INSTITUTE OF VISUAL COMPUTING

TU Graz

SCIENCE
PASSION
TECHNOLOGY

Progressive Visual Segmentation of Vibration Signals to Detect Change Points

Julian Rakuschek, Johanna Schmidt, Udo Schlegel, Michèle Posch, Jutta Isopp, Tobias Schreck

VMV 2026 - 25.09.2026

All about vibrations

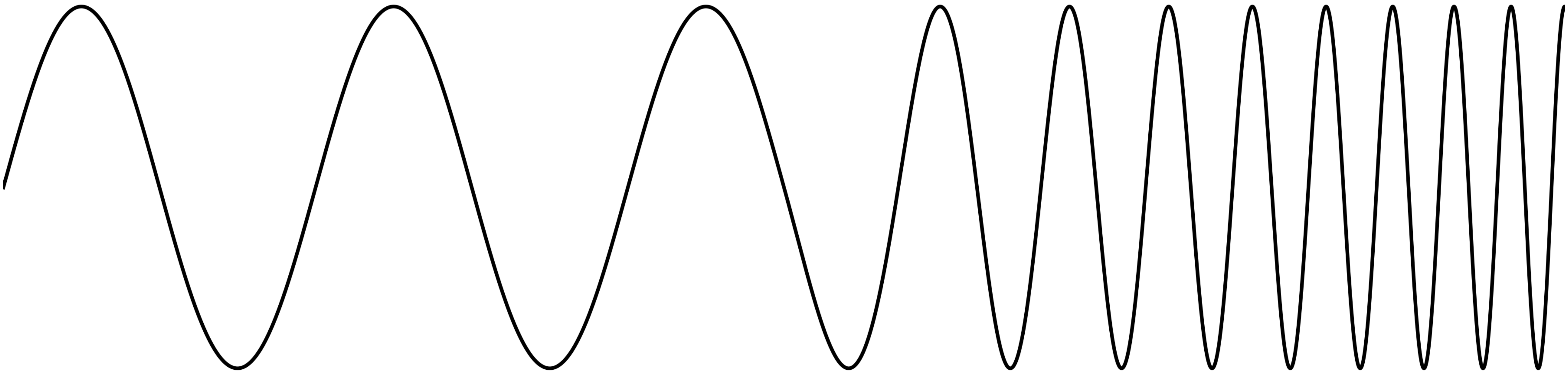

+


=


Engine Sensor Vibration signal

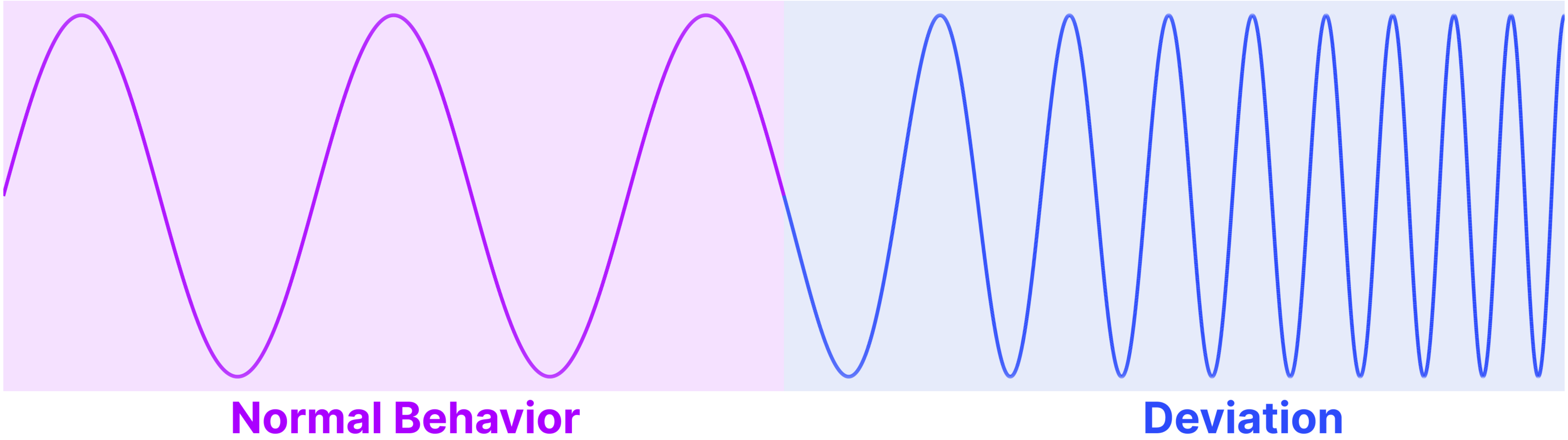
Why analyze vibrations?

Find long-term deviations



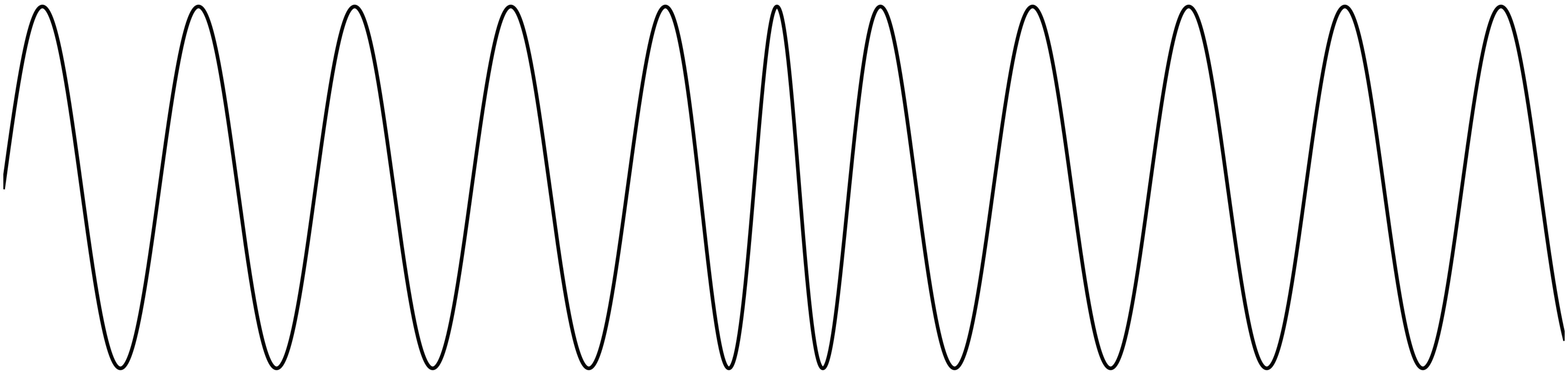
Why analyze vibrations?

Find long-term deviations



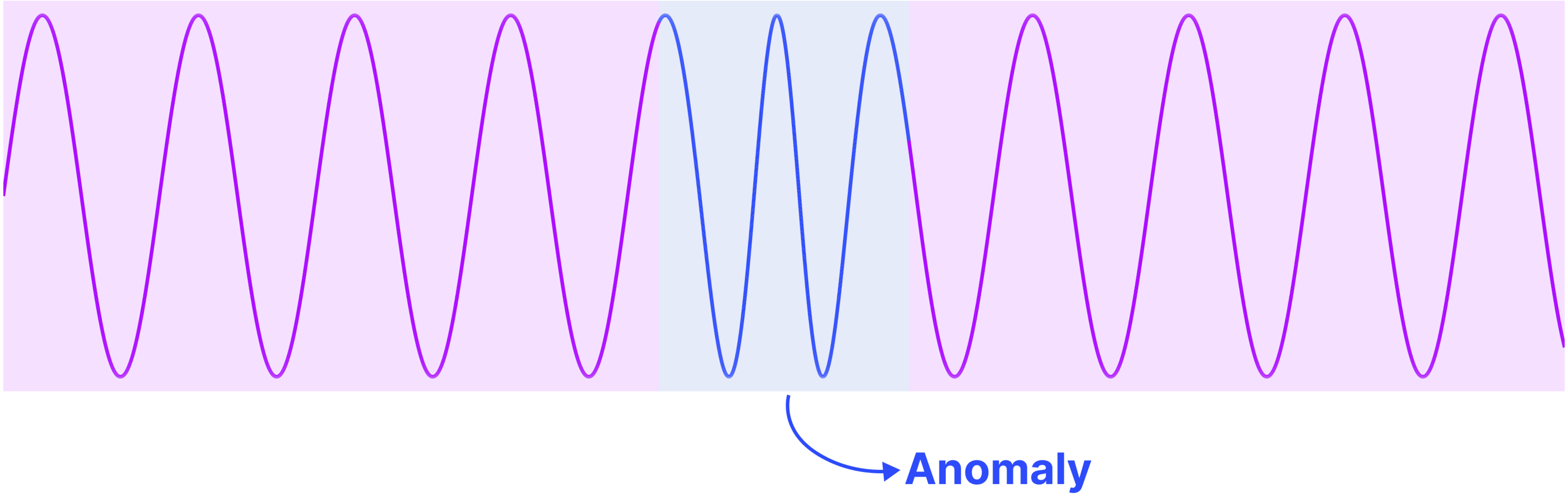
Why analyze vibrations?

Find anomalies (short-term, unexpected)



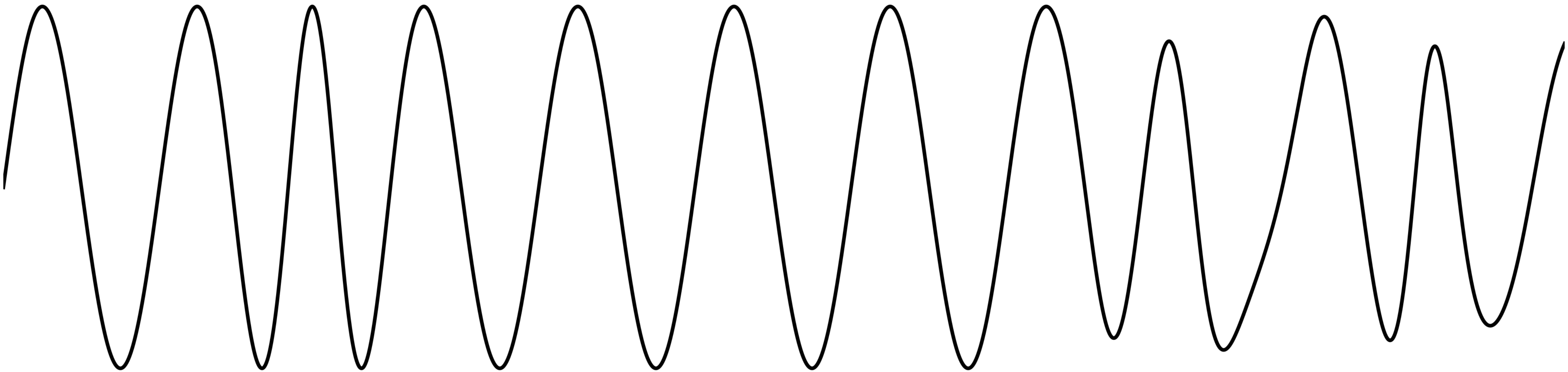
Why analyze vibrations?

