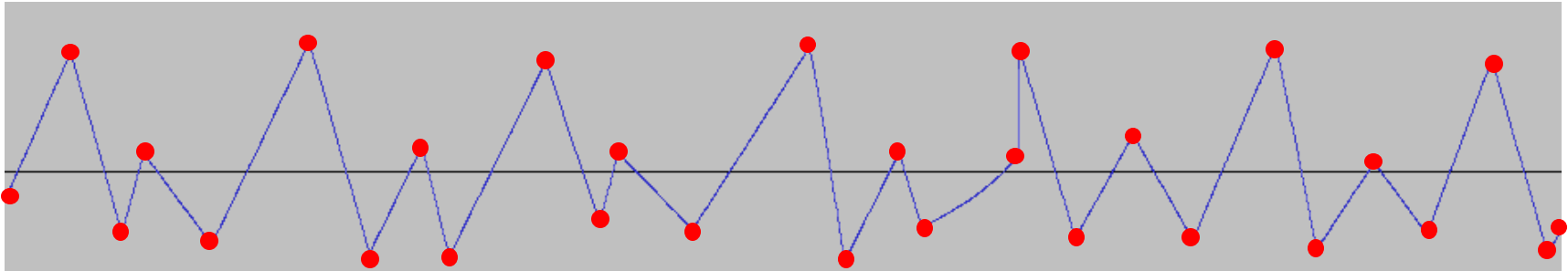


Input stream: audio waveform (mono).

STEP 3: feature detection

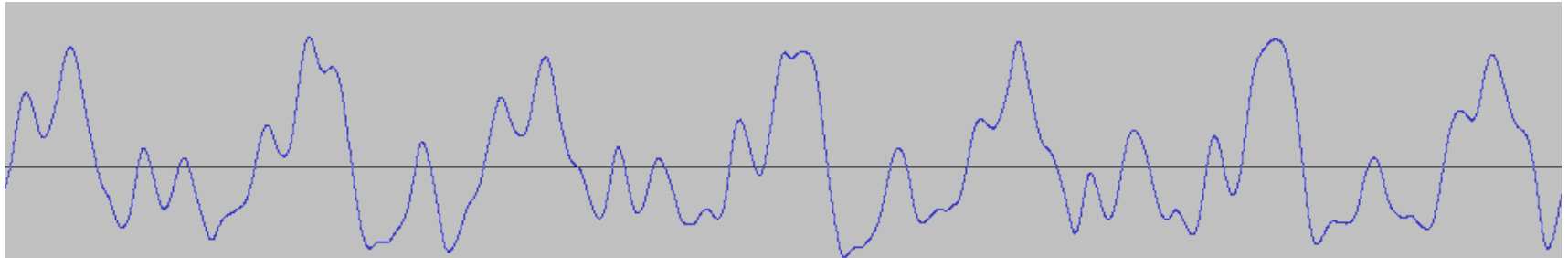


Input stream: audio waveform (mono)

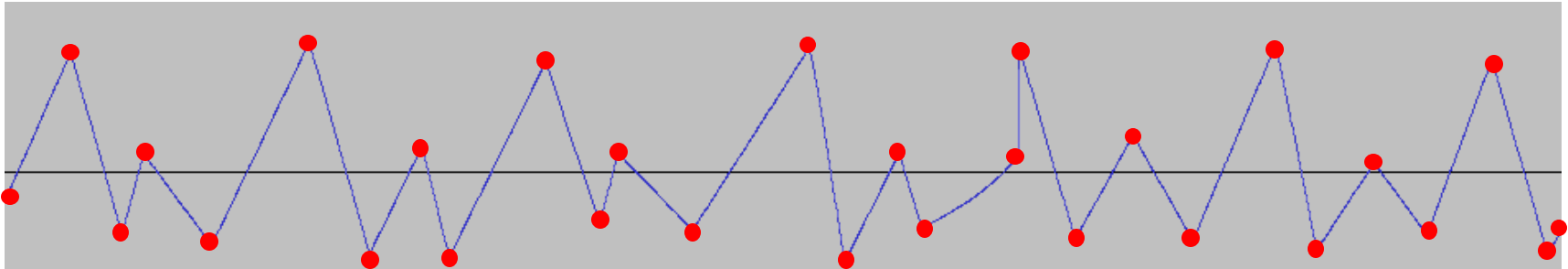


Features (**extremes**) and the **stream restored** from features.

