Institute for Cognitive Systems
Department of Electrical and Computer Engineering
Technical University of Munich

Comparative Signal Quality Analysis of EEG Recording Systems

Engineering Practice Report

Karahan Yilmazer ✉

Department of Electrical and Computer Engineering, Technical University of Munich (TUM), Arcisstr. 21, 80333 Munich, Germany

✉ karahan.yilmazer@tum.de

July 26, 2021

1 Introduction

There are metrics researchers use to benchmark electroencephalography (EEG) signal qualities but none of them are standardized or widely accepted, "there currently are no community-accepted or standard metrics for EEG signal quality as observed within real-world domains" [1]. One study even mentions a medical technical assistant visually inspecting the EEGs of subjects using a "skill-based state-of-the-art procedure"[2] as a reliable way of assessing signal qualities. Although visual inspection is the most widely used technique for this reason, these statements show the need for standardizing and automating the signal quality assessment of EEG signals and recordings. This is a serious concern, because not being able to distinguish between "good" and "bad" EEG signals is also a central problem for many brain-computer interface (BCI) applications. Arguably, one of the biggest benefits of such metrics would be to allow researchers to check the signal quality of a recording before even processing it. Being able to pre-check the signal quality prevents possible backtracking after processing data and seeing that the data is not good enough for a given application. Another huge benefit of such metrics would be the ability to compare different EEG systems according to their signal qualities. This could be particularly useful when deciding which system to use for a given application or when investing in new EEG systems. In this report, I will be presenting how metrics acquired from simple measurements such as correlation matrices and relative band powers can give insight into EEG signal qualities.

2 Equipment

The two EEG systems studied in this project are Unicorn Hybrid Black by g.tec neurotechnology GmbH and Smarting 24 by mBrainTrain. Unicorn Hybrid Black (UHB) is a consumer-grade wireless EEG system that has 8 conductive rubber electrodes (Fz, C3, Cz, C4, Pz, PO7, Oz, PO8) with 24 bits resolution and input sensitivity of ± 750 mV that can be used for both dry or wet (with conductive gel) measurements. It has a Bluetooth 2.1 interface and a sampling rate of 250 Hz per channel. It also has a 3-axis (x, y, z) accelerometer and a gyroscope to detect head movements.

On the other hand, Smarting 24 is also a wireless EEG system that has 24 channels as its name suggests (Fp1, Fp2, Fz, F7, F8, FC1, FC2, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, M1, M2, Pz, P3, P4, O1, O2). They are wet electrodes of also 24 bits resolution. Smarting has a better input sensitivity than UHB with ± 100 mV. It also has a Bluetooth 2.1 interface that supports EDR (Enhanced Data Rate), which is capable of transmitting data 2 or 3 times faster than previous versions of Bluetooth. One other perk of Smarting is that it has two recording modes with sampling frequencies 250 Hz and 500 Hz for the user to choose freely.

Most EEG systems have a built-in impedance measuring which allows researchers to use it as the primary way of assessing signal quality before beginning to record data. But not all systems have quantifiable impedance measuring. For example, the Unicorn Recorder app that allows to visualize and record data with UHB, has a built-in signal quality checker GUI (graphical user interface) but it only has two measures: good and bad. Furthermore, these are not calculated using the impedance values but rather using the standard deviation and the bandpower mean difference of the signal. These are also non-quantified measures, so the user is only left with the GUI which can be seen on the left side of Figure 1. This is not the case for Smarting Streamer though. This app shows quantifiable impedance values which help with applying the conductive gel to the electrodes and also indicates a good signal quality once the impedance values decrease.

After this brief introduction to these two systems, one would assume that Smarting 24 is a better system than UHB. One of the main goals of this project is to compute and compare different metrics to validate this assumption. If the results and the methods to compute them are robust enough, this project can be generalized to other systems as well.