Find anomalies (short-term, unexpected)



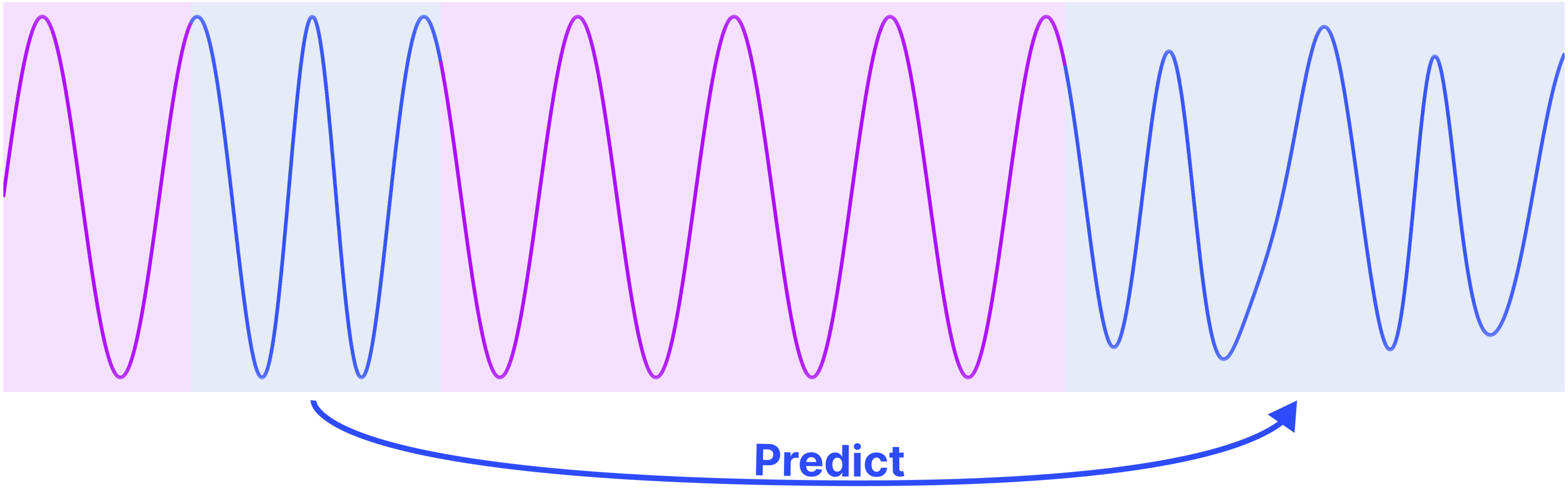
Why analyze vibrations?

Predict faults

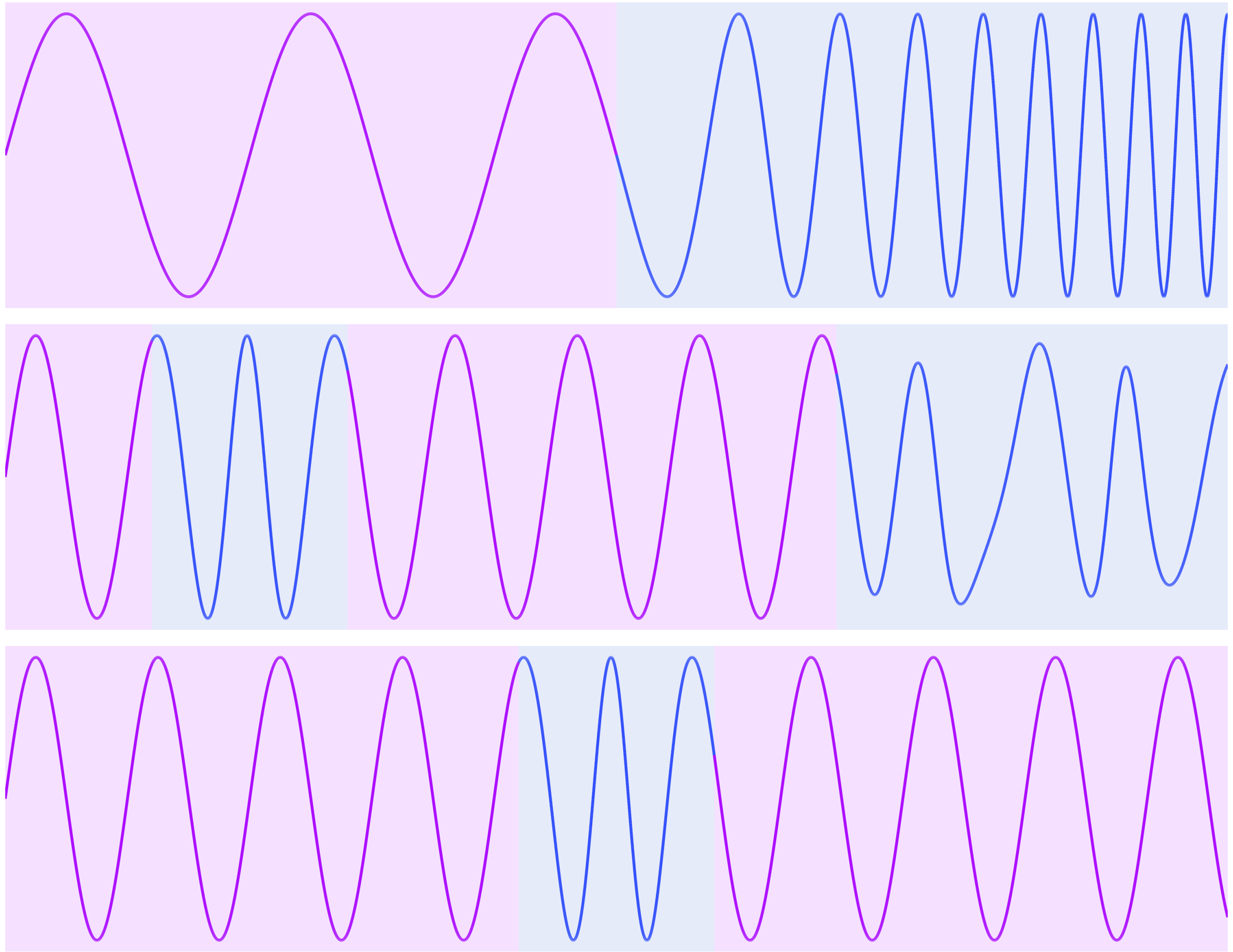


Why analyze vibrations?

Predict faults



A good segmentation can help solve these tasks!



