Exploring the effect of caffeine consumption on EEG sleep signals using ML

by Philipp Thölke

Index

- Introduction
- Methods
- Results
- Discussion
- Possible future work

Introduction - Motivation

- the quality of sleep has a direct impact on health
  - lack of sleep and sleep disorders can lead to deterioration of sleep-related brain processes
  - bad sleep quality increases the risk for depression, weight gain, hypertension, cardiovascular diseases and diabetes
  - (Short et al., 2013; Patel and Hu, 2012; Gangwisch et al., 2006; Cappuccio et al., 2010; Gottlieb et al., 2005)
- caffeine as a psychostimulant and antagonist to adenosine reduces the natural circadian sleep pressure by attaching to adenosine receptors (Ribeiro and Sebastiao, 2010)
  - feeling of higher alertness and invigoration
- a better understanding of the effects of caffeine on brain activity during sleep is a major health concern due to high levels of consumption in the population (Garriguet, 2008)

Introduction - Previous work

- caffeine affects spectral power, the amount of contribution to the signal by different frequencies, of the EEG during sleep (Drapeau et al., 2006; Robillard et al., 2015)
  - decreases in delta power and increases in beta have been found by multiple studies
  - has been detected in humans and Cynomolgus monkeys (suggested to be representative due to diurnal nature and similar proportion of sleep stages compared to humans; Authier et al., 2014)
- sleep variables are disturbed by the ingestion of caffeine before sleep
  - increase in sleep latency, decrease in efficiency, decrease in total sleep duration and shift in sleep stage distribution
- increased resting brain entropy due to caffeine found in an fMRI study (Chang et al., 2018)
  - entropy is a measure of complexity, indicating how unpredictable or random a signal is
  - higher resting brain entropy after caffeine ingestion might indicate an increase of information processing capacity

Introduction - Hypotheses

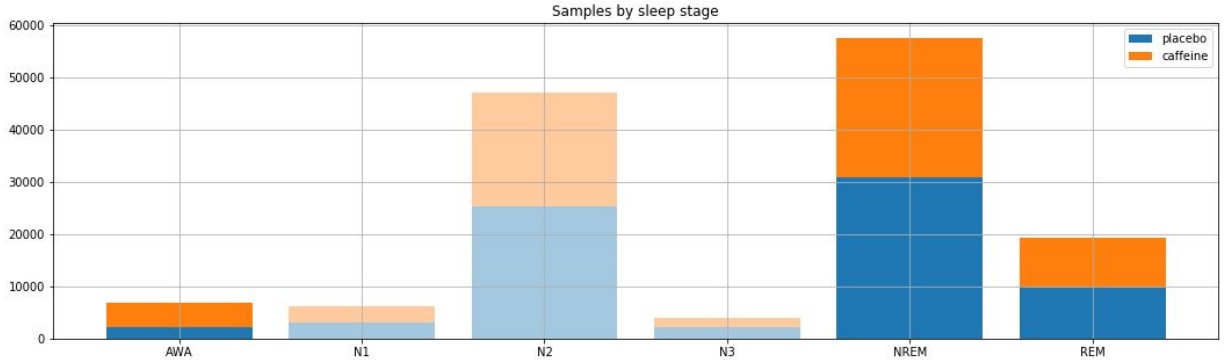
- caffeine might also increase brain entropy in electrophysiological recordings, not only in the BOLD signal
- exploration of EEG features using a data-driven approach may quantify the separation between caffeine and placebo
- new biomarkers could be identified, helping the understanding of the influence of caffeine on cerebral dynamics during sleep

Methods - Data acquisition

- sleep EEG from 40 participants
  - from 28 to 58 years old, mean age 35.3 ± 14.3 years
  - 21 male, 19 female
- randomized, double-blind, cross-over design
  - subjects spent two nights at sleep laboratory, receiving 200 mg caffeine or placebo (lactose) capsule before regular bedtime
- no caffeine consumption after noon on day of recording
- recording was done with 20-electrode EEG cap at 256 Hz
  - arranged in international 10-20 system
  - referential montage with linked ears