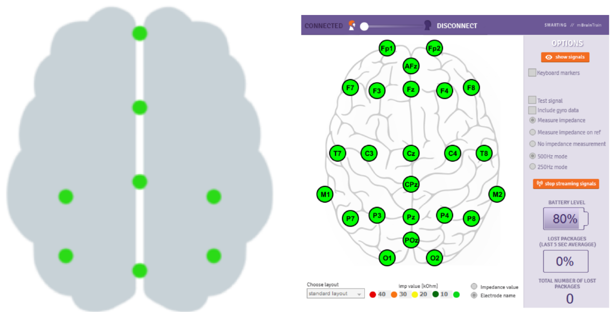


Figure 1 Left: Signal quality GUI of Unicorn Recorder, Right: Signal quality GUI of Smarting Streamer

3 Recordings

Previous work has shown that three simple measurements can help with checking and comparing the signal qualities of different EEG systems. [2] In the beginning, the participant is resting with eyes closed, then with eyes open and lastly performing a cognitively challenging task. First of all it is worthwhile to emphasize that it actually is very intuitive to use resting measurements to check the signal quality. While the participants are resting electrical fluctuations in the range 8-13 Hz called the alpha waves can be measured [3] and it is one of the most commonly used methods among researchers [2]. Normally this process is not automated and the researcher asks the participant to close his/her eyes and monitors the signal for alpha waves. But this is not always easy to detect because for example, some participants cannot produce strong alpha waves or maybe the researcher simply does not have enough experience to detect alpha waves. Secondly, one has to acknowledge that in the ideal case most probably performance-based signal quality will most be the most accurate way of comparing different systems. But this method also has its shortcomings. The participant may not be "BCI literate" to begin with. Also, making the participant perform a BCI task to check if his/her signal is good enough to perform a BCI task is counter-intuitive. It would have been much better to use simpler and faster measurements to come to the same solutions. Which leads us back to the starting point with the resting calibration measurements.

The mentioned work suggests that devices with good signal quality should be able to pick up "a significant Berger effect between measurements with the eyes open and those with the eyes closed"[2]. The Berger effect [4] states that the alpha band power should decrease once the participants open their eyes after resting with eyes closed. When it comes to the cognitive load tasks, although this work assumes that devices with good signal quality should show "a significant increase in the frontal theta power when comparing the easy and more demanding cognitive tasks" [2] this was not directly tested in this project. Instead, for the sake of simplicity, it was assumed that good systems should be able to pick up a relatively higher theta power than worse systems. That is why only one cognitive load task was used for all the measurements: multiplying two two digit numbers together in mind.

In total, there were six participants whose data was recorded in this project. The first three participants' (1, 2, and 3) data was recorded with both of the systems, UHB and Smarting 24. Whereas, rest of the participants' (4, 5 and 6) data was recorded using only UHB. In the recordings of participants 1 and 2, all 24 electrodes of Smarting 24 were used but in the recording of participant 3 only 16 electrodes were used. Figure 2 shows the topography maps of the used electrodes in both systems. Red dots represent the 8 electrodes not used in the recording of participant 3. Furthermore all participants' UHB recordings were made using the Unicorn Recorder app, except for participant 2.

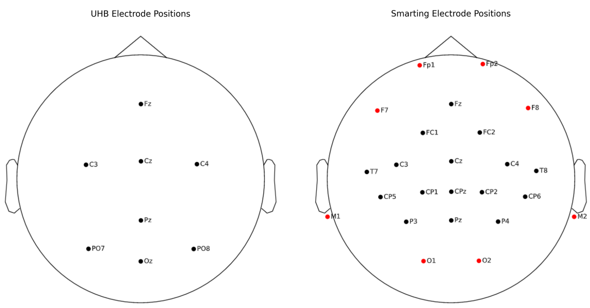


Figure 2 Left: UHB electrodes, Right: Smarting electrodes where the red dots represent the 8 electrodes not used in the recording of participant 3

To visualize the data in the Unicorn Recorder app, one has to pick a pre-defined band pass filter and a notch filter. The recordings can also be done using this app but unfortunately I was not aware that the filtering to visualize the data was also performed on the recording itself until all five participants' recordings were done. After I figured this out, I set up the Lab Streaming Layer (LSL) to record the data from participant 2 without any filtering. It was a setback that five participants' data was pre-filtered because it made it impossible to compare unfiltered signals. That is why only the data from participant 2 was used for raw UHB signals.

