NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of Physics, Astronomy and Informatics

Non-linear features for EEG based on recurrence analysis and short-term Fourier transform.

W. DUCH[1], Ł. FURMAN[1], L. MINATI[2], K. TOŁPA[3]

1 Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University, and Neurocognitive Laboratory, CMIT, 87-100 Toruń, Poland
2 Center for Mind/Brain Sciences (CIMeC), University of Trento, 38123 Trento, Italy Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan
3 Faculty of Philosophy and Social Sciences, Nicolaus Copernicus University, 87-100 Toruń, Poland

Introduction

Electroencephalography (EEG) records neural activity with high temporal resolution. This technique has now many applications, from medical diagnostics, therapeutic neurofeedback, to the brain-computer interfaces. Although many sophisticated techniques are applied for classification of such signals, results are difficult to interpret and far from optimal. Finding informative features in the complex, non-stationary multi-channel time-series is a daunting task. Two competing approaches are compared here.

Taken has shown that it is possible to recreate high-dimensional attractor dynamics using time-delayed embeddings of a single component of signal $u_i=u(t_i)$. This method creates vectors $x_i = [u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}]$ and requires estimation of optimal dimension m and time delays τ. It may be used to create binary recurrence matrix $\mathbf{R}_{i,j} = \Theta(\varepsilon-\|x_i-x_j\|)$ that indicates whether two states of the system are closer to each other than the similarity threshold ε. Recurrence plots are used to visualize this matrix, and Recurrence Quantification Analysis (RQA) provides non-linear features of EEG signals characterizing dynamics of the system. RQA calculates values of features **FR** from the **R** matrix that estimate recurrence rate (RR), trapping time (TT) in meta-stable states, laminarity of trajectories (LAM), and other characteristics of neural dynamics.

EEG signal is a complex mixture of waves with frequencies going to the gamma range, in practice usually cut at 50 Hz. An alternative way to generate RQA features is to create another recurrence matrix from vectors that represent power spectra. Defining Hamming time window short-time Fourier transform (STFT) is used to calculate power spectra for this window. Shifting these windows generates vectors y_i = STFT($[u_i,u_i']$) that can be used to create recurrence matrix $\mathbf{S}_{i,j} = \Theta(\varepsilon'-\|y_i-y_j\|)$. This representation of the signal shows peaks of power at certain frequencies. RQA on this matrix provides another set of features **FS**. There are two pipelines here, starting with the time-series signals for each EEG channel: **U⇒X⇒R⇒FR**, and **U⇒Y⇒S⇒FS**. The purpose of our work is to compare features **FS** with the features **FR**, and show which are easier to interpret and are more useful in extracting information from the EEG signals.

Methods

We have used the data from BCI2000 platform, recorded from 64 channel EEG for 109 people, with the sampling rate of 160 Hz. Two one-minute baseline runs were extracted, one with eyes open, and one with eyes closed. 31 seconds of signal from both baselines were left, each containing 5000 samples per channel. To calculate recurrence matrices **R** for each of the 64 channels we have experimented with different embedding dimensions, time delays, and similarity thresholds. Using a combination of these 3 parameters we have calculated sets **FR** of non-linear RQA features, including TT, RR, LAM, entropy of diagonal lines Lentr, determinism DET, average diagonal line length Ldiag.

In the second approach Hamming windows were used with 240 samples sliding window by a single sample to cover all data samples. For each window STFT spectrum is calculated, covering frequencies from 0-50 Hz. A vector with 257 points represents this spectrum with 1/3 Hz resolution. These vectors were used to create **S** recurrence matrices, and **FS** sets of features.

With these two sets of features we could compare their histograms, try to interpret them, and use these features to classify which segments of the EEG resting state signal corresponds to the eyes opened or eyes closed condition.