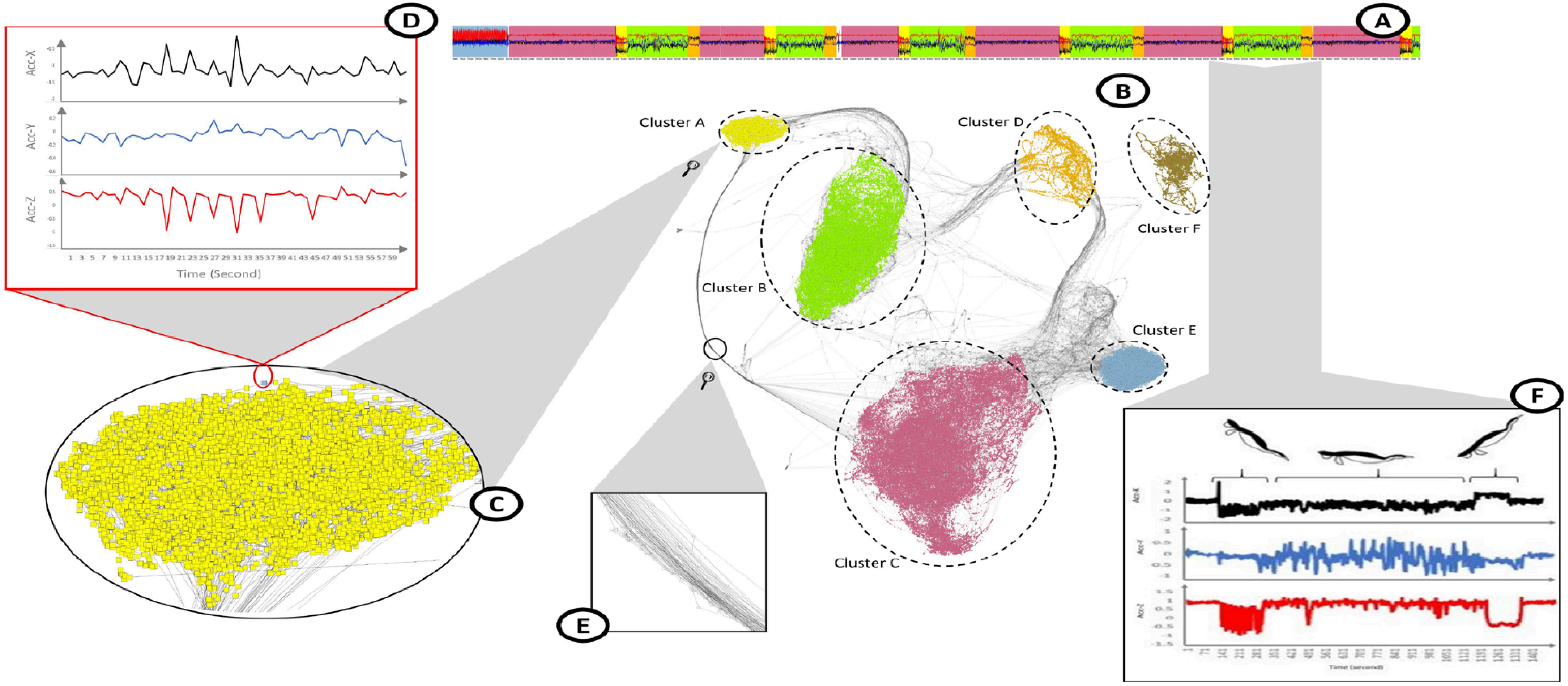
Condition Monitoring

Predictive Maintenance

Anomaly Detection

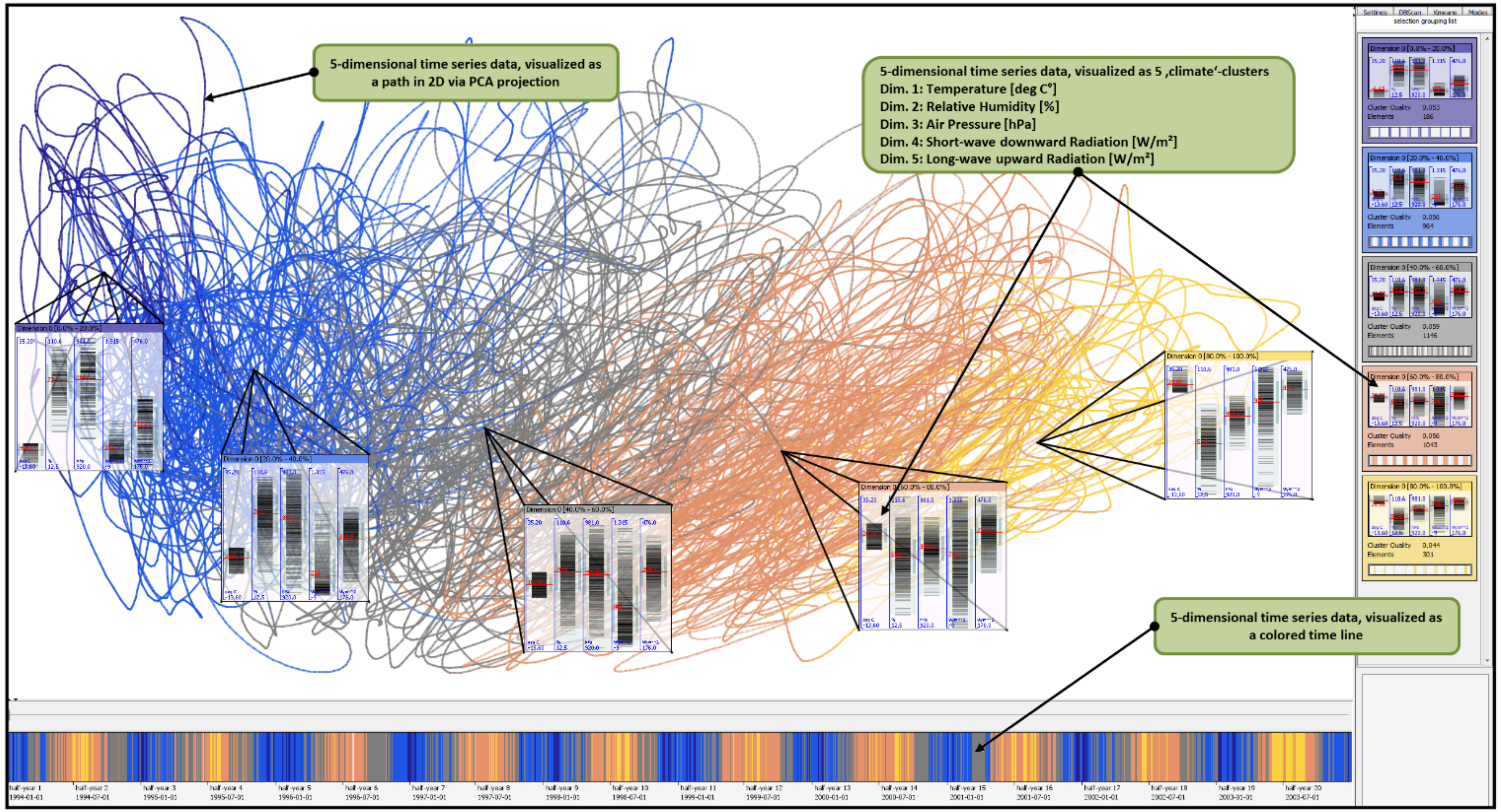
The transitions between segments mark **change points**

Related Work (1)



Ali, Mohammed et al. “TimeCluster: dimension reduction applied to temporal data for visual analytics.”
The Visual Computer 35 (2019): 1013 - 1026.

Related Work (2)



Bernard, J. et al. "TimeSeriesPaths : Projection-Based Explorative Analysis of Multivariate Time Series Data." International Conference in Central Europe on Computer Graphics and Visualization (2012).

Related Work (3)

DendroTime: Progressive Hierarchical Clustering for Variable-Length Time Series

Sebastian Schmidl
Hasso Plattner Institute, University of Potsdam
Potsdam, Germany
sebastian.schmidl@hpi.de

Ferdinand Rewicki
German Aerospace Center (DLR), Friedrich Schiller University Jena
Jena, Germany
ferdinand.rewicki@dlr.de

Felix Naumann
Hasso Plattner Institute, University of Potsdam
Potsdam, Germany
felix.naumann@hpi.de

Thorsten Papenbrock
Philipps University of Marburg
Marburg, Germany
papenbrock@informatik.uni-marburg.de

Abstract

Many effective dissimilarity measures for variable-length time series, such as DTW, MSM, or TWED, are expensive to compute because their runtimes increase quadratically with the time series' lengths. When used in hierarchical agglomerative clustering algorithms that need to compute all pairwise time series dissimilarities, they cause slow runtimes and do not scale to large time series collections. However, there are use cases, where fast, interactive hierarchical clustering is necessary. For these use cases, progressive hierarchical clustering algorithms can improve runtimes and interactivity. Progressive algorithms are incremental algorithms that produce and continuously improve an approximate solution, which eventually converges to the exact solution.

In this paper, we present DendroTime, the first (parallel) progressive clustering system for variable-length time series collections. The system incrementally computes the pairwise dissimilarities between the input time series and supports different ordering strategies to achieve progressivity. Our evaluation demonstrates that DendroTime's progressive strategies are very effective for clustering scenarios with expensive time series dissimilarity computations.

Keywords

Hierarchical Clustering, Time Series, Progressive Algorithm, Dendrogram, EDEN ISS, Pattern Mining, Classification

1 Clustering time series

Time series clustering is an active and popular research area with hundreds of publications per year [1, 21, 22, 24]. The objective is to group similar time series, such that the time series within a group are more similar to each other than time series of different groups. Time series clustering is often the starting point for exploratory analysis and, thus, applied in many application domains, such as greenhouse monitoring [51], speech recognition [18], environmental monitoring [9], healthcare [56], and energy management [19]. A clustering can be calculated in different ways, including feature-based (e. g., Time2Feat [8]), partitional sequence-based (e. g., k-Means [22] or k-Shapes [41]), or hierarchical sequence-based (e. g., hierarchical agglomerative clustering (HAC) [1, 24, 41]).

Figure 1: Dendrogram quality (WHS) and convergence indicator (*#CumClusterChanges@k*) of DendroTime's *ada* (progressive) and *fcfs* (non-progressive) strategies for the ACSF1 dataset with *Move-Split-Merge* [57] (MSM) and *average*-linkage. The vertical dashed line marks the switch from the approximate to the exact dissimilarity computation phase.

In this paper, we propose a novel HAC technique because HAC is particularly useful for the (interactive) clustering of time series [1, 14, 24, 41]: It can leverage elastic dissimilarity measures to cluster variable length input time series, it does not require users to specify the number of clusters upfront [1], and it achieves high comparative qualities in benchmarks with elastic dissimilarity measures [24, 61]. However, because HAC has to call a usually costly (elastic) dissimilarity function for all pairs of time series, it scales poorly and works for only small datasets [1].

To deal with the high computational complexity of HAC, researchers have proposed approximate HAC algorithms [12, 25, 26, 61] that compute the dissimilarities for a small selection of time series pairs and infer them for the remaining ones. Because this strategy works only for certain dissimilarity measures and linkage methods, existing approximate HAC algorithms are limited in their effectiveness and applicability. So to control the computational complexity of HAC and integrate it into interactive clustering processes, we propose to calculate a hierarchical clustering progressively.

Progressive algorithms [5, 33, 44, 45, 55, 62, 68] are incremental algorithms that try to achieve better early qualities than traditional incremental algorithms; they produce the same results as exact algorithms when they terminate. Progressive algorithms are different from online or streaming algorithms, in that they process static datasets but intend to gradually improve the result quality. A progressive clustering algorithm should, thus, quickly compute an approximate dendrogram that rapidly converges to the exact dendrogram. It allows scientists to monitor the dendrogram over time and stop the clustering process early, shortening

EDBT '26, Tampere (Finland)

Schmidl, Sebastian et al. "DendroTime: Progressive Hierarchical Clustering for Variable-Length Time Series." International Conference on Extending Database Technology (2026).

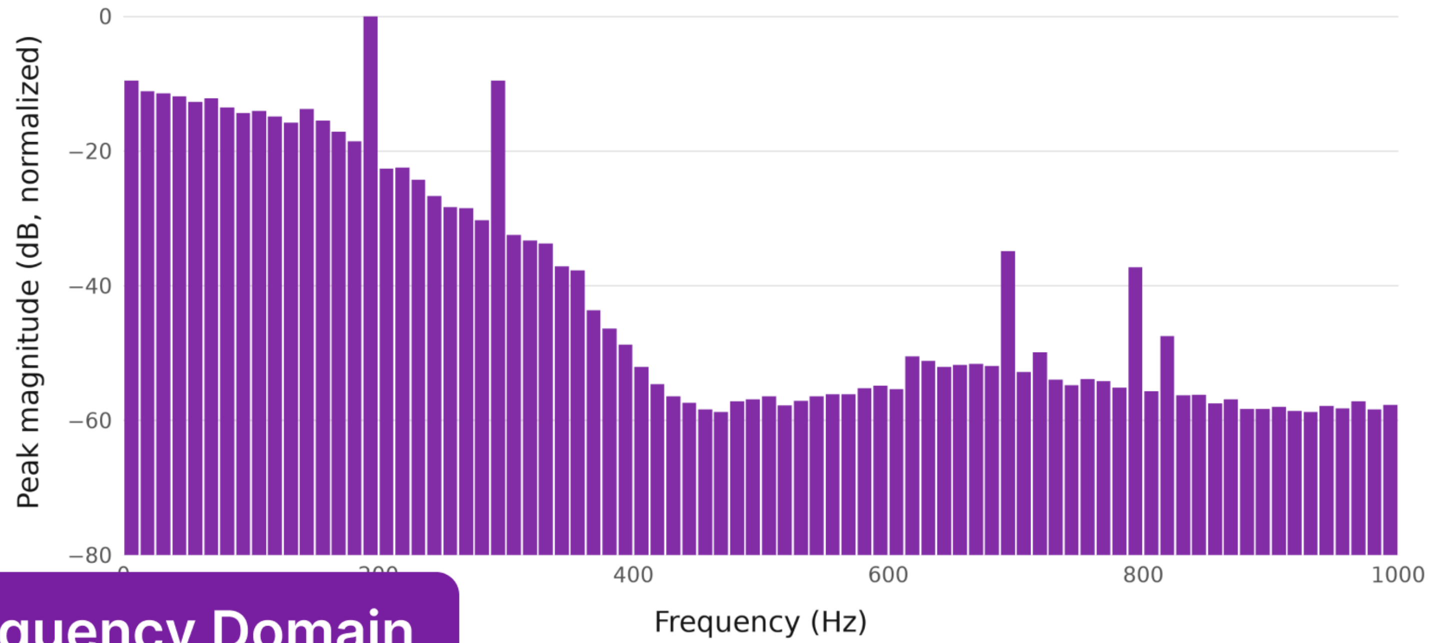
This is hard to read

We transform it to the frequency domain

Time Domain

Fast Fourier Transform

Frequency Domain



The big problem with vibrations: Large data volumes

High sampling rate = large data volumes

Example: 30 kHz (30 000 samples per second)