4 Pipeline

4.1 Filtering

After the data is recorded, the first thing to do is to crop out the first and last few seconds of it, as these parts do not capture neurophysiological phenomena. The first few second can be thought as the time span that is needed for the subject to focus on the task after the start signal and the last few seconds as the time that passes by until the researcher can stop the recording after desired amount of data is gathered. In the ideal case, the time points of where to crop the data can be set with a marker stream. Alternatively it can be approximately guessed from the time series by visual inspection. In this project, there was no marker stream and that is why each recording was cropped manually after visual inspection. But the problem with this method is that the electrical line noise (50 Hz in Germany) and its harmonics (100 Hz, 150 Hz, 200 Hz, ...) overlay with the EEG signal and it is not possible to determine the best time points to crop the data. That is why the actual first step is to filter out these frequencies with a notch filter. The results of such a filtering can be seen in Figure 3.

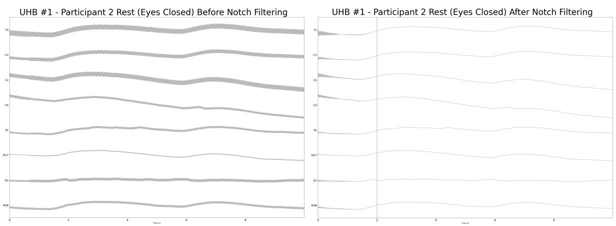


Figure 3 Time domain signals of eyes closed resting measurement (#1) of participant 1 before and after notch filtering

4.2 Computing Correlation Matrices

After this initial cleanup, the correlation matrix of the channels is computed for each measurement. This is a (num_channels) x (num_channels) matrix which captures the Pearson correlations of each electrode with the others. It was assumed that neighboring EEG sensors (electrodes) pick up approximately the same signal, just like microphones placed in a crowded room full of chatter. So, even though the ground truth of what the actual signals picked up from the electrodes should look like still remains a secret, it is fair to expect that the computed correlation matrices should show high correlations with a slight gradient representing the distance of the electrodes from each other. The problem at this point is that not all the recordings were done in the same manner as explained before. UHB measurements of the participants 1, 3, 4, and 5 were already notch filtered at 50 Hz and band-pass filtered at 0.1-60 Hz in some cases and 0.5-60 Hz in others. But the Smarting measurements were done through the Lab Streaming Layer (LSL) which returned unfiltered data. That is why the color-encoded correlation matrices in Figure 4 belong to Smarting measurements, to show the effect of band-pass filtering on the these matrices.

The color-coded correlation matrices provide a visual intuition to how different channels are correlated with each other and also which channels might be bad. Figure 4 shows the correlation matrices of eyes closed resting measurement (#1-5) of participant 2 before and after band-pass filtering. The correlation of electrode P3 with the other electrodes is very low in both cases. This might suggest that P3 is a bad channel. To confirm this suspicion, one might take a look at the time domain signal of this electrode. As it can be seen in Figure 5, P3 does not deviate from its baseline like the other channels do. In hindsight, it is not surprising because the impedance value of this electrode was constantly higher than 40 kOhm while making these recordings.

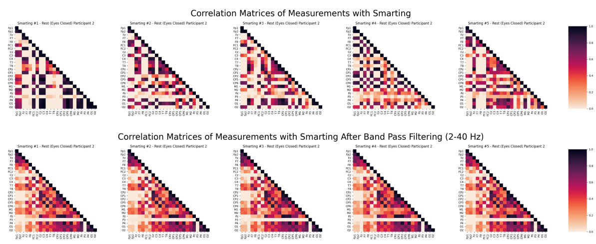


Figure 4 Color-coded correlation matrices of eyes closed resting measurements (#1-5) of participant 2 before and after band pass filtering

This example shows that the correlation matrices do not only give a visual intuition but it might also help with detecting bad channels. I have also experimented with an automated process to mark channels as bad if the absolute value of channel correlation means were lower than a predefined threshold. However, this process was not robust enough to present in this report. Nearly in all cases electrodes on the pre-frontal cortex were marked as bad channels, most probably due to eye artifacts (even though the participants' eyes were closed in the majority of measurements). If this is really the case, one possible solution would be to