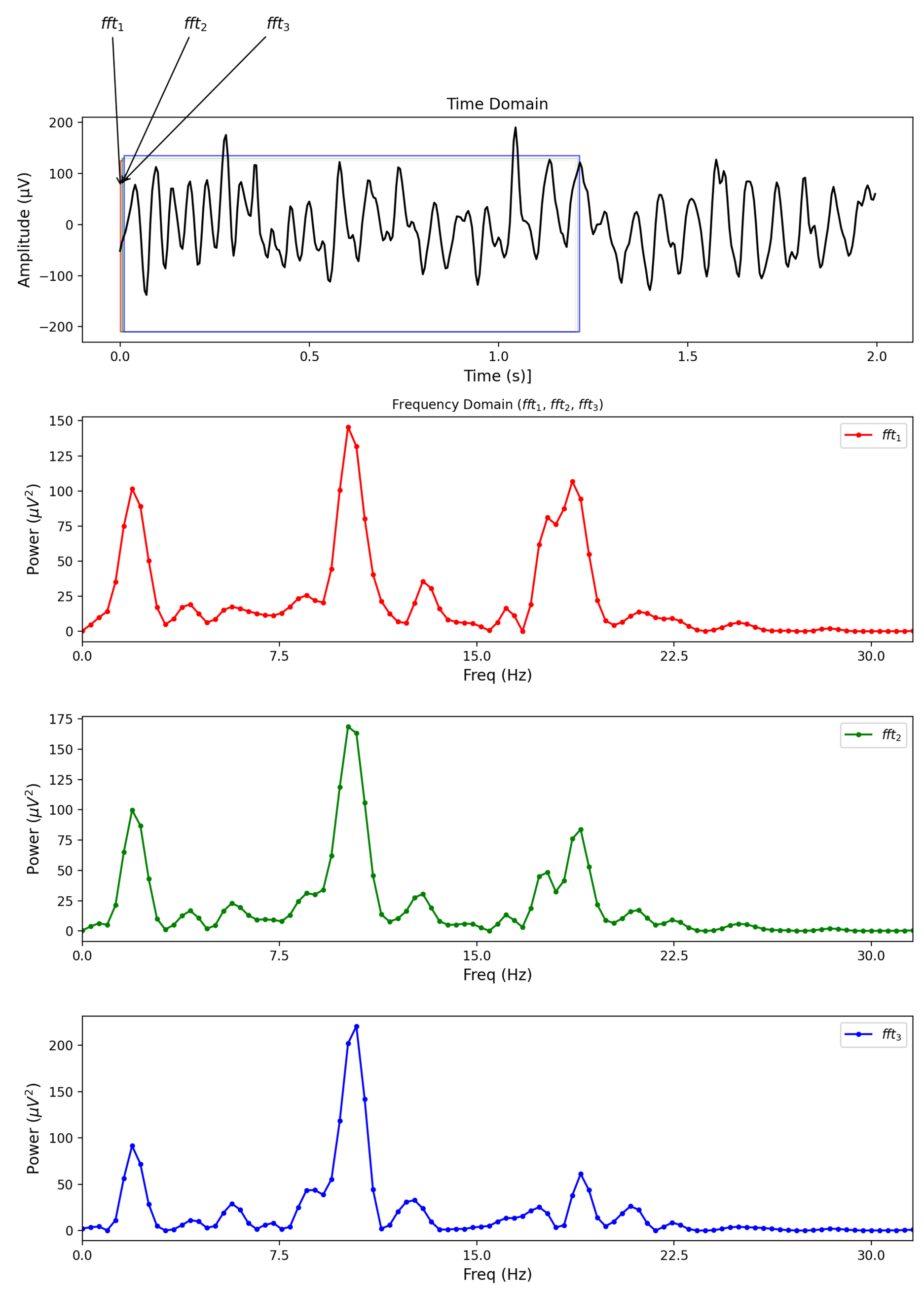
Results

Histograms of the six individual **FR** features show high overlap of values for the two conditions. Histograms of the **FS** features show less overlap and are easier to interpret. Trapping times and recurrence rates are longer in the eyes closed case than in the eyes open (more stable alpha waves). **FR** visualization using UMAP shows better separation than visualization of **FS**. Linear SVM calculations based on the **FR** and **FS** features was used to estimate direction of the best hyperplane that can divide this data, showing greater separation in **FS** case. The difference of classification accuracy was quite significant. For the best combination of parameters representation based on embedding reaches about 61% of overall accuracy, while STFT about 90%.