Assumption: 8 byte per sample

After one minute: 14,4 MB

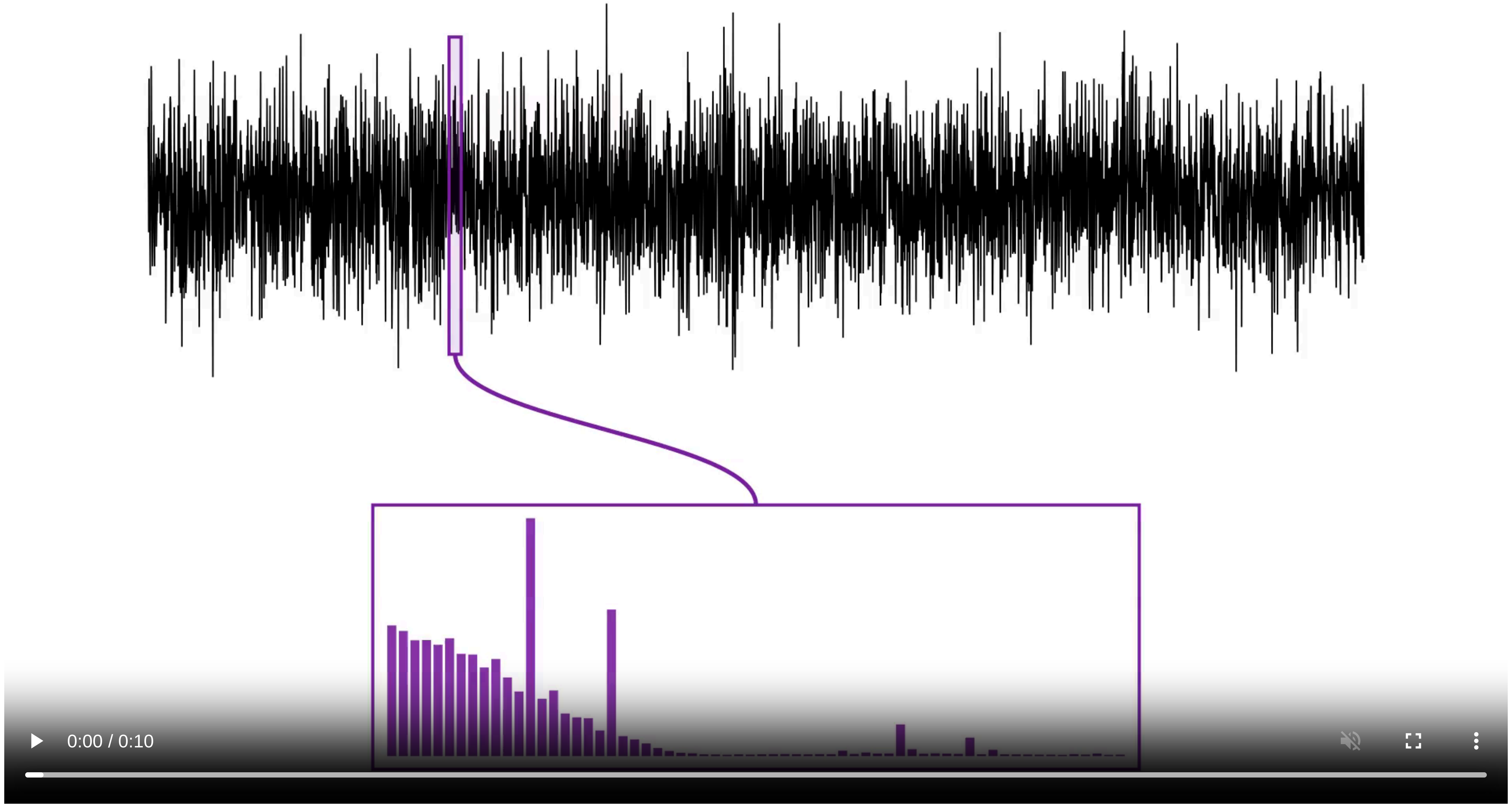
After one hour: 864 MB

After one day: 20,7 GB

After a year: 7,5 TB

How do domain experts approach this?

Signal Analysis via Manual Window Positioning

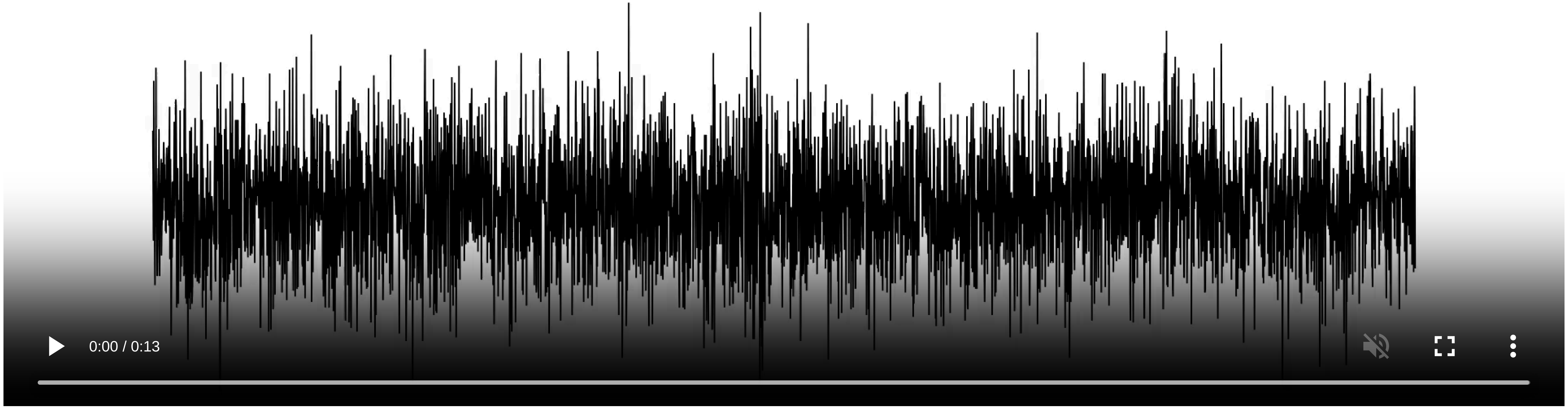


How about 10TB?

Manual scanning is tedious!

Our approach: Sampling

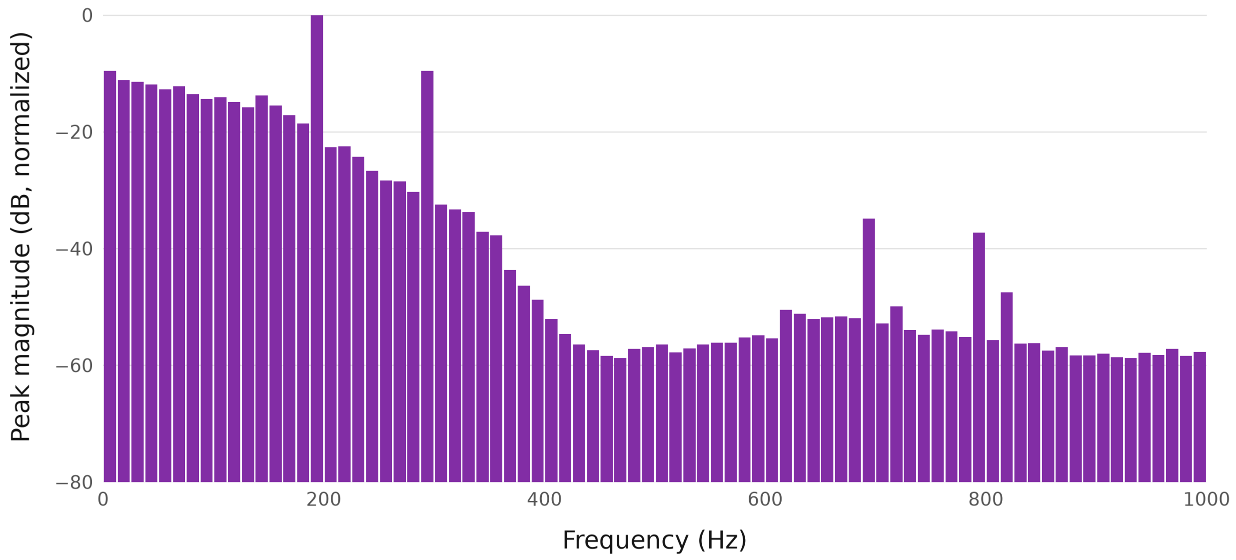
We iteratively construct an overview



In many cases a small fraction of the data suffices!

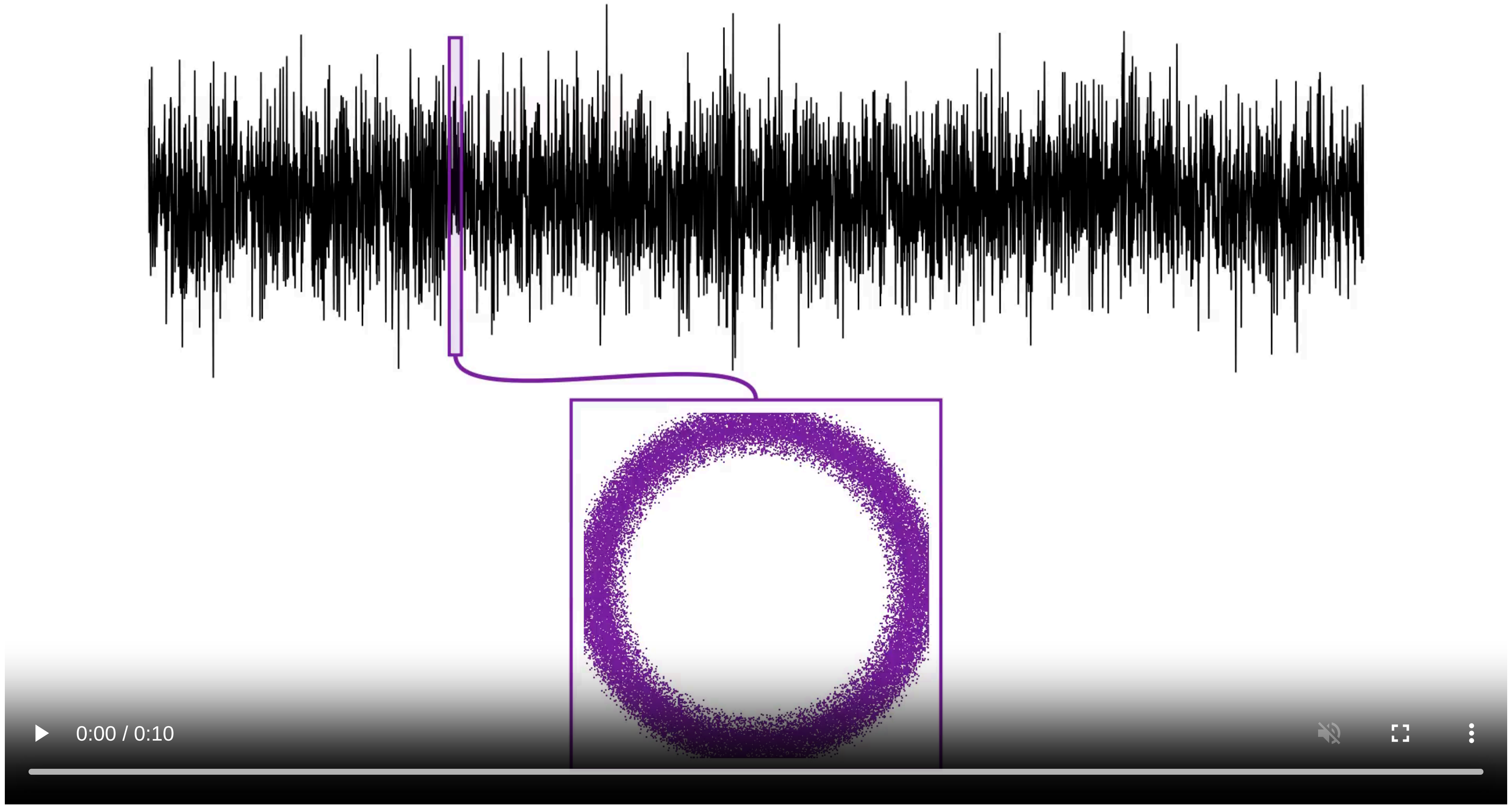
Do we need to save the full signal?