Methods - Preprocessing

- artifact removal, spindle and slow wave extraction
- data divided into 20s epochs
- segmentation into sleep stages using hypnogram (AWA, N1, N2, N3, REM)
    - N1, N2, N3 were combined into single NREM stage
    - AWA stage only contains wake epochs after sleep onset (no data for two subjects, excluded)



Methods - Feature extraction

Power spectral density (PSD)

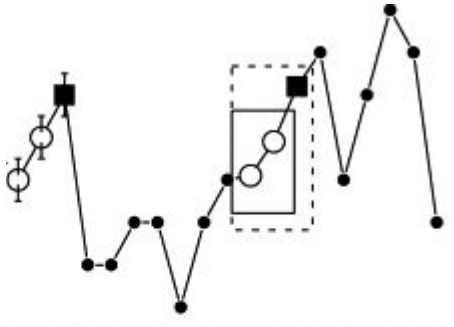
- computation using Welch's method of averaged periodograms
  - Hamming window on six segments without overlap (Bartlett's method)
- extraction of power bands from six frequency intervals by averaging
  - delta: 0.3 - 4.0 Hz
  - theta: 4.0 - 8.0 Hz
  - alpha: 8.0 - 12.0 Hz
  - sigma: 12.0 - 16.0 Hz
  - beta: 16.0 - 32.0 Hz
  - low gamma: 32.0 - 50.0 Hz

Methods - Feature extraction

Spectral entropy (SpecEn)

- PSD extracted in the same way as for power bands
- Shannon entropy was applied to full power spectrum (instead of frequency bands)
- estimation power spectrum complexity (higher entropy means higher complexity/more randomness)
- Shannon entropy looks at the relative power of each frequency bin
    - uniform distribution over all bins corresponds to maximal entropy
    - 100% power in one frequency bin means no complexity/randomness → minimal entropy

Methods - Feature extraction



Sample entropy (SampEn)

- estimates entropy by evaluating probability of temporal continuation of a window somewhere else in the signal
  - negative logarithm of probability that if two windows of size m match by a distance threshold r times the standard deviation of the signal, the two windows also match with size m+1
- similar to approximate entropy (ApEn; often used for biomedical time series)
- advantages of SampEn over ApEn:
  - does not count self-matches → less bias
  - not dependent on signal length due to normalization term
  - higher consistency concerning parameter choice
- parameter choice: m=2, r=0.2 * signal standard deviation

Methods - Feature extraction

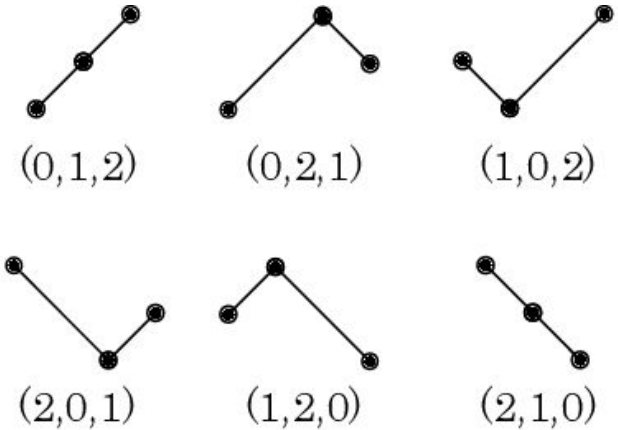
Spectral sample entropy (SpecSampEn)

- very similar to SpecEn but using SampEn on the power spectrum instead of Shannon entropy
- able to look at similar frequency powers at once
- looks for similar patterns across frequencies in the power spectrum
- parameters for SampEn: m=2, r=0.2 * signal standard deviation

Methods - Feature extraction

Permutation entropy (PermEn)

- evaluates times series by computing Shannon entropy on the probability distribution of different ordinal patterns of length n with a delay of τ
- occurrences are counted for n! different ordinal patterns
- looks at diversity in the ordering of values in the signal
    - random ordering leads to high PermEn
- chosen parameters were embedding dimension n=3, sample delay τ=1



Methods - Feature extraction

Spectral permutation entropy (SpecPermEn)