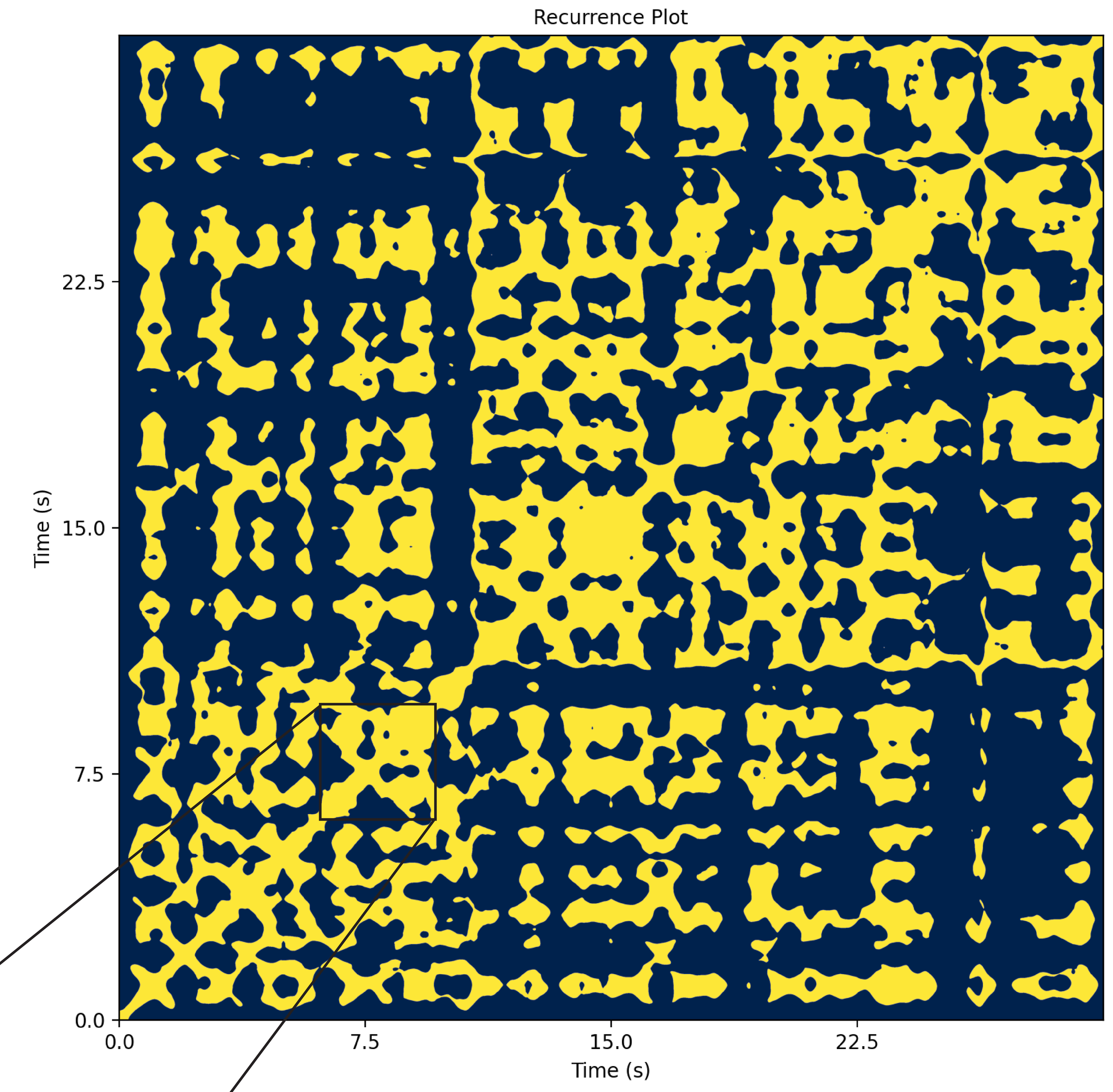
References

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. Physics Reports, 438(5–6), 237–329.