No!



We only save a FFT histogram per sample into the database.

Further representations possible



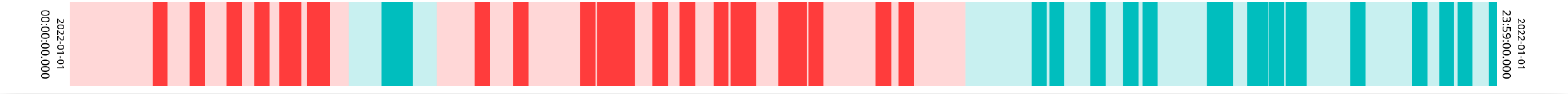
Time Series Projections

We start with an empty timeline

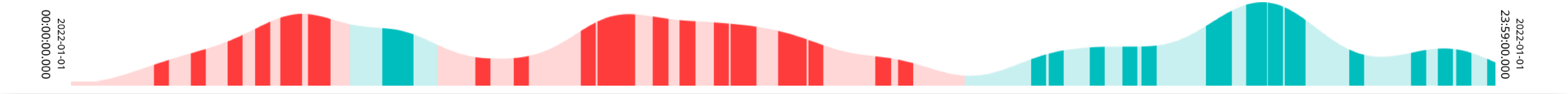
Next we randomly draw samples

The subsequences are groupes by similarity

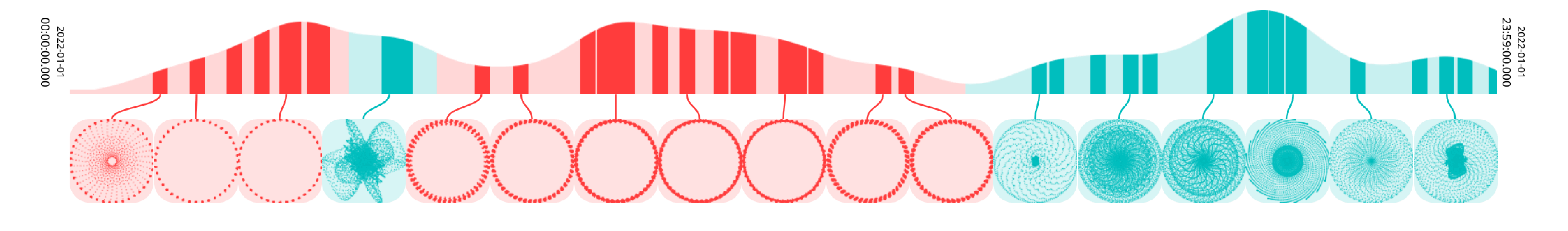
The background is colored by closest label



The height is adjusted based on the density



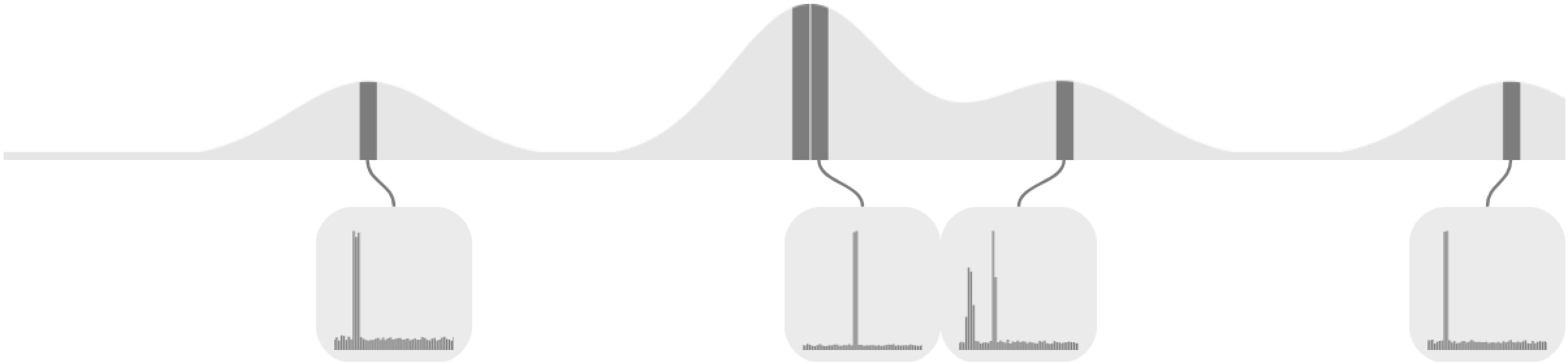
For a subset of samples we can show representations as thumbnails.



We obtain a segmentation

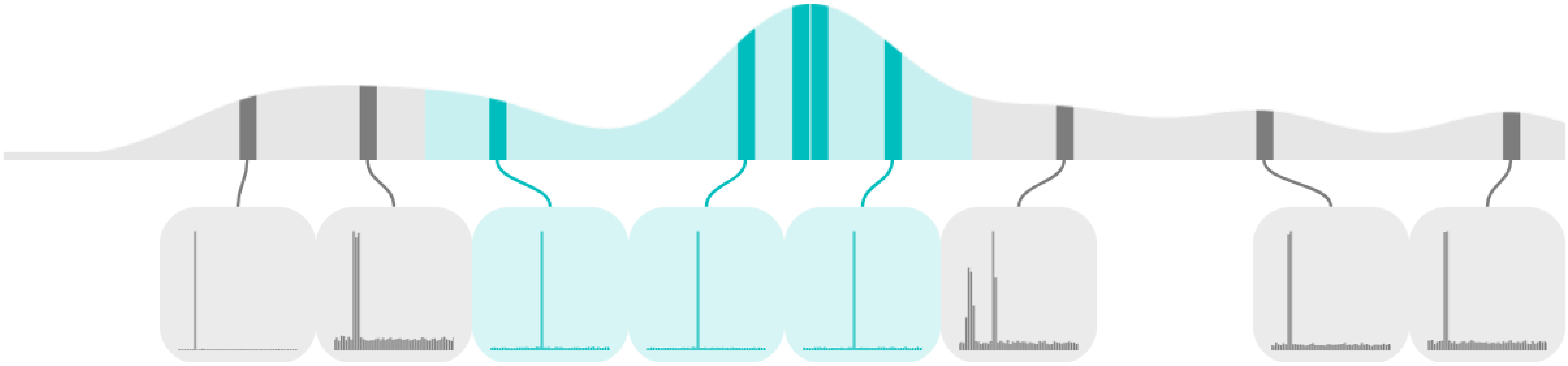
With more samples, the segmentation becomes more accurate.

