Artifact Removal in EEG Data

Maie LE GUILLY - Rasmus HAARSLEV - Troels THOMSEN

1 Purpose

Our purpose is to remove noise from EEG data, in order to allow for real-time signal analysis. The objective is to reproduce results from a method using principal component analysis (PCA) that has been successful in previous studies [1]. The EEG is typically noised by artifacts produced by patient's muscular movements such as eye blinking or head movement.

2 Method

During real-time signal analysis data will be treated in small time windows. We slice our dataset into similar small windows, in order to simulate this behavior.

Given a window of multi-channel EEG input, we subtract the means of each channel in order to normalize. We calculate the eigenvectors and eigenvalues from the covariance of the window. We keep the eigenvectors, whose eigenvalues are below a certain threshold and discard the ones above. Components with eigenvalues above the threshold are likely to represent artifacts, since artifacts will increase variance significantly. We project the window onto the remaining eigenvectors and then back into the original channel space. Lastly we add the subtracted means.

The eigenvalue threshold is chosen from a calibration period on the input source, where we assume all windows of data to be artifact free. We calculate the eigenvectors and eigenvalues of the artifact-free windows, and based on these eigenvalues we can calculate the following three different eigenvalue thresholds:

- The maximum eigenvalue of all the calibration windows. We expect to only reject components representing artifacts, since their variance is assumed to be above the variances found during calibration.
- The average eigenvalue of all the calibration windows. We expect to reject all components representing artifacts, with the risk of also rejecting some components not representing artifacts.
- The average of the largest eigenvalue in each calibration window. We expect similar behavior to that of the maximum eigenvalue threshold, with the advantage of evening out accidental high-variance during calibration and the disadvantage of the risk of rejecting some components not representing artifacts.

Since we assume calibration windows to be free of artifacts, we also assume their variance to be significantly smaller than windows with artifacts.

3 Results

The following experiments are done on an EEG dataset recorded using an *Emotiv* [2] device. It records on a total of 14 channels. The initial dataset is assumed to be cleaned, e.g. free of artifacts.

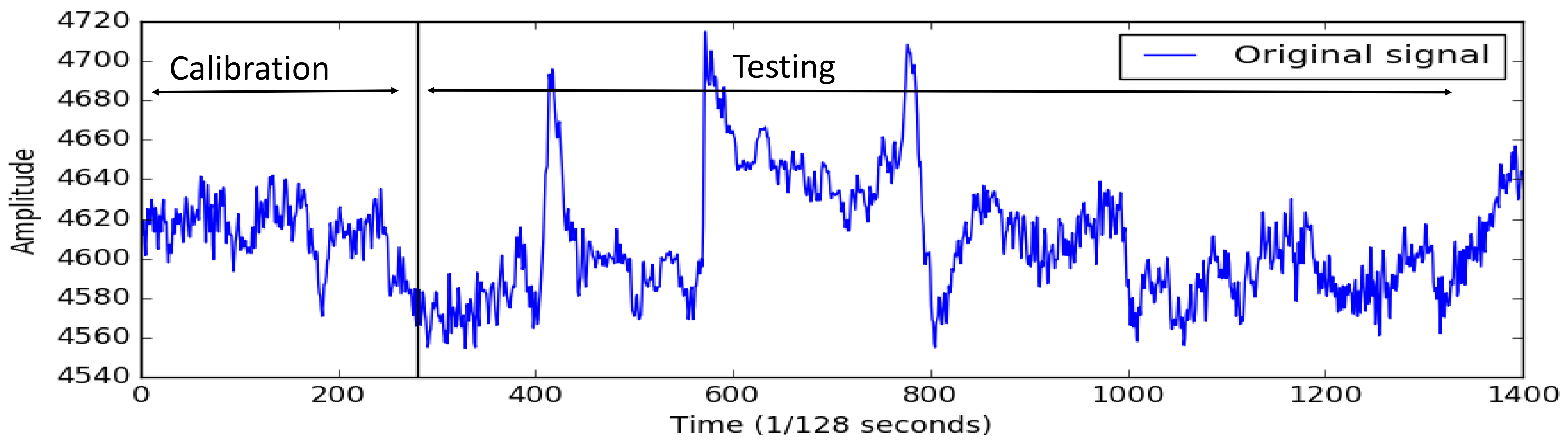

Figure 1: The 14th channel of the original dataset is represented, among the 14 recorded one. It contains 1400 samples. 20% of the dataset has been allocated for the calibration, the remaining is used as testing set.

