EEG signal segmentation for assessing the time-course of brain response to stimuli

M Lotka[1,2], JK Ochab[2]
marta.lotka@doctoral.uj.edu.pl

[1]Doctoral School of Exact and Natural Sciences, Jagiellonian University, Kraków, Poland
[2]Institute of Theoretical Physics, Jagiellonian University, Kraków, Poland

Abstract

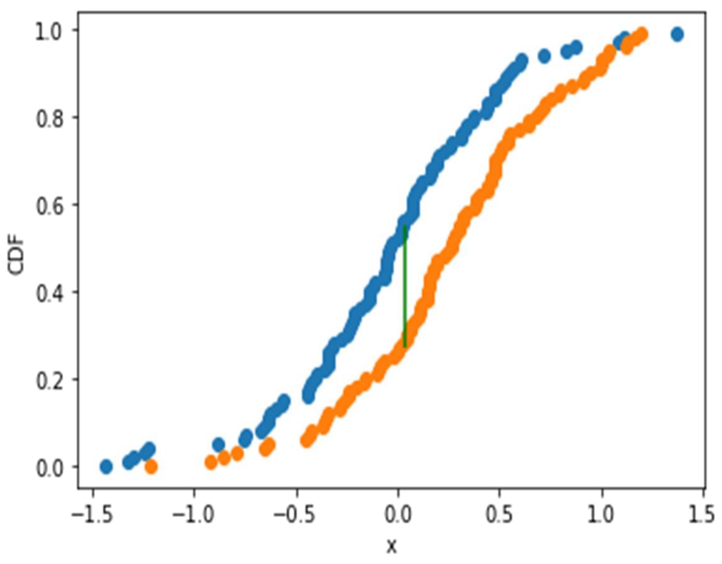
Signal detected by electroencephalography (EEG) exhibits a power spectrum with a predominant 1/f component. As such, the signal is nonstationary. When EEG is applied to the study of cortical response to stimuli, the event-related potential (ERP) technique is commonly used. It led to innumerable insights into the mechanisms of cognition. However, it has significant limitations. We propose an alternative technique utilizing an algorithm introduced by [1], which recursively divides the series into segments based on maximizing the Kolmogorov-Smirnov (KS) distance between them. The method's performance was tested on standardized EEG recordings included in the ERP CORE dataset [2] and compared with the ERP technique.

Segmentation algorithm: from non-stationary time series into quasi-stationary spells

As introduced by [1]:

1. A pointer is moved along a segment of a time series to find the position that maximizes the Kolmogorov-Smirnov statistic.
2. After this position of the pointer is found, a numerically determined statistical significance criterion is applied.
3. If the statistical significance criterion is fulfilled one verifies whether the segments are at least of a fixed minimal size. If both conditions are fulfilled, the cut is accepted.

The entire procedure is then performed recursively (until there are no segments left to be further divided).



The Kolmogorov-Smirnov test for simulated Gaussian data. The Kolmogorov-Smirnov statistic is indicated by the green line.

EEG dataset and preprocessing steps

The ERP CORE dataset [2]: 40 subjects performing the flanker task. Data downsampled to 1024 Hz and re-referenced to the average of near-mastoid electrodes. Further preprocessing steps:

1. *HPFILT* - DC offset removal, high-pass filtering (non-causal Butterworth impulse response function, half-amplitude cut-off at 0.1 Hz, 12 dB/oct roll-off).
2. *REMICA* - remove ICA compontent(s) corresponding to ocular artifacts (identified by visual inspection).
3. *EPOCH* - epoching from -600 to 400 ms with respect to the response, with baseline correction (baseline -400 to -200 ms).
4. *REMART* – automated removal of epochs containing artifacts(voltage threshold algorithm, moving window peak-to-peak algorithm, step-like algorithm for removal of horizontal eye movements).