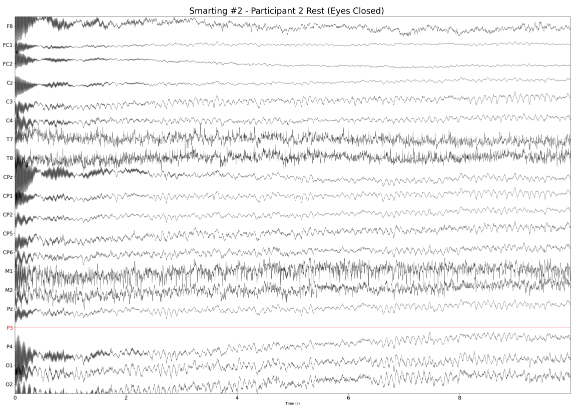


Figure 5 Time domain signals of eyes closed resting measurement (#2) of participant 2 with the bad channel P3

filter out eye artifacts, for example with Independent Component Analysis (ICA), and then compute the correlation matrices again from the cleaned up data. This was not performed in this project but I believe that with further adjustments similar to this, correlation matrices can be easily used to automatically detect bad channels.

As mentioned, since the UHB recordings could not be converted back to raw, unfiltered signals any more, all UHB and Smarting recordings were band pass filtered with the cutoff frequencies of 2 Hz and 40 Hz. Based on the following quote this cutoff frequency of 2 Hz is actually not desired: "When it comes to high-pass filtering, using corner frequencies above 0.1 Hz were found to be generate a systematic bias easily leading to misinterpretations of neural activity." [5] But since the time accuracy of the signals do not play a role in these calibration recordings, it was desirable to filter out low frequency drifts in data, like breathing of the subjects. The effect of this filtering can be seen in the middle plot of Figure 6.

After this filtering the correlation matrices of UHB and Smarting measurements could be compared with each other. To have a common ground, only the electrodes present in both systems were used to compute the correlation matrices, these were: Fz, C3, Cz, C4, Pz. Also, the component wise mean values of all the correlation matrices for the same type of measurement (e.g., cognitive load) were computed for each participant. The results can be seen in Figure 7. Curiously enough, the UHB correlation matrices are much darker than the Smarting matrices. This could be due to a better separation of source signals by the Smarting device. Since Fz is the electrode highly uncorrelated with the rest of the electrodes, it could be that the brighter colors in Smarting matrices do not necessarily mean worse signal quality, because as mentioned before, frontal channels were rather sensitive in these matrices.