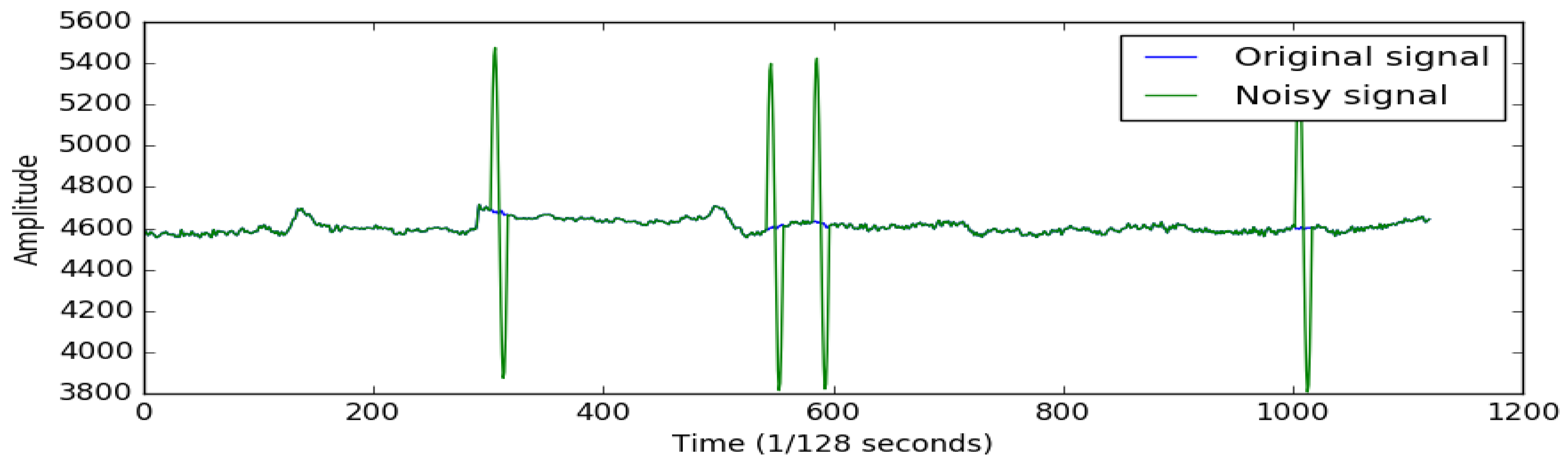

Figure 2: Artifacts are simulated using a sinusoidal function parametrized by an amplitude and a period, and randomly added to the testing set. Each channel is affected differently, the amplitude of the artifact being smaller when the channel's sensor is far from the source of the artifact. This one is the most affected channel (same channel as figure 1).

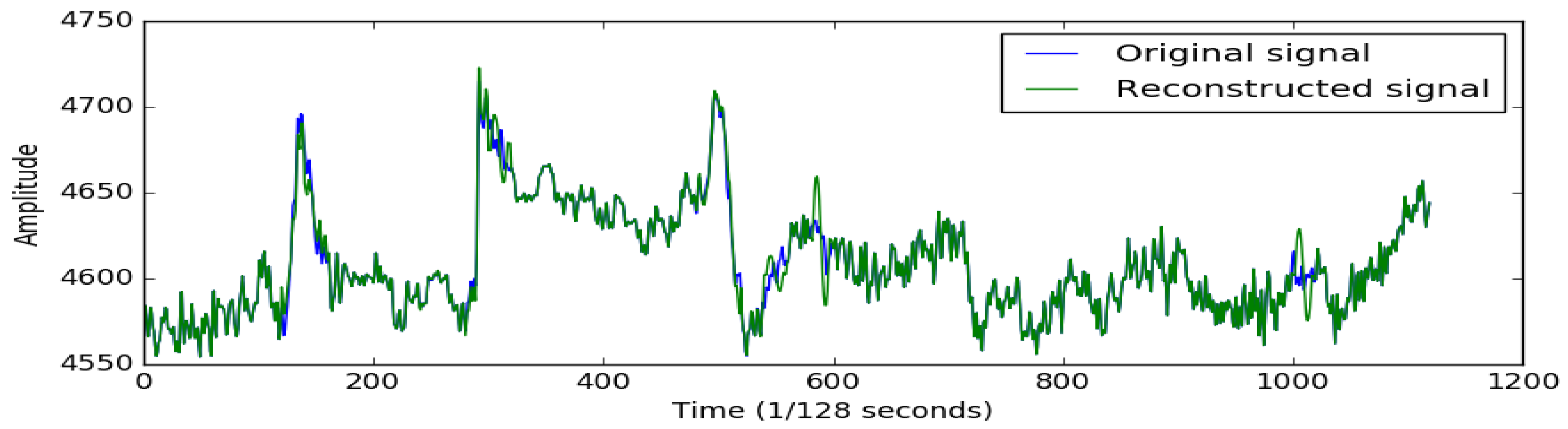

Figure 3: The same channel as figure 2, the original signal is represented together with the reconstructed signal. The maximum eigenvalue threshold has been used for PCA.

In order to find the optimal window size, we cross-validate the window size and the threshold with mean square error.