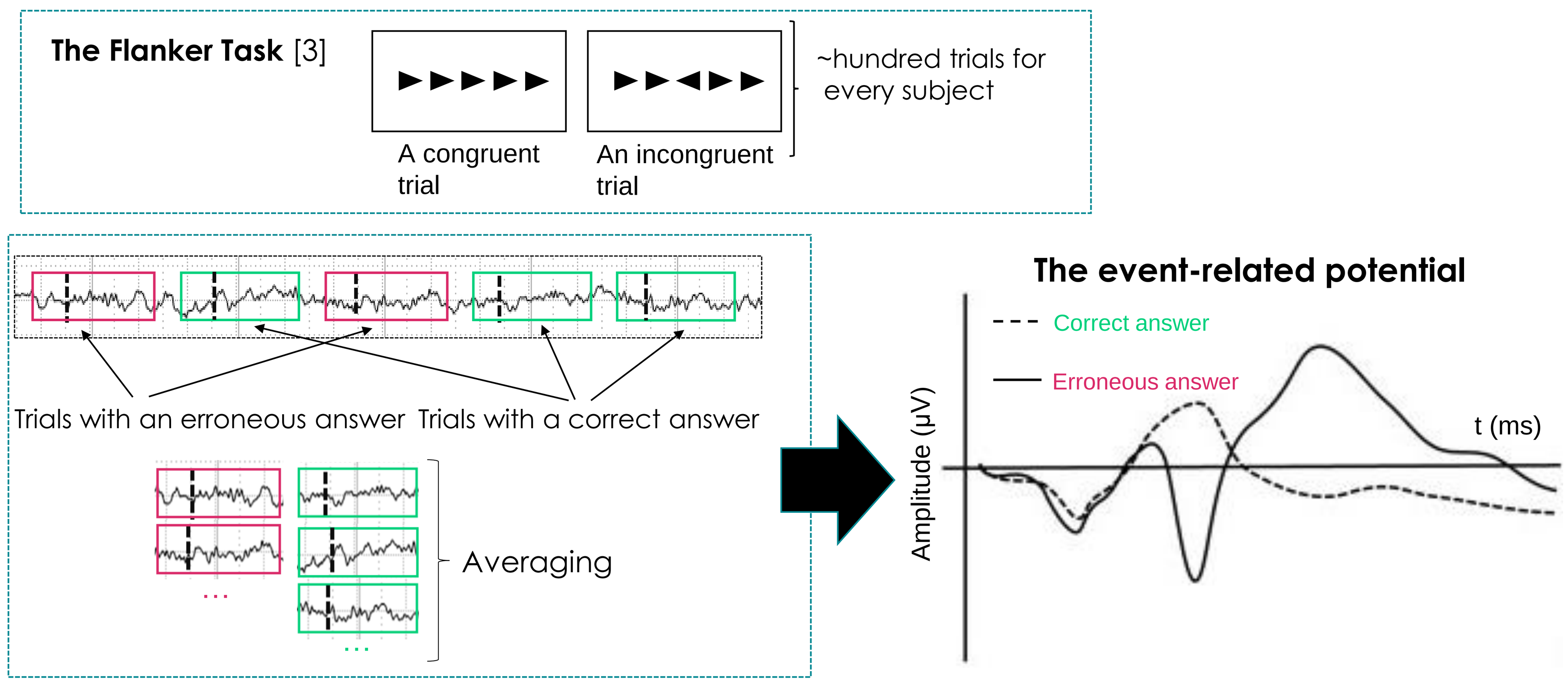
Different preprocessing pipelines were tested:

Label	I	II	III	IV	V
HPFILT	No	Yes			
REMICA	No	No		Yes	
EPOCH	Yes				
REMART	No	No	Yes	No	Yes

Kernel density estimation

Probability density of a KS segmentation algorithm partition was estimated for each subject and electrode separately by kernel density estimation with a Gaussian kernel. The bandwidth was chosen by leave-one-out cross-validation (log-likelihood maximization).

The event-related potential technique



Normalized mutual information

The partitions produced by the KS segmentation algortithm were compared using normalized mutual information $I(A,B)$ [4,5]. Let us assume two partitions A and B of a series X: into n_a and n_b segments.