FFT Feature Descriptors



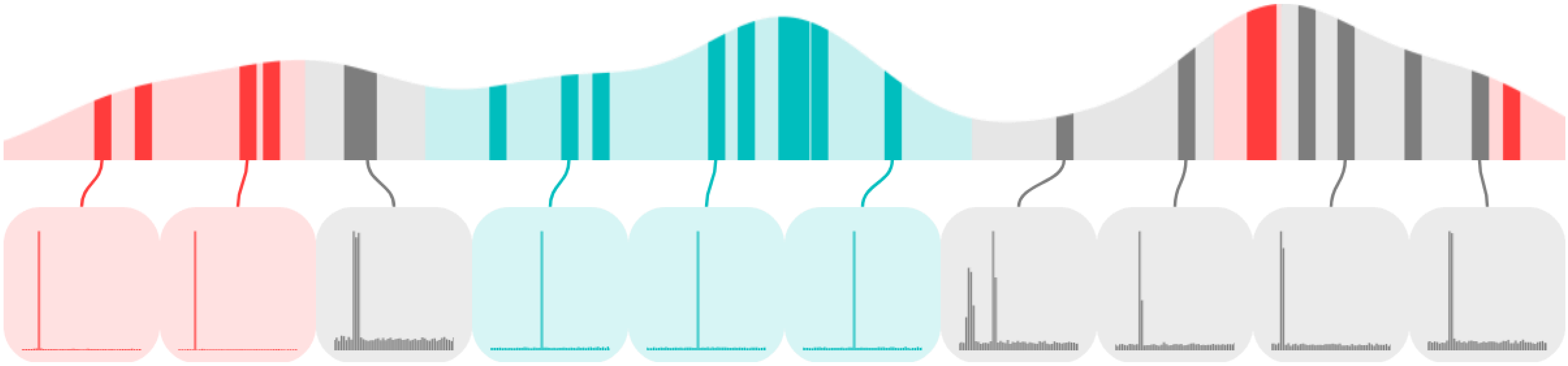
5 samples

FFT Feature Descriptors



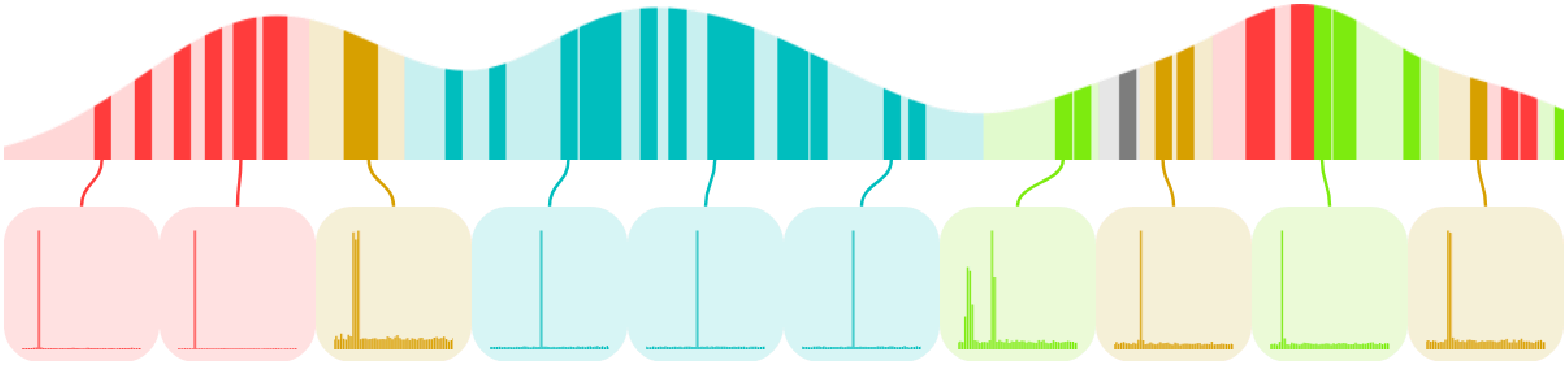
10 samples

FFT Feature Descriptors



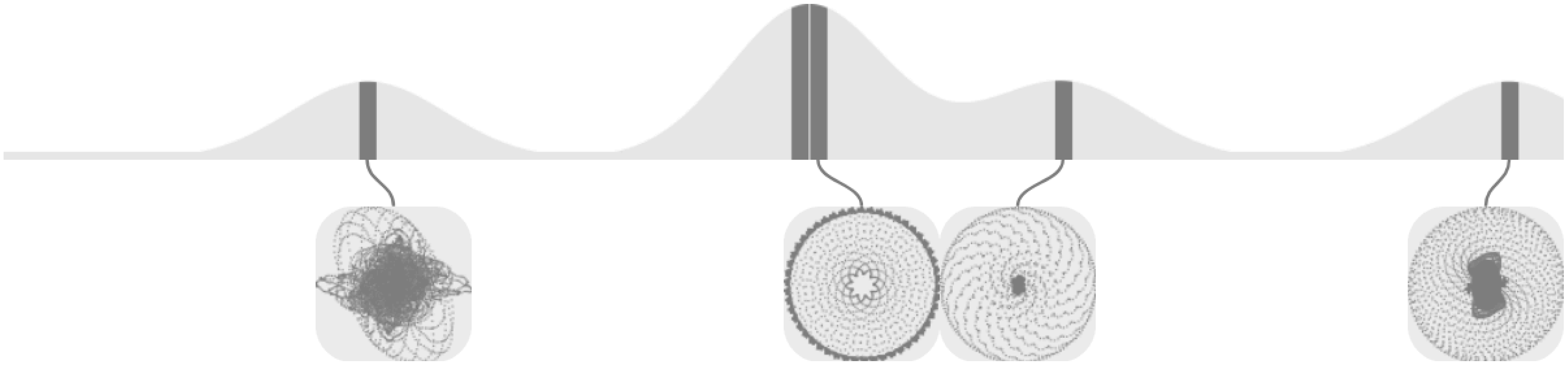
25 samples

FFT Feature Descriptors



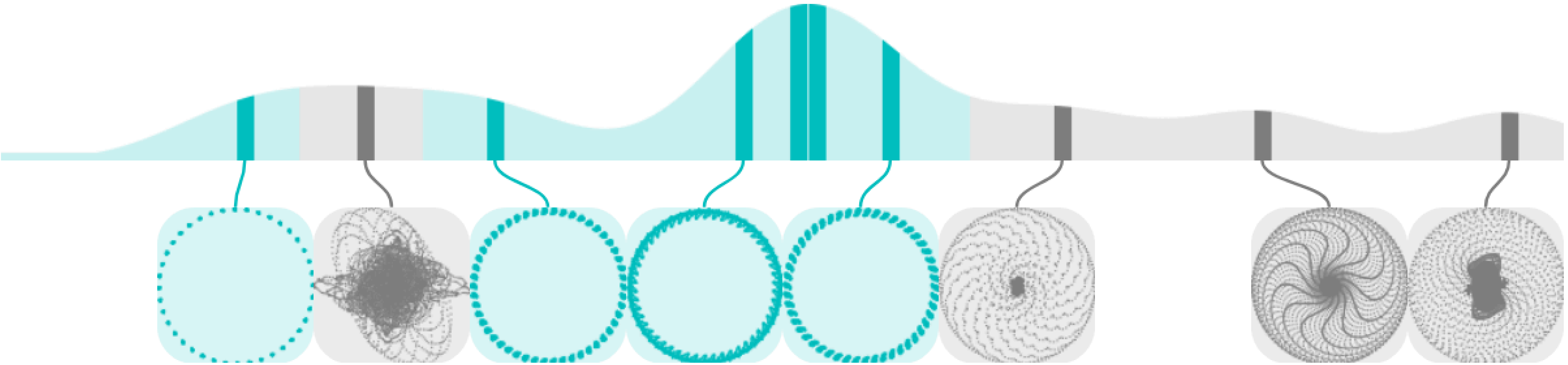
50 samples

Projection-Based Feature Descriptors



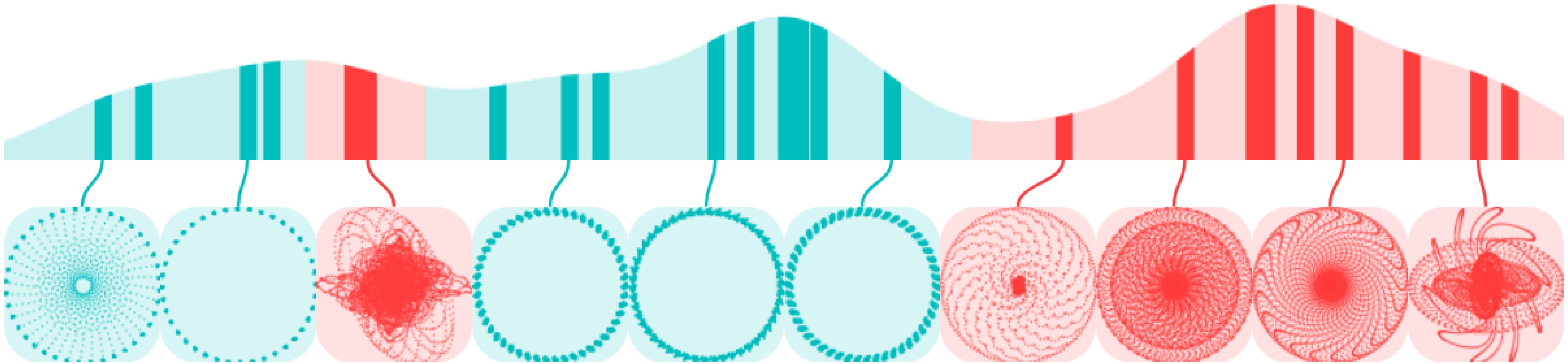
5 samples

Projection-Based Feature Descriptors



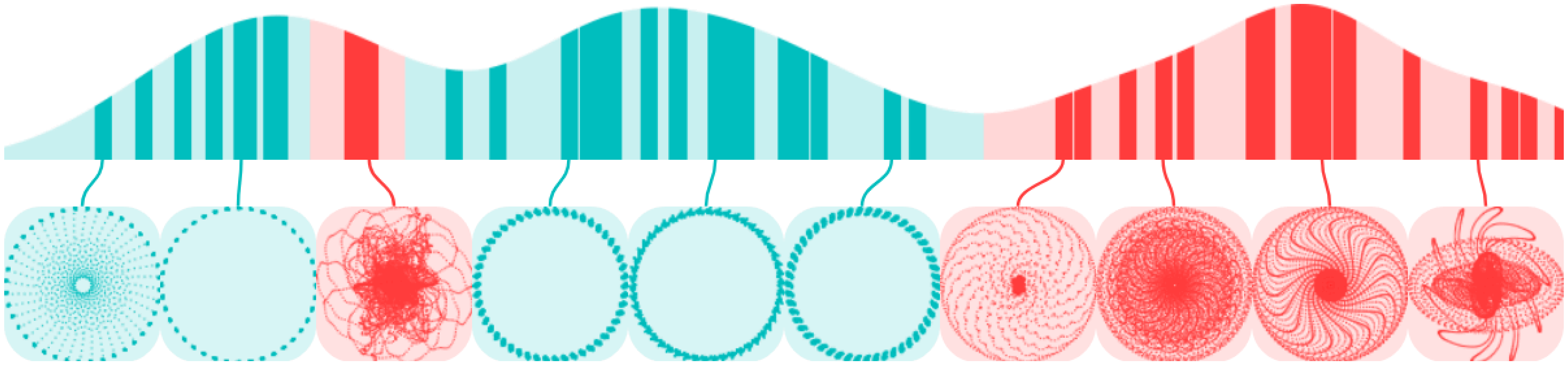
10 samples

Projection-Based Feature Descriptors



25 samples

Projection-Based Feature Descriptors

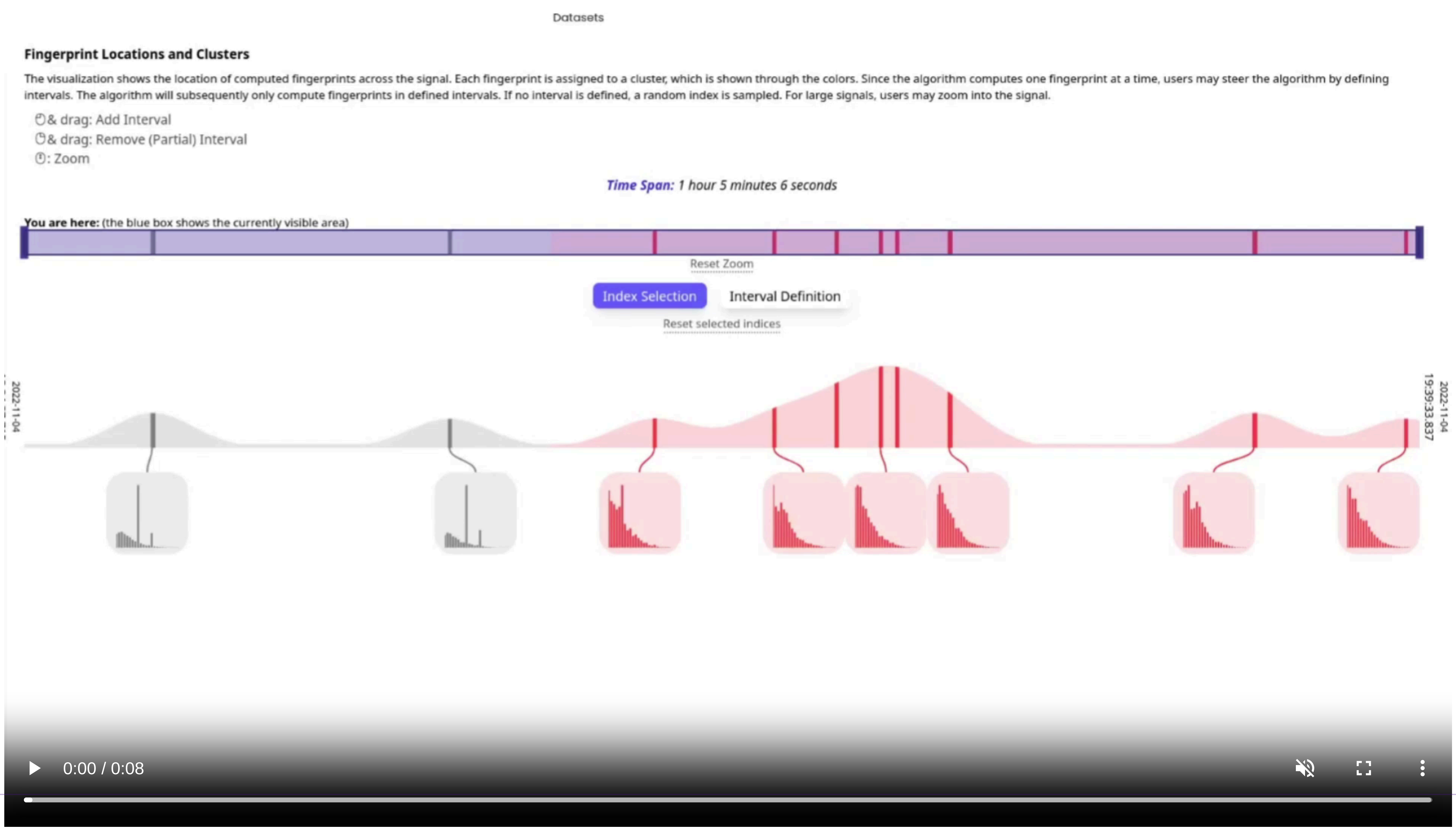


50 samples

How can users steer the sampling?