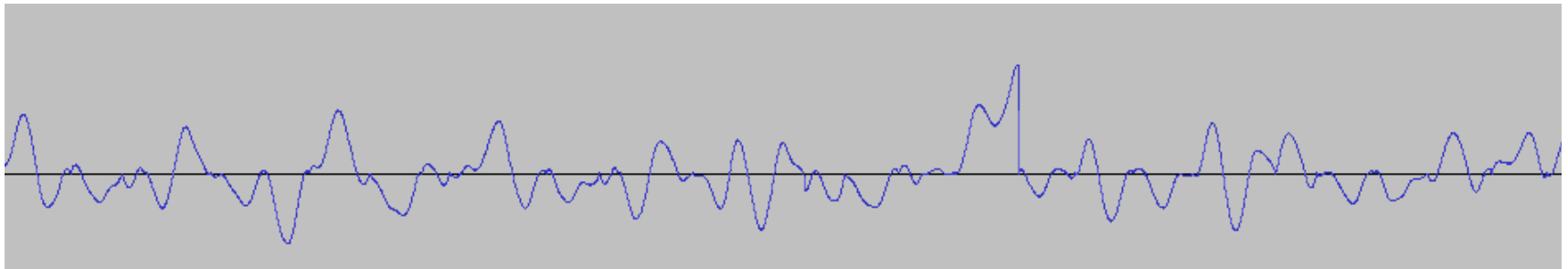
STEP 4: feature selection and residual determination



Input stream: audio waveform (mono).

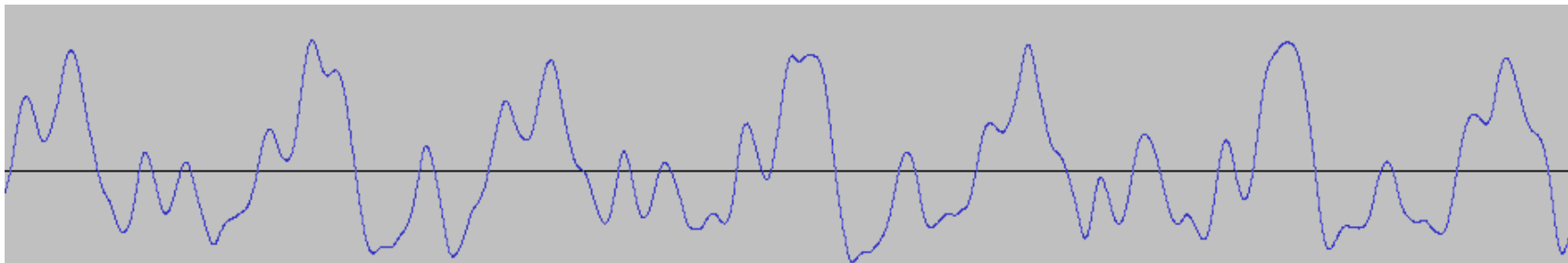


Features (**extremes**) and **stream restored** from features only.

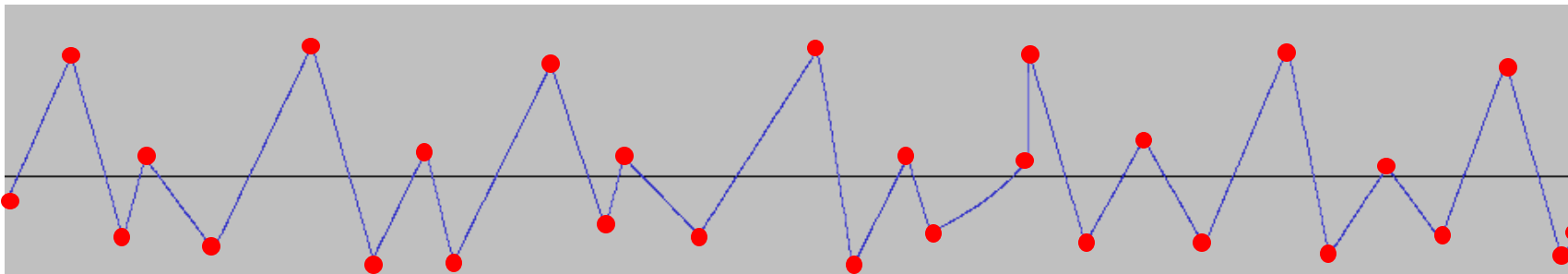


Residuals: Difference Input stream – Stream restored from features.