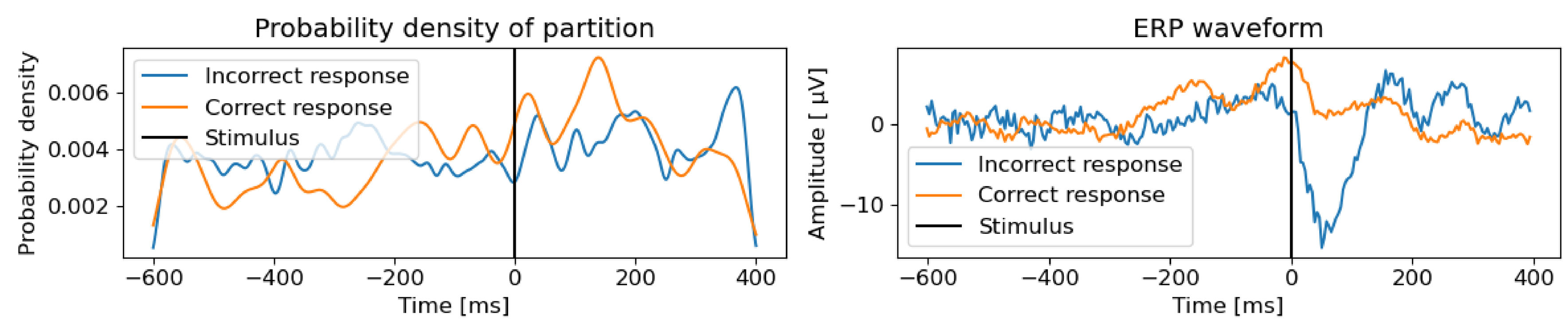
$$I(A,B) = \frac{-2\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} N_{ij}\log\left(\frac{N_{ij}N}{N_i N_j}\right)}{\sum_{i=1}^{n_a} N_i log\left(\frac{N_i}{N}\right) + \sum_{j=1}^{n_b} N_j log\left(\frac{N_j}{N}\right)}$$

where N_{ij}- the numer of elements in the ith segment of A present in the jth segment of B. N_i and N_j are sums of N_{ij} over, respectively, j and i.

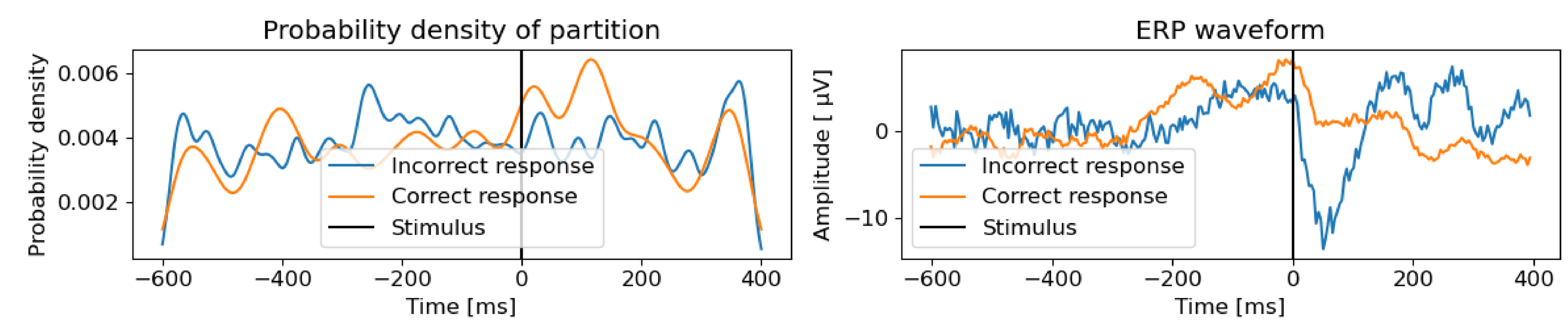
Kernel density estimations vs ERP waveforms

For one subject and one electrode (FCz).

No high-pass filtering, ICA, artifact removal (I)



High-pass filtering, ICA, artifact removal (V)



Normalized mutual information – comparison of preprocessing pipelines

	I	II	III	IV	V
I	1	0.85	0.85	0.72	0.72
II	0.85	1	1	0.75	0.75
III	0.85	1	1	0.75	0.75
IV	0.72	0.75	0.75	1	1
V	0.72	0.75	0.75	1	1

Average normalized mutual information between partitions of epochs for one subject and electrode (FCz).

Bibliography

[1] Camargo, S., Queirós, S., & Anteneodo, C. (2011). Nonparametric segmentation of nonstationary time series. *Physical review. E*, Statistical, nonlinear, and soft matter physics. 84. 046702.
[2] Kappenman, E. S., Farrens, J. L., Zhang, W., Stewart, A. X., & Luck, S. J. (2021). ERP CORE: An open resource for human event-related potential research. *NeuroImage*, 225, 117465.
[3] Eriksen, B.A., Eriksen, C.W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* **16**, 143–149 (1974).
[4] Ana, L. N. F. and Jain, A. K. Robust data clustering. volume 2, pages II–II, 2003.
[5] Kuncheva, L.I and Hadjitodorov, S. T. Using diversity in cluster ensembles. In 2004 IEEE International Conference on Systems, Man and Cybernetics, volume 2, pages 1214–1219, The Hague, Netherlands, 2004. IEEE