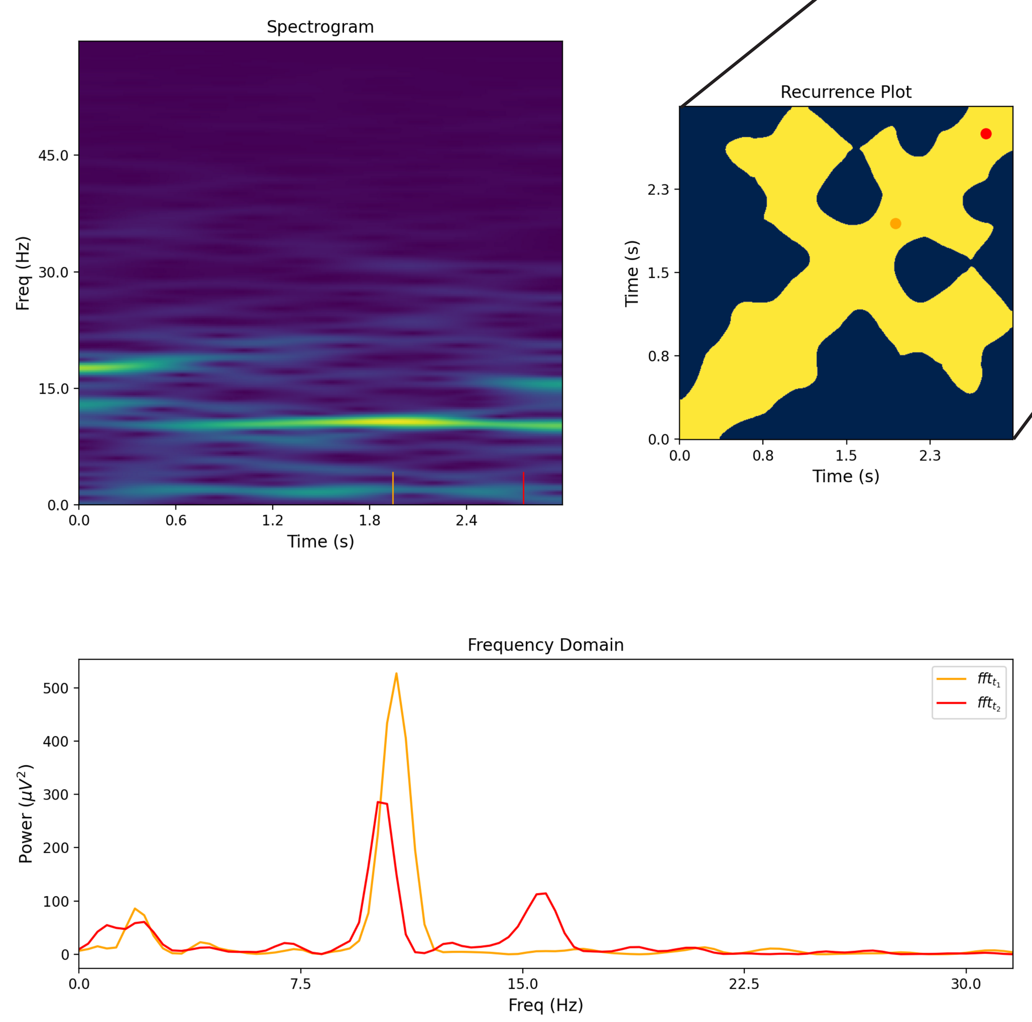
Marwan, N., Schinkel, S., & Kurths, J. (2013). Recurrence plots 25 years later - Gaining confidence in dynamical transitions. Europhysics Letters, 101(2), 20007.