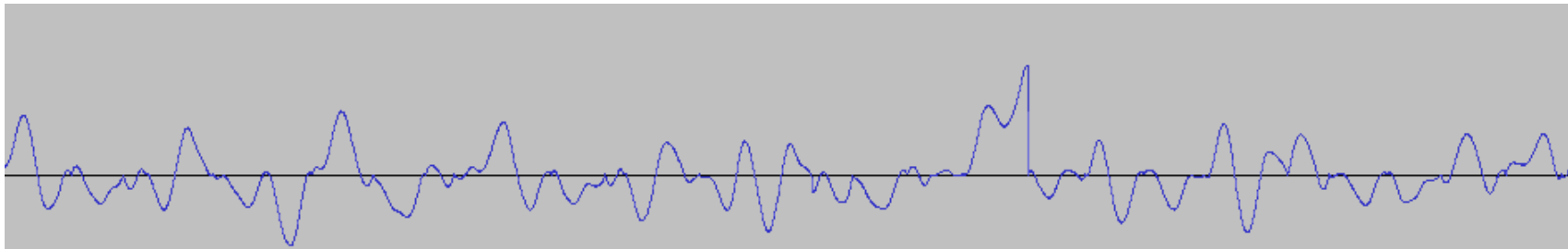
STEP 4: feature selection and residual determination



Input stream: audio waveform (mono).

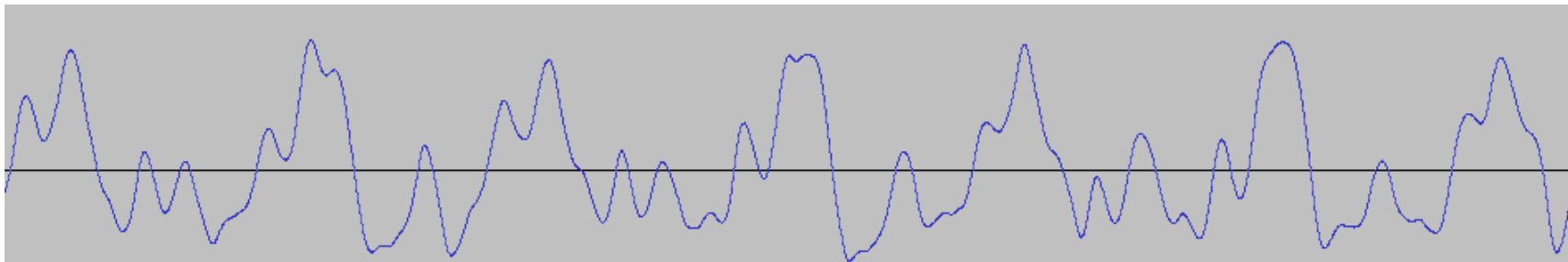


Features (**extremes**) and **stream restored** from features only.



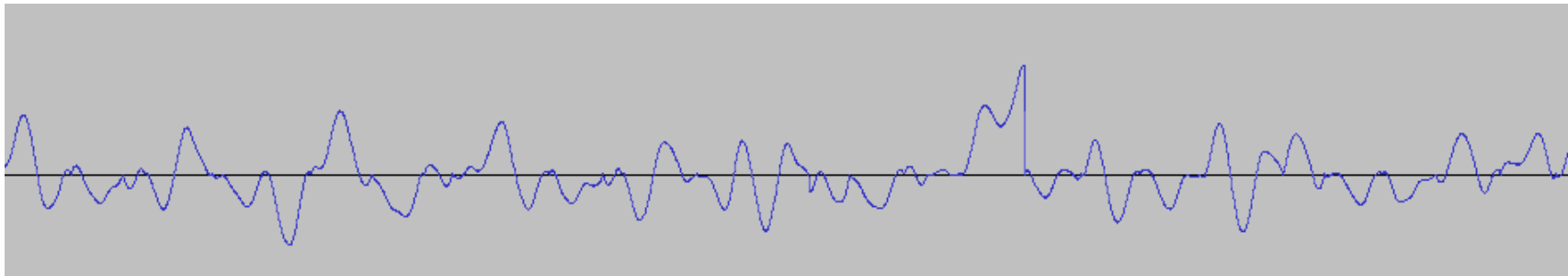
Residuals: Difference Input stream – Stream restored from features.