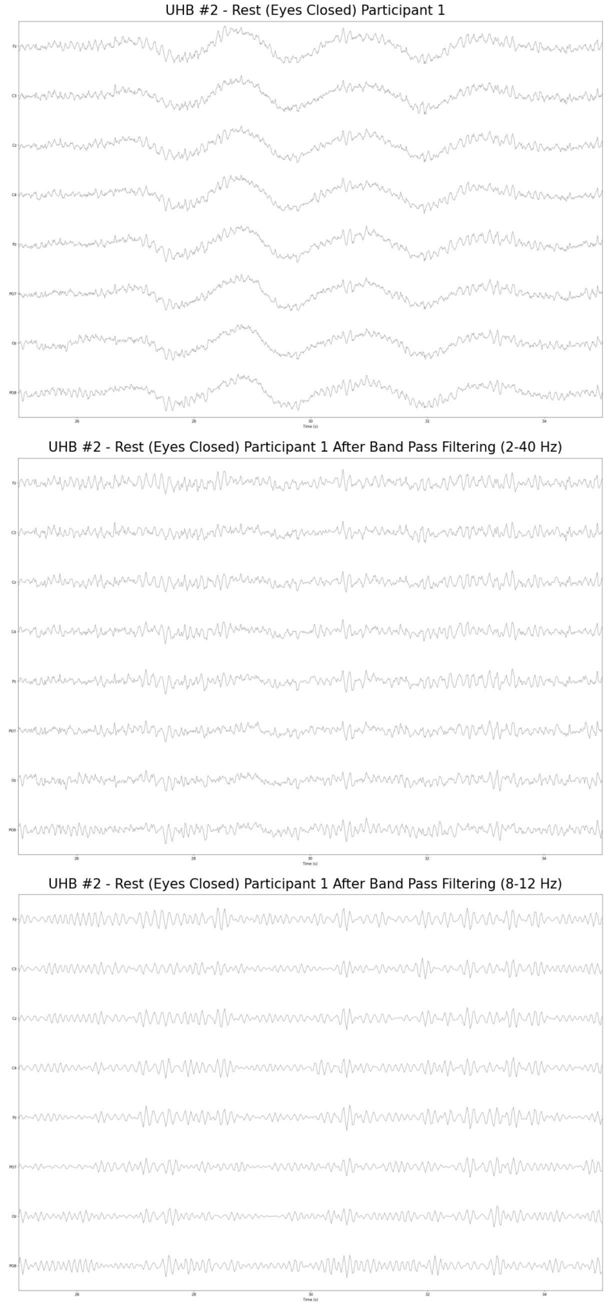


Figure 6 Top: Time domain signals of eyes closed resting measurement (#2) of participant 1 where the low frequency drifts because of breathing are clearly visible, Middle: Time domain signals of eyes closed resting measurement (#2) of participant 1 after band pass filtering (2-40 Hz) where the low frequency drifts were filtered out, Bottom: Time domain signals of eyes closed resting measurement (#2) of participant 1 after band pass filtering (8-12 Hz) where the alpha waves are clearly visible

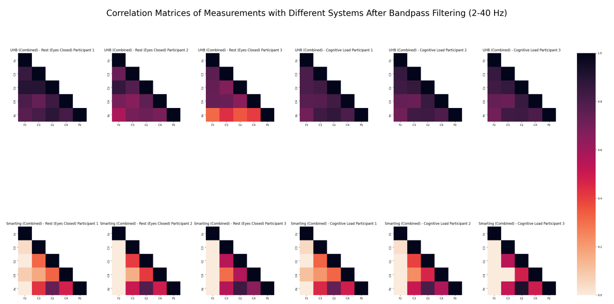


Figure 7 Averaged correlation matrices of each type of measurement for each participant, Top row: UHB, Bottom row: Smarting

4.3 Relative Band Power

Finally the relative band powers were computed. This was done by first normalizing the data to values between 0 and 1, since the scaling of each recording was different. Normalized data was then used to compute the power spectral density (PSD) of the signal using Welch's method [6]. All the computed PSDs of different channels were then averaged. With the help of logical indexing, the desired bands were isolated and the area under the curve was approximated using Simpson's method. The computed value was the absolute band power. The total band power was computed analogously. The relative band power for a given band was just the ratio between these two values. As the scalings were off in some measurements, the relative band power was a more robust result. The described PSD plot can be seen in Figure 8.

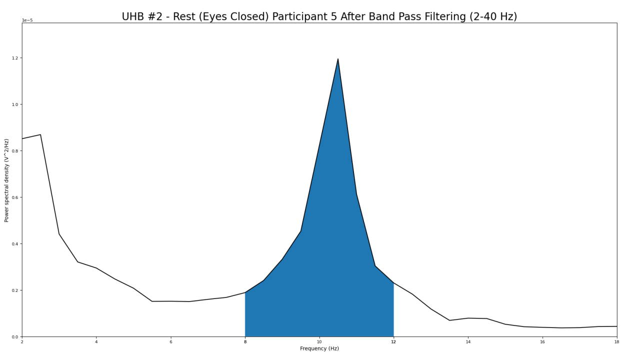
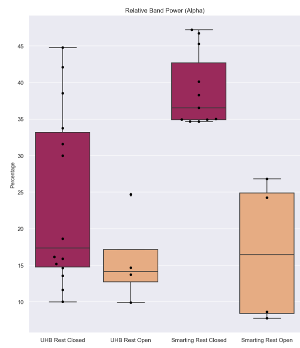


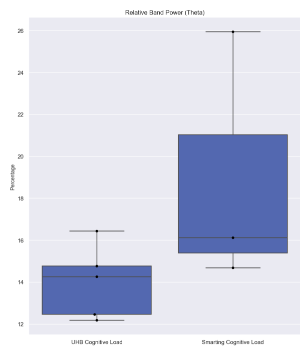
Figure 8 Power spectral density of eyes closed resting measurement (#2) of participant 5, blue area: alpha band (8-12 Hz)