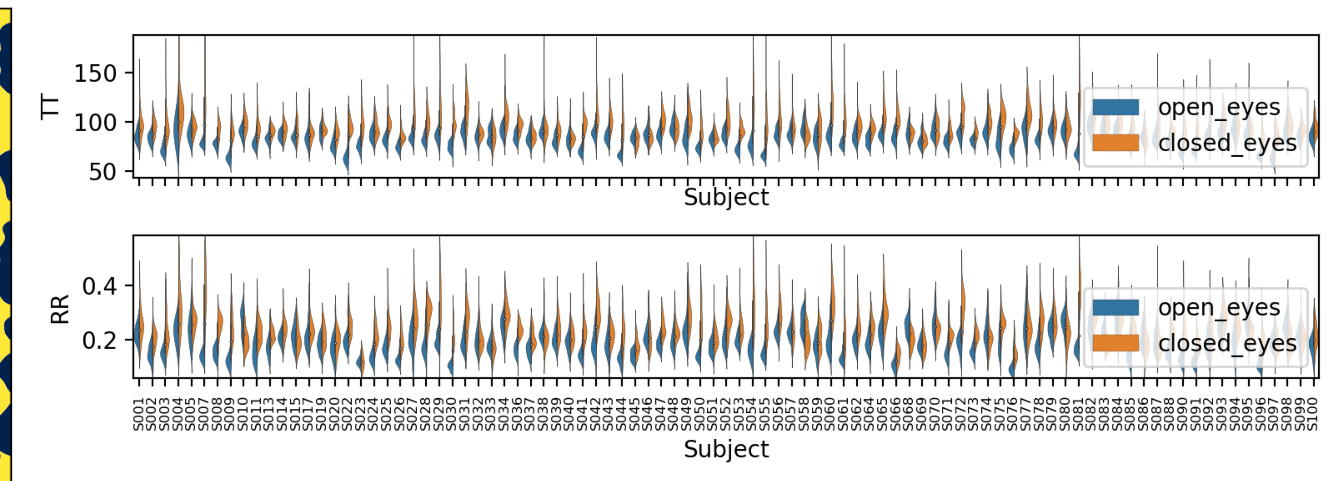
Example of STFT vectors stored in the **Y** matrix, showing changes of power spectra after 100 ms and 200 ms, measured by the O1 electrode.



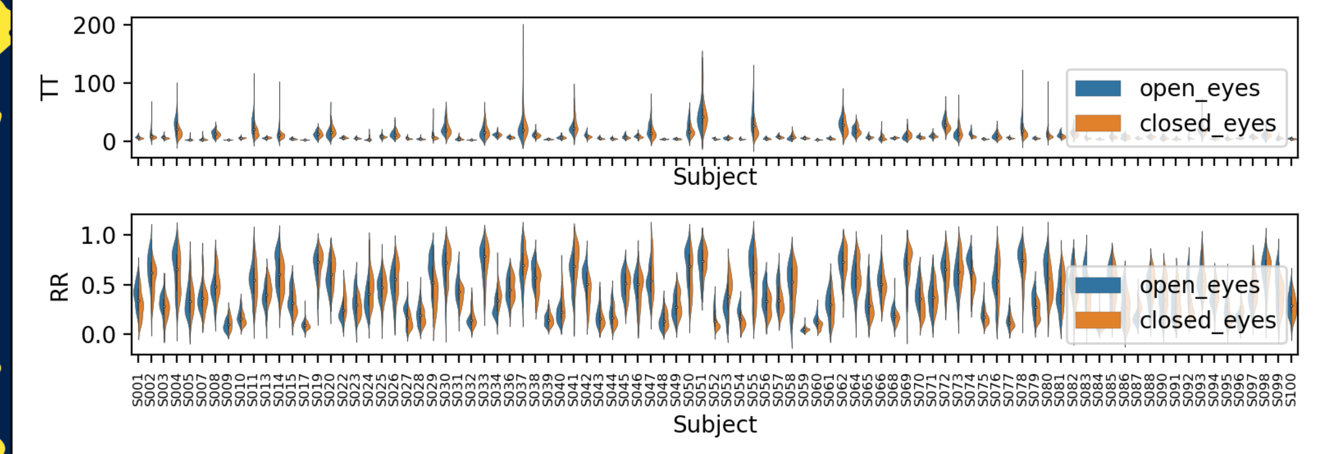
Example of **FS** matrix, STFT vectors for window size = 240 samples, eyes open condition, ε = 4.25, 31 seconds, electrode O1, subject S001. Dark dots show distances outside the ε neighborhood.