STEP 4: feature selection and residual determination



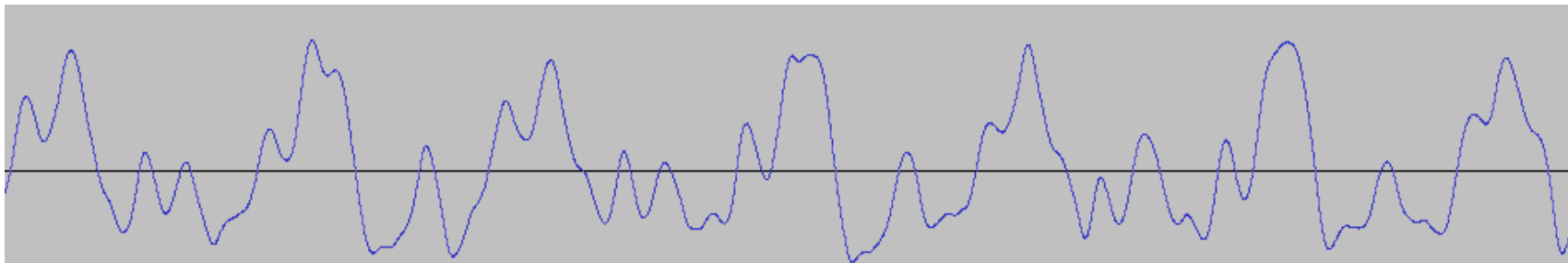
Input stream: audio waveform (mono).

- ▶ Features: expectedly below 10% of output data.
- ▶ Residuals are more critical.
 - FLAC compresses the input stream a bit more successfully than the residuals. And we still need 10% for features!



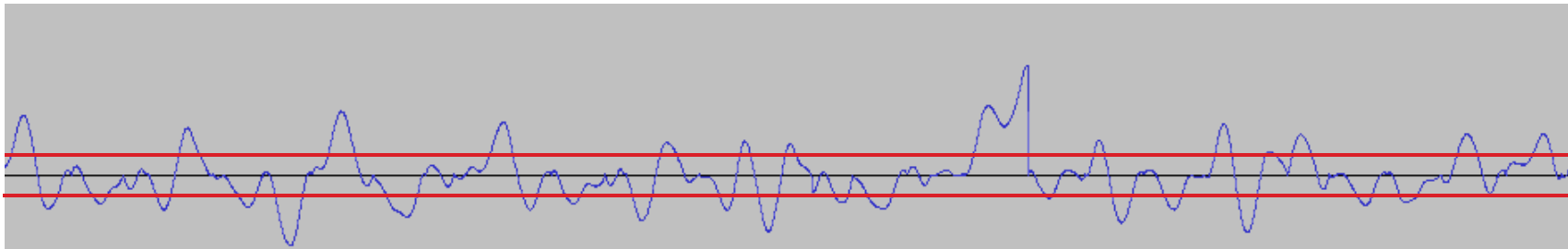
Residuals: Difference Input stream – Stream restored from features.

STEP 4: feature selection and residual determination



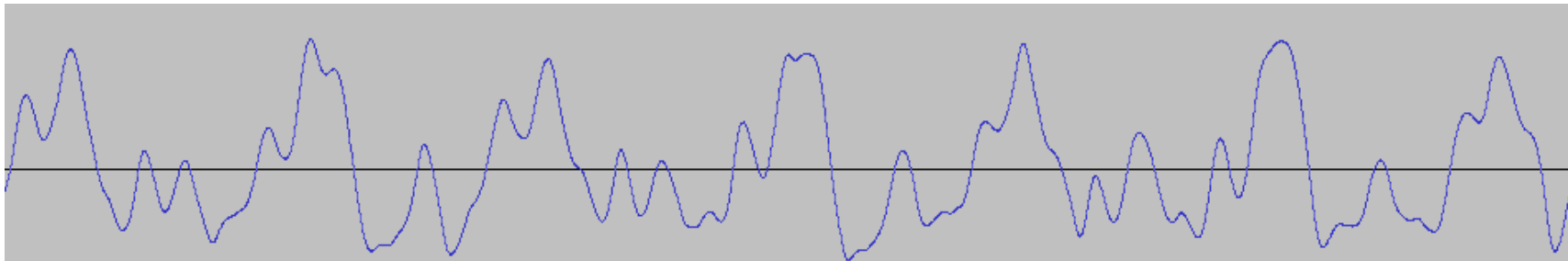
Input stream: audio waveform (mono).