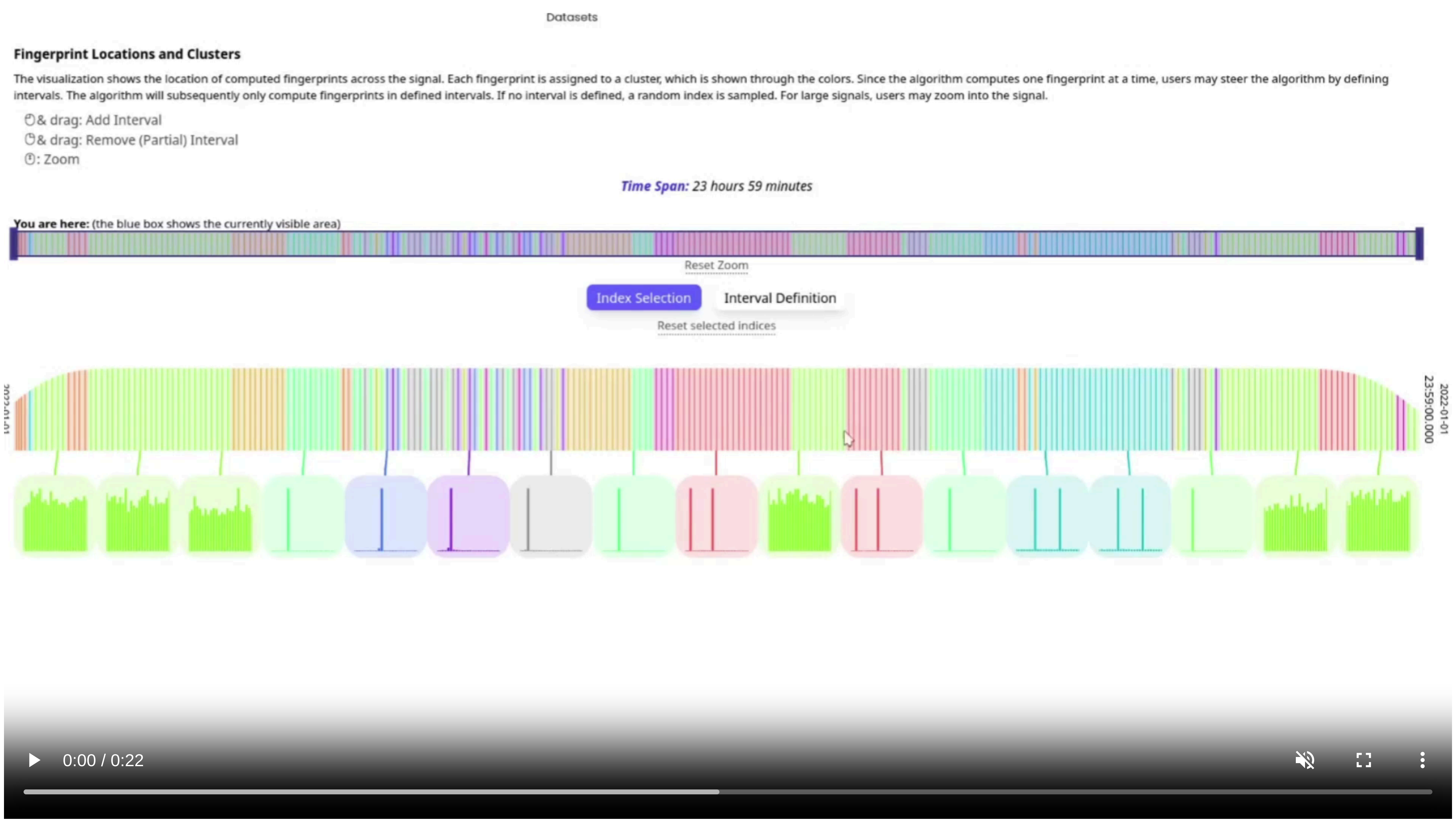
Defining intervals to constrain the sampling

(Accelerated video footage)

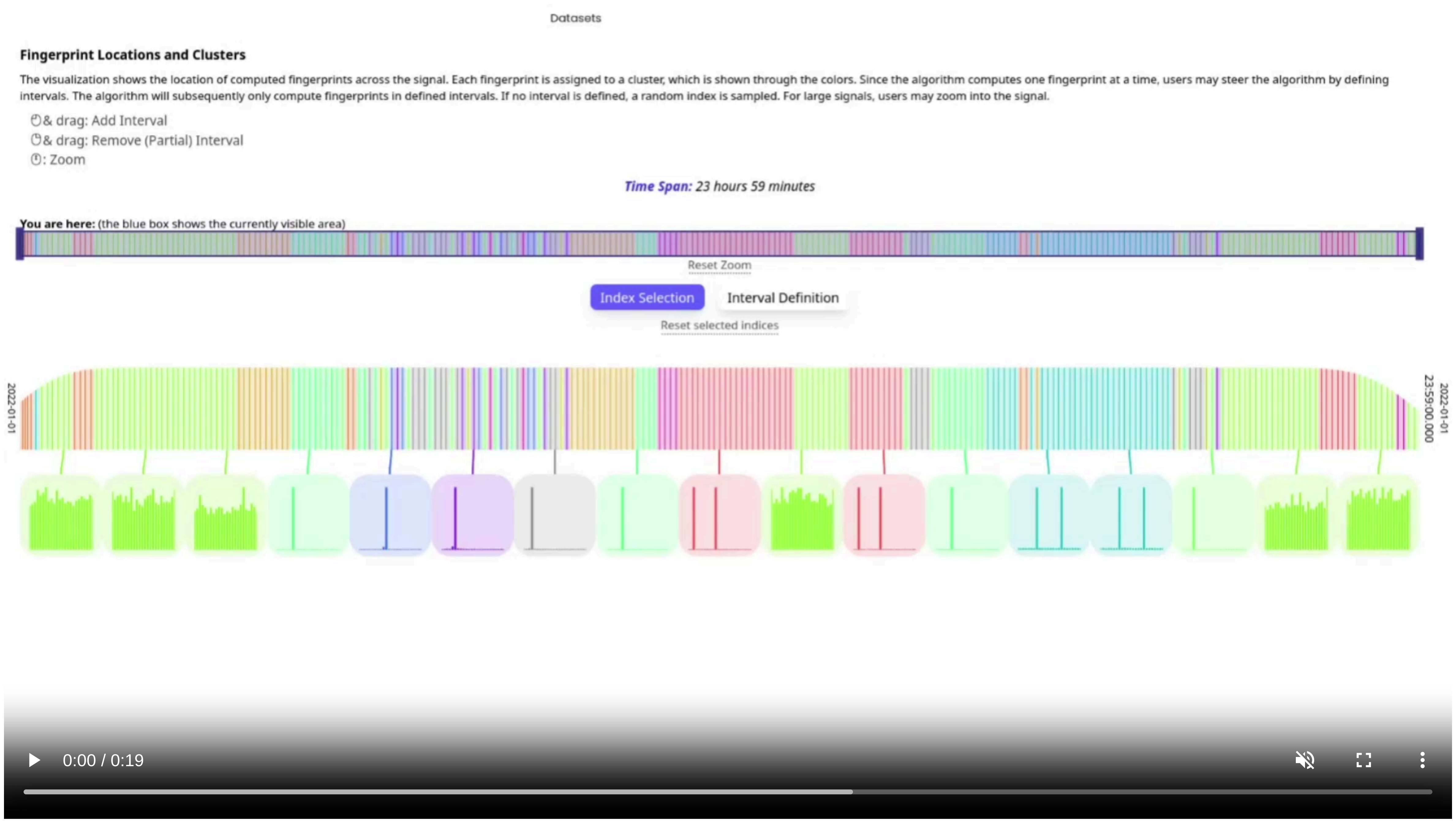


How can users explore the result?

Classic zooming and panning

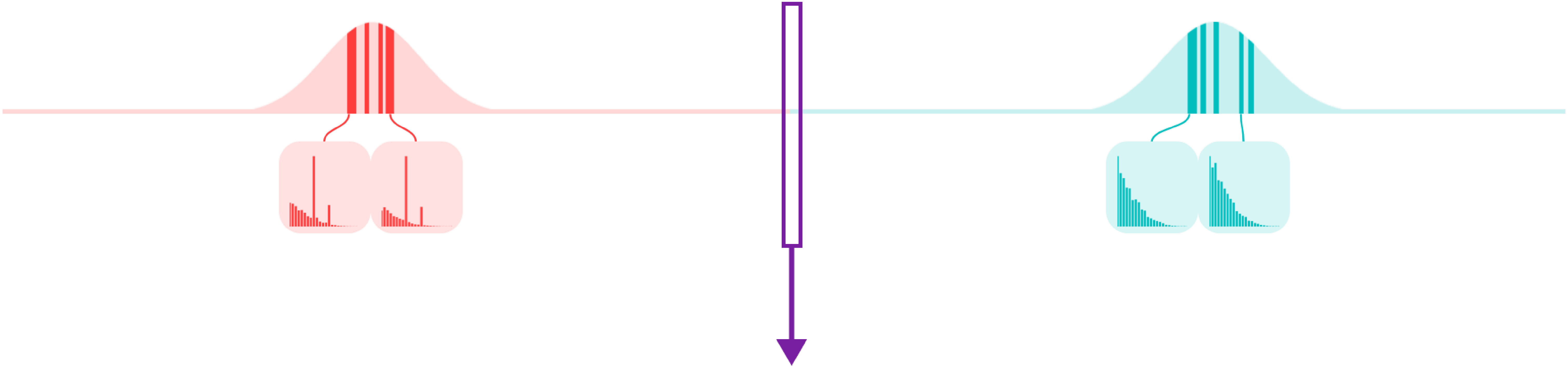


Inspecting samples in detail



Other sampling strategies

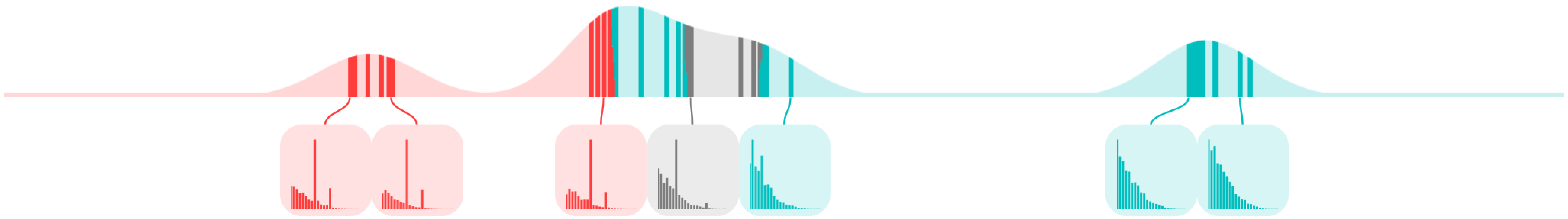
Binary Sampling (1)