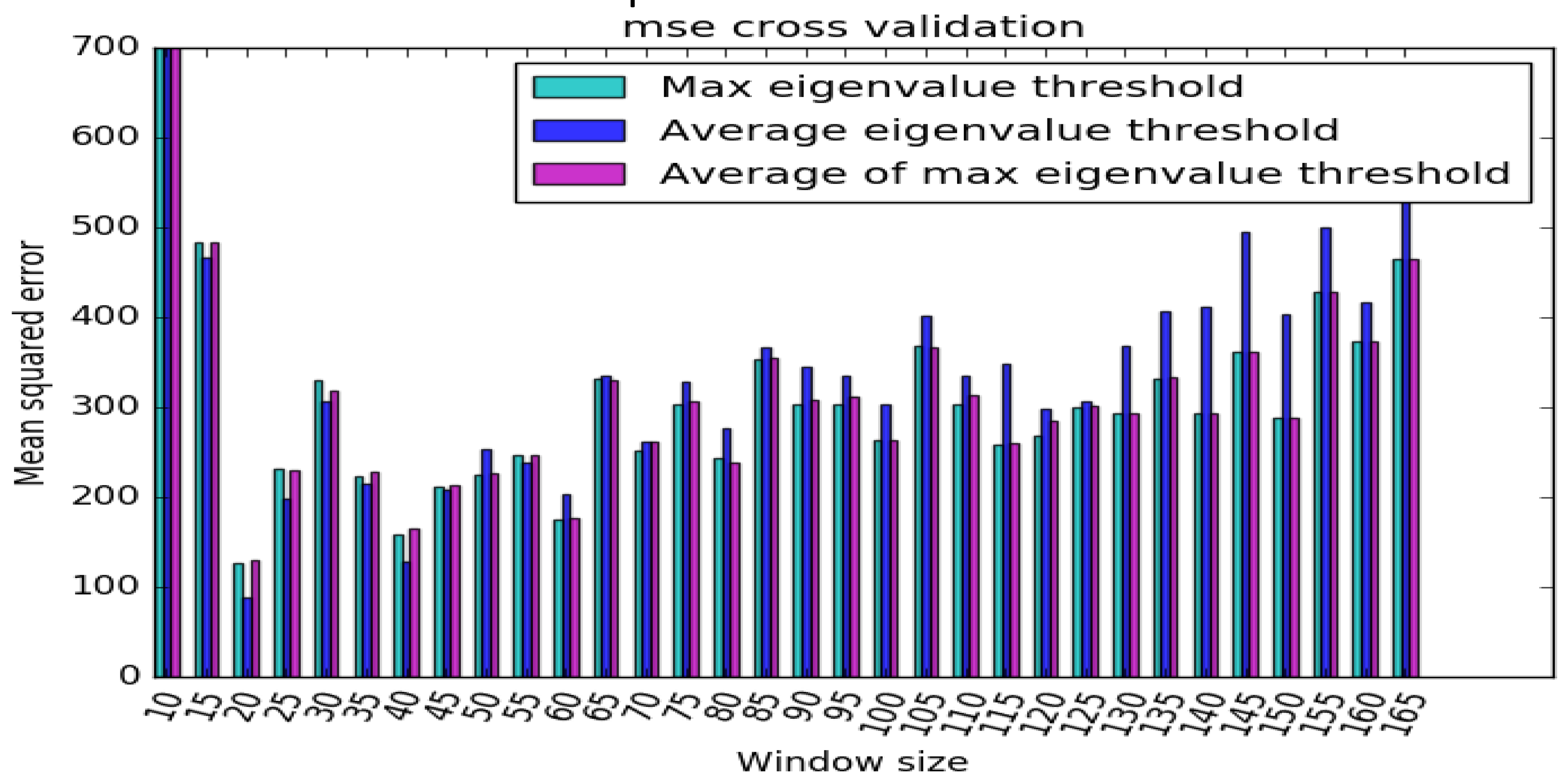

Figure 4: cross-validation between thresholds and window size on a 1400 sample dataset. A window size of 20 with MAX threshold is optimal.

For a window with no artifacts we expect no component to be rejected, such that the reconstructed and the original window are identical. For a window with an artifact, we expect one or more components to be rejected.

Threshold	Max	AVG	MAX_AVG
Sensitivity	1.0	1.0	1.0
Specificity	0.82	0.0	0.78

Figure 5: Sensitivity and specificity for the different threshold and window size 20

For the same window size we can remove artifacts at a rate around 200 Hz allowing real time artifact removal

4 Conclusion

We are able to remove the simulated artifacts. We found as expected that the maximum eigenvalue threshold gives the best results, while the average eigenvalue threshold gives the worst results . By our estimate the timing measure indicate that our method would be fast enough for real time analysis.

[1] Mullen, T.R., et al 2015. Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *Biomedical Engineering, IEEE Transactions on*, *62*(11), pp.2553-2567

[2] https://github.com/codehunks/emotiv/tree/master/csv-eeg-data

[3] Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T.O. and Makeig, S., 2010. MATLAB-based tools for BCI research. In *Brain-Computer Interfaces* (pp. 241-259). Springer London.

[4] Hsu, S.H., Mullen, T., Jung, T.P. and Cauwenberghs, G., 2015. Real-time Adaptive EEG Source Separation using Online Recursive Independent Component Analysis.