- ▶ Iterative optimization process.
- ▶ E.g. constrain local errors (amplitudes of residuals), refresh features in critical areas, compute new residuals...



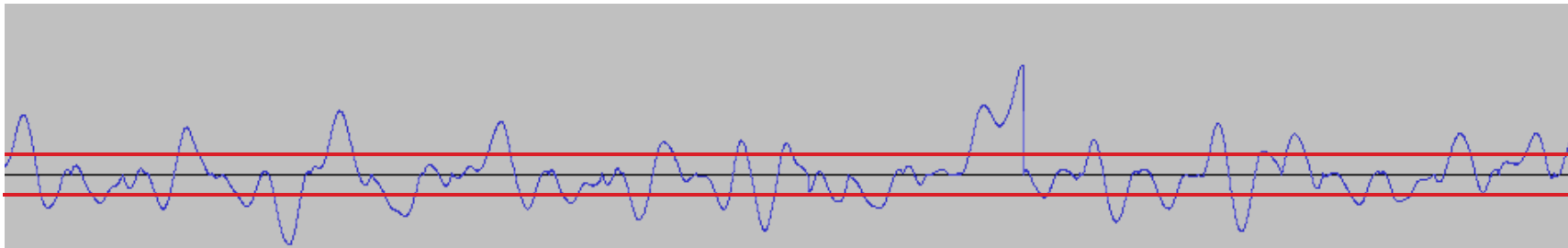
Residuals: Difference Input stream – Stream restored from features.

STEP 4: feature selection and residual determination



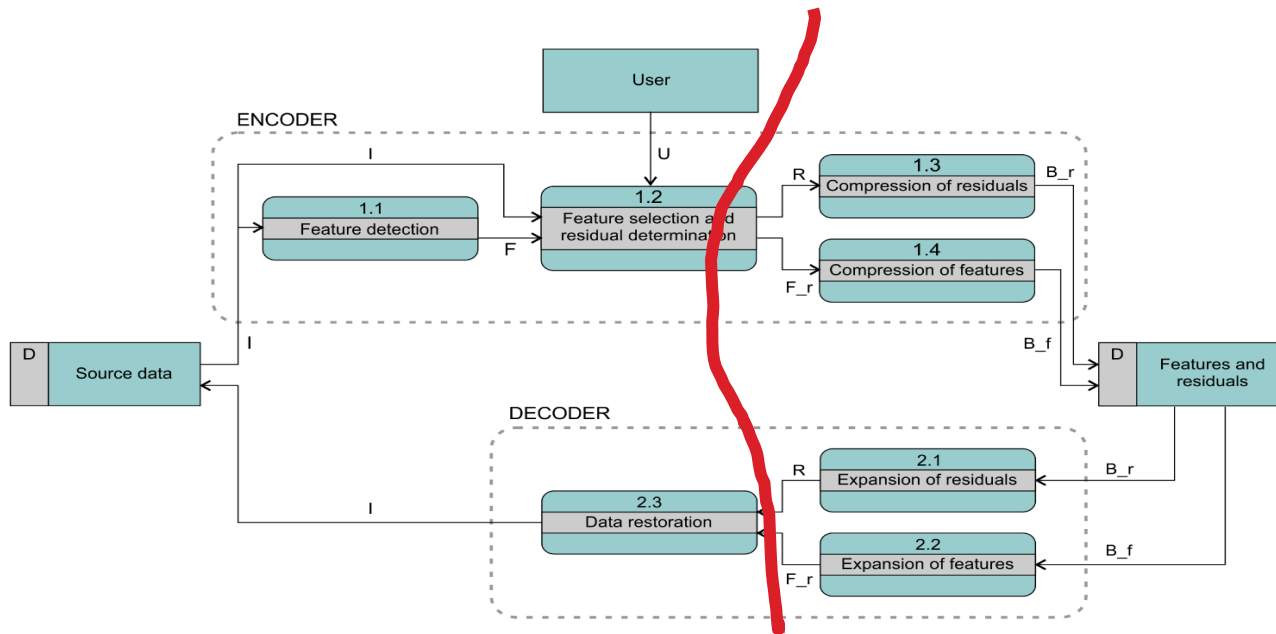
Input stream: audio waveform (mono).