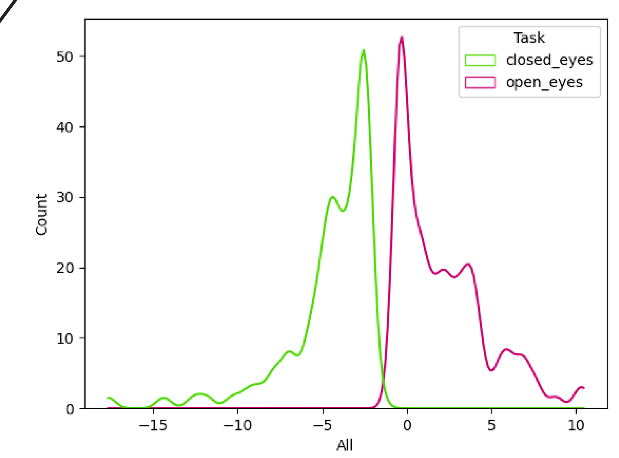
Example of a spectrogram that contains STFT vectors for each point in time, showing periods of rapid changes and relative stability, when the system is trapped in a meta-stable state. Red and orange color dots/markers correspond to the two time points shown in the spectrogram plot and two power distributions.



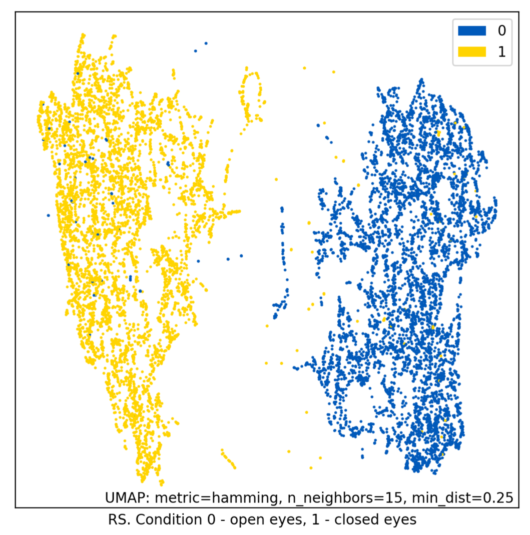
Examples of histograms of the two **FS** feature values (from the STFT vectors, window size=240) extracted from the RQA analysis (other features have similar histograms), showing for all subjects distribution of values for 64 electrodes, with recurrence threshold ε=4.25.



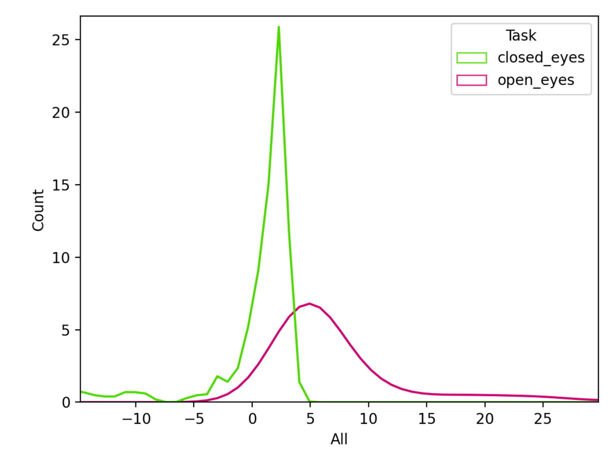
Examples of histograms of the two **FR** feature values (X vectors for m=2, **Δ**t=9) extracted from the RQA analysis (other features have similar histograms), showing for each subject distribution of values for 64 electrodes, with recurrence threshold ε=0.035.