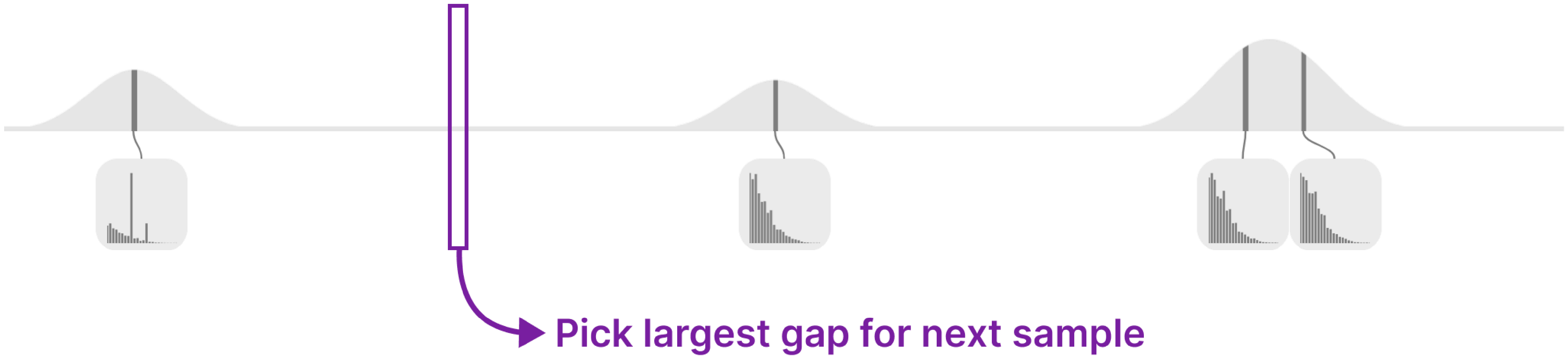
Sample between different segments

Binary Sampling (2)

The result: Sharp boundaries

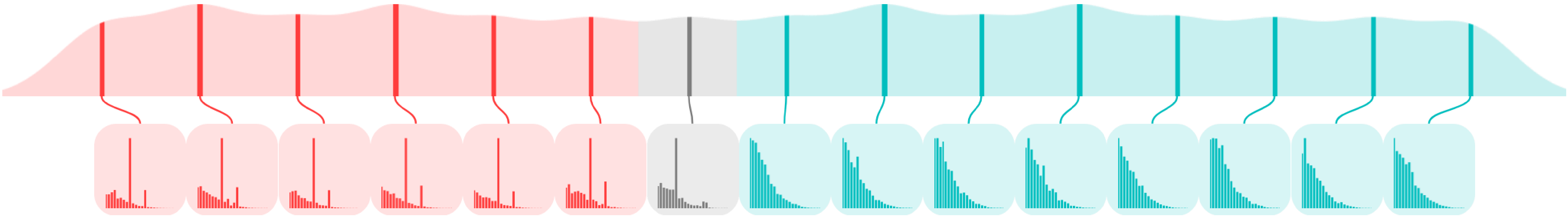


Gap Filling (1)



Gap Filling (2)

The result is a uniform distribution of samples



Linear Sampling

Pick samples in chronological order

(Accelerated video footage)

Formative Evaluation (1)

With two domain experts from the field of vibration analysis

Round 1 / 3

- Discusssion of initial idea.
- Formulation of requirements.

Round 2 / 3

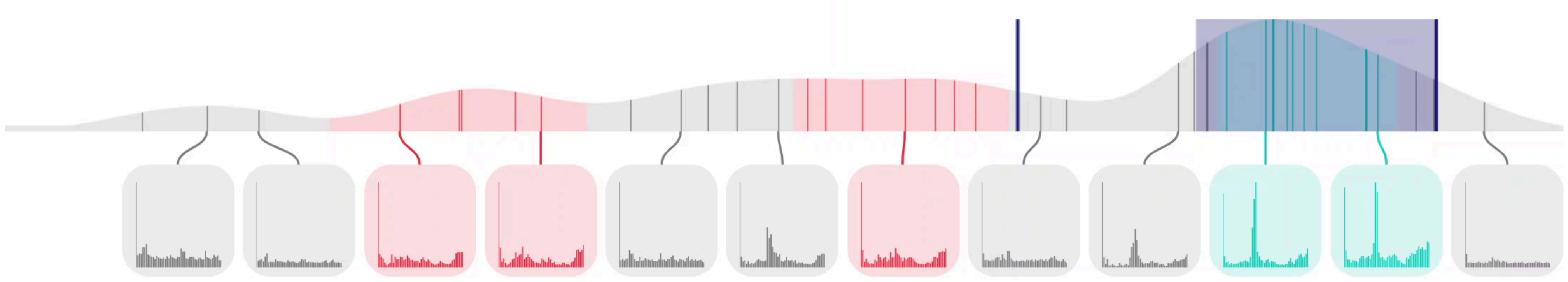
- Experts used the system to solve a task.
- The task was to identify one change point in a vibration signal.
- We identified several points for improvement.

Formative Evaluation (2)

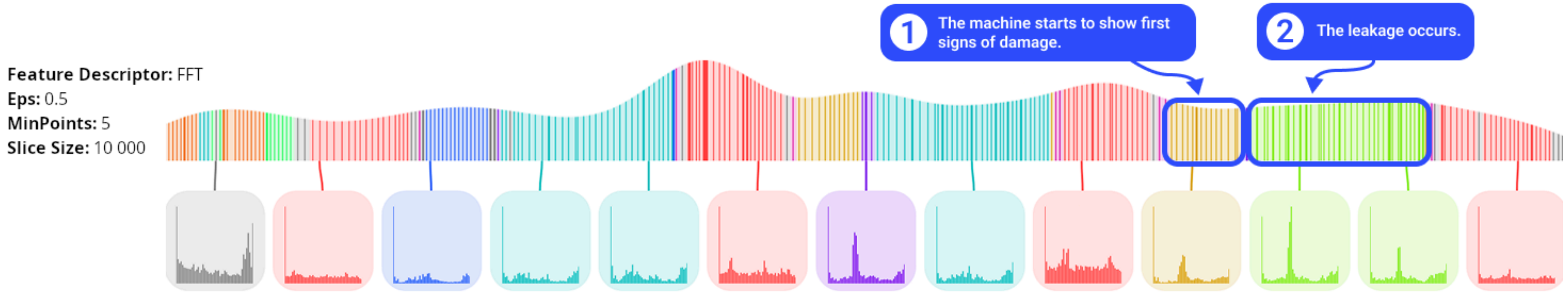
Round 3 / 3

- Experts used own dataset.
- Tried to replicate a finding.
- Successfully used system to identify leakage in a machine.

Early analysis results:

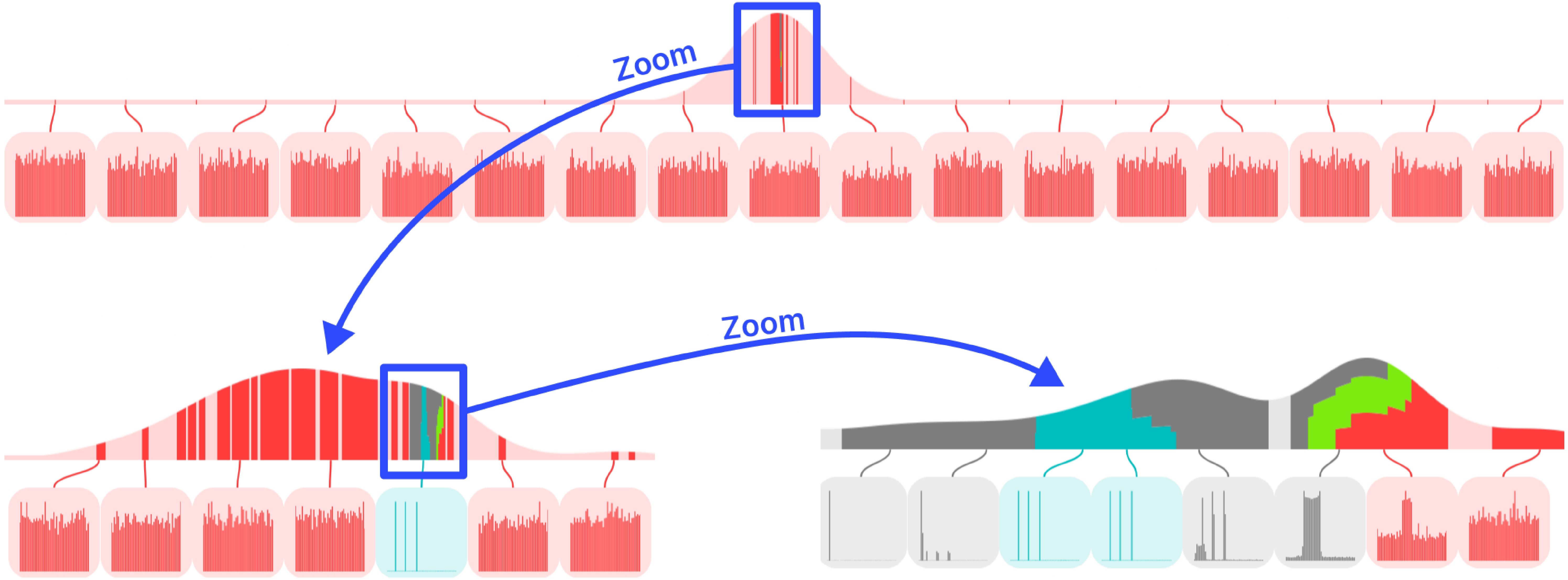


Final analysis results:



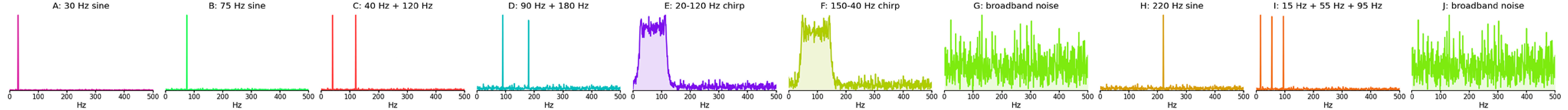
Where does it not work?

Small patterns (e.g. Anomalies) can be hard to detect



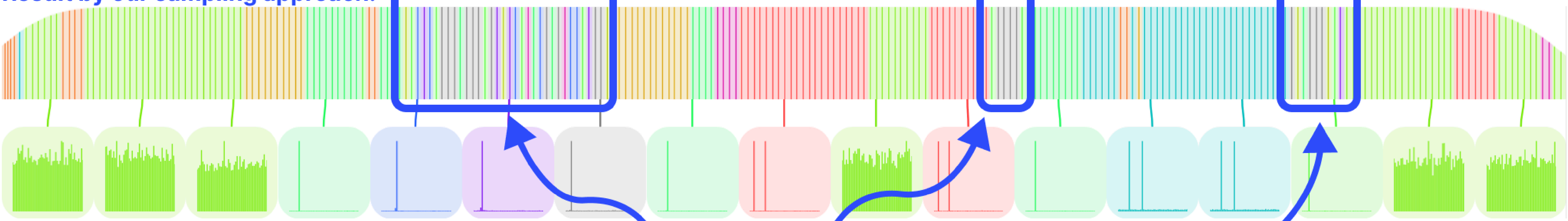
Chirp signals are challenging

Pattern Types (FFT Histograms):



Ground Truth:

Result by our sampling approach:



Feature Descriptor: FFT
Eps: 0.5
MinPoints: 5
Slice Size: 10 000

Chirp signals are particularly challenging

Future Work

- **Multi-resolution visualization** to reveal small, localized patterns
- **Uncertainty-aware segment boundaries** with gradual transitions
- **Broader user studies** to validate findings and segment discovery
- **Transient Visual Analytics** to remove redundant samples and reduce storage
- **Scale to longer signals** and larger analysis windows

Thank you!

Julian Rakuschek

Johanna Schmidt

Udo Schlegel

Michèle Posch

Jutta Isopp

Tobias Schreck

FFG Forschung wirkt.

PRESENT and **ENVIRON-HYDRO**

Open Source

Slides