- ▶ Perform optimization on the domain-dependent level.
- ▶ Lossless, near lossless and lossy separation made here.
- ▶ „Translate“ features and residuals to the unified level.
- ▶ Further optimize at unified level (maybe).



Residuals: Difference Input stream – Stream restored from features.

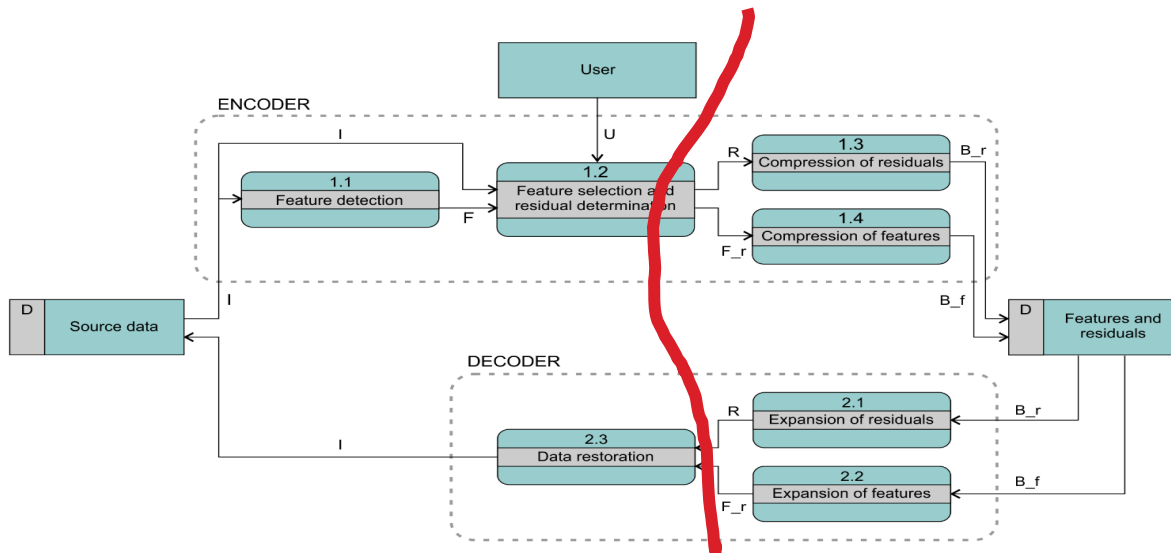
STEP 4: feature selection and residual determination



▶ Red line: feature interpretation.

- → translate domain-dependent features to unified taxonomy.
- → no further changes on residual stream expected.
- ← reverse from the above.

Step 5...: beyond the red line



- ▶ Study distributions of residuals and individual unified feature types and choose optimal lossless compression method.
- ▶ Ensembles and pipelines of different lossless data transformation and compression methods also acceptable.
- ▶ Can be and **must be partially done in parallel with 1.1** (and then in parallel with 1.2 and 2.3) – see WP3 and WP4 in workplan.

Deliverables

- ▶ At least 3 papers in international open access journals,
- ▶ at least 6 conference papers,
- ▶ (above papers not inked to individual WPs. Domain-dependent results also count.)
- ▶ organisation of 2 dedicated presentation events,
- ▶ 1 patent application,
- ▶ eventual additional requirements from GAČR(?),
- ▶ Website and a profile on at least 1 social network (after M6)
 - <https://gemma.feri.um.si/projects/slovene-national-research-projects/j2-4458-data-compression-paradigm-based-on-omitting-self-evident-information/eng/>
- ▶ ongoing results at the end of individual WPs (plans, reports, instructions, software, test datasets...).