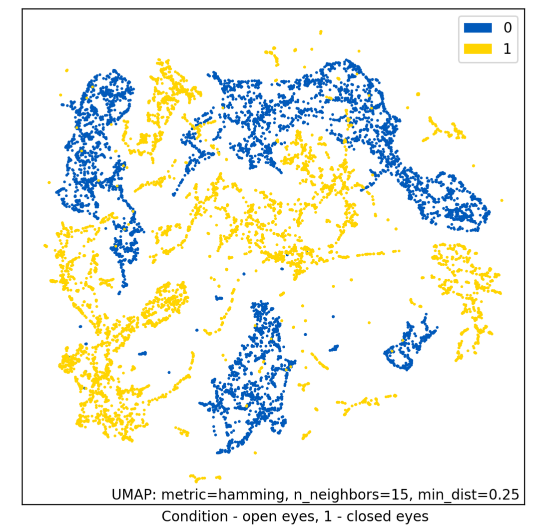
Histograms of the projection of 384 **FS** feature values (6 RQA features for 64 electrodes), for all subjects, in the direction perpendicular to the LSVM hyperplane, for all data.



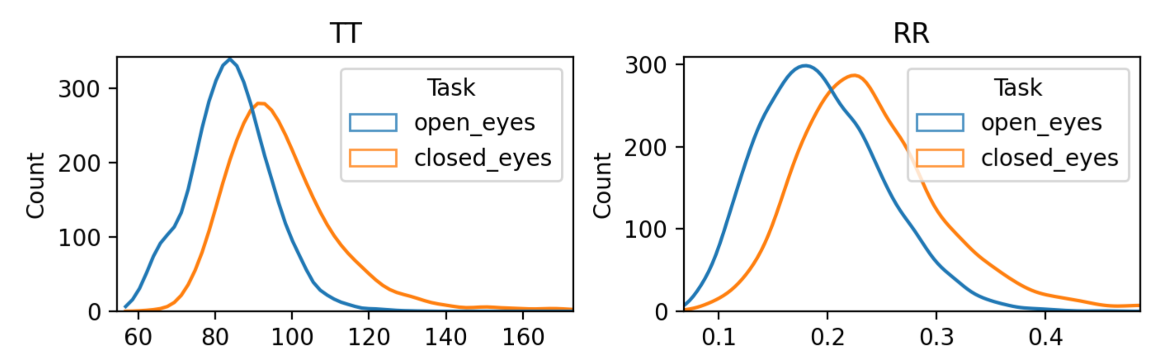
UMAP Visualization of 6 **FS** features for 64 electrodes and 90 subjects.



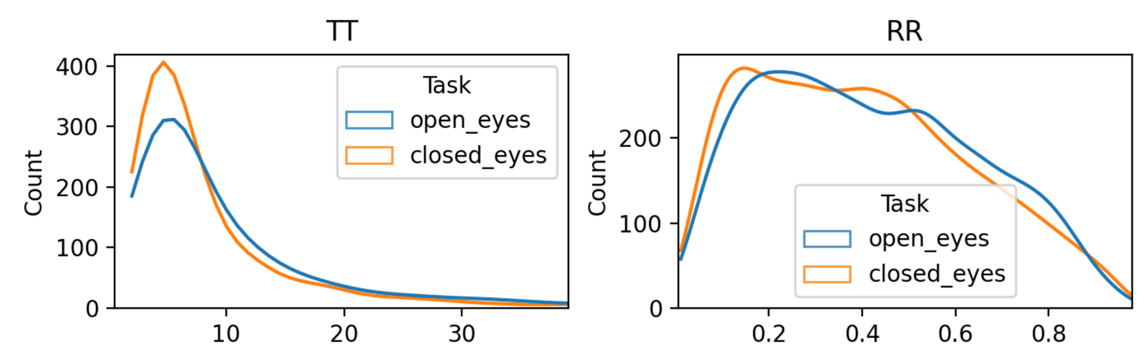
Histograms of the projection of 384 **FR** feature values (6 RQA features for 64 electrodes), for all subjects, in the direction perpendicular to the LSVM hyperplane, for all data.



UMAP Visualization of 6 **FR** features for 64 electrodes and 90 subjects.



TT and RR histograms for the **FS** features, for all subjects.



TT and RR histograms for the **FR** features, for all subjects.