Once relative band powers were computed for each measurement, it was time to test whether there would be any differences between the two systems. Figure 9 shows this comparison for the alpha band and Figure 10 show for the theta band. It is comforting to see that eyes closed resting measurements with Smarting are more robust with less variance than UHB. Furthermore, it seems like the Berger effect is more prominent with Smarting measurements as there is no overlap between eyes closed and eyes open measurements. Looking at the relative theta band powers, it looks like Smarting provided better results also in this case. But one thing to consider is that the sample sizes differ significantly from measurement to measurement. Although the results look promising enough, it would be exciting to perform statistical tests on this data once there are more samples gathered.



	UHB Rest Closed	UHB Rest Open	Smarting Rest Closed	Smarting Rest Open
Mean	24.0	15.71	38.92	16.83
STD	12.22	6.32	5.11	10.08
Sample Size	14.0	4.0	11.0	4.0

Figure 9 Left: Distribution of relative alpha band powers of resting measurements for UHB and Smarting, Right: Mean, standard deviation and sample size of each type of measurement



	UHB Cognitive Load	Smarting Cognitive Load
Mean	14.01	18.9
STD	1.75	6.13
Sample Size	5.0	3.0

Figure 10 Left: Distribution of relative theta band powers of cognitive load measurements for UHB and Smarting, Right: Mean, standard deviation and sample size of each type of measurement

5 Conclusion

The first thing this project has shown is that matrices of correlations between the electrodes provide a visual intuition to how the data behaves and that they might be efficient tools to check for bad channels in the recorded data. Since the correlation matrices look nearly the same for same kinds of measurements independent from the participant, it might be of good practice to compare correlation matrices of new data with older ones. This might show if the new recording is statistically similar to previous data. Also this would be a much simpler feature to use for statistical tests, especially when working in higher dimensions like with EEG data. Another finding of this project came from relative band powers. Relative alpha and theta band powers computed from simple measurements such as resting states or multiplying two two digit numbers together provided promising results to compare different systems with each other. Although the scope of this project was rather small, same measurements can be repeated in future studies with different systems and participants, also building a database for more accurate results in the meantime. I believe that such simple computations and statistical testing are the easiest and most reliable way of assessing signal qualities of EEG systems, until we discover the "ground truth" behind brain signals.

References

[1] Anderson S. Oliveira et al. "Proposing Metrics for Benchmarking Novel EEG Technologies Towards Real-World Measurements". In: *Frontiers in Human Neuroscience* 10 (2016), p. 188. ISSN: 1662-5161. DOI: 10.3389/fnhum.2016.00188. URL: https://www.frontiersin.org/article/10.3389/fnhum.2016.00188.

[2] Thea Radüntz. "Signal Quality Evaluation of Emerging EEG Devices". In: *Frontiers in Physiology* 9 (2018), p. 98. ISSN: 1664-042X. DOI: 10.3389/fphys.2018.00098. URL: https://www.frontiersin.org/article/10.3389/fphys.2018.00098.

[3] Rajesh P. N. Rao. *Brain-Computer Interfacing: An Introduction*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139032803.

[4] Hans Berger. "Über das Elektrenkephalogramm des Menschen". In: *Archiv für Psychiatrie und Nervenkrankheiten* 87.1 (Dec. 1929), pp. 527–570. ISSN: 1433-8491. DOI: 10.1007/BF01797193. URL: https://doi.org/10.1007/BF01797193.

[5] David J. Acunzo, Graham MacKenzie, and Mark C.W. van Rossum. "Systematic biases in early ERP and ERF components as a result of high-pass filtering". In: *Journal of Neuroscience Methods* 209.1 (2012), pp. 212–218. ISSN: 0165-0270. DOI: https://doi.org/10.1016/j.jneumeth.2012.06.011. URL: https://www.sciencedirect.com/science/article/pii/S0165027012002361.

[6] P. Welch. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73. DOI: 10.1109/TAU.1967.1161901.