- PermEn applied to the power spectrum
- same idea as SpecSampEn
- looks at distribution of ordinal patterns inside the PSD
- parameters for PermEn: n=3, τ=1

Methods - Feature averaging

- low comparability between machine learning performances across stages when sample count varies strongly
- classifier performance may suffer from skewed class distributions
- features were averaged subject-wise, leaving two samples per subject
    - 80 samples in NREM and REM stages (40 caffeine and 40 placebo)
    - 76 samples in AWA stage (38 caffeine, 38 placebo)
- feature-wise z-transformation of samples using mean and standard deviation across electrodes individually for each subject

Methods - Statistical analyses

- assessing the statistically significant differences between caffeine and placebo conditions for each electrode
- placebo samples were subtracted subject-wise from the caffeine condition
- two-sided paired permutation-based pseudo t-tests (tmax correction) applied to all extracted features
- exhaustive permutations with the number of permutations n=10000
- significance was evaluated at $p<0.05$ and $p<0.01$

Methods - Single-feature machine learning

- single-feature, single-electrode classification between caffeine and placebo condition
- four different algorithms for more reliability:
  - support vector machine (SVM), linear discriminant analysis (LDA), Gaussian process, perceptron
- permutation tests (n=1000) were applied to estimate the confidence of classifier scores (significant at $p<0.05$ and $p<0.01$)
  - retraining the classifier n-1 times with permuted labels to determine if the classifier learned a feature-label dependency or is guessing randomly
- 10-fold cross-validation in each permutation, score is averaged across test folds

Methods - Multi-feature machine learning

Single-electrode classification

- determine overall effect of caffeine on the features for each electrode
- classifiers were trained on all extracted feature combined (11 features)
  - 6 PSD bands, SpecEn, SampEn, SpecSampEn, PermEn, SpecPermEn
- four different algorithms for more reliability:
  - support vector machine (SVM), linear discriminant analysis (LDA), random forest, multilayer perceptron (MLP)
- permutation tests (n=1000) with 10-fold cross-validation (significance at $p<0.05$ and $p<0.01$)
- an ensemble classifier was created from the four classifiers' predictions using majority vote (no permutation tests)

Methods - Multi-feature machine learning

Multi-electrode classification

- a random forest was trained on combined features from all electrodes
  - 11 extracted features * 20 electrode = 220 total features
- feature importance was used to estimate the effect of caffeine on different electrodes and features
  - importance for one feature can be calculated by averaging the height of the feature in all decision trees
  - higher features (closer to the root) contribute more to the predictive decision of decision trees
- training of the random forest was repeated 1000 times
  - determine variance in feature importance and classification accuracy
  - 7-fold cross-validated hyperparameter grid search inside a leave 5 subjects out cross-validation (left out subjects different in each iteration of training a random forest)

Results - Statistical analyses

- awake
  - decrease in left theta electrodes, central sigma and beta
- non-REM
  - decrease in central electrodes in delta
  - increase in motor areas and occipital in beta, increase in central electrode in low gamma
  - strong increase in SpecEn, SampEn and SpecSampEn across the cortex
- REM
  - decrease in left electrodes for theta
  - increased sigma and beta power and SpecEn in occipital

Results - Single-feature machine learning

- Perceptron classifier performs worse than Gaussian process, LDA and SVM
- awake
  - best scores for theta and sigma power (mostly central to frontal)
  - also good results in delta power (central) and SpecPermEn (right occipital)
- non-REM
  - good scores in delta power in different electrodes across cortex
  - best accuracy and most significant electrodes in SpecEn, SampEn and SpecSampEn
  - few significant electrodes in beta power
- REM
  - theta power is the best feature
  - good scores also visible in sigma power (frontal) and SpecEn and SpecPermEn (both left parietal)

Results - Multi-feature machine learning

Single-electrode classification

- non-REM classification results in bests scores/significance (around 70% accuracy, $p<0.01$)
- worst performance in REM (55-60% accuracy)
- classification in wake is most significant in occipital, reaching about 65% accuracy
- ensemble classifier does not improve scores much

Results - Multi-feature machine learning

Multi-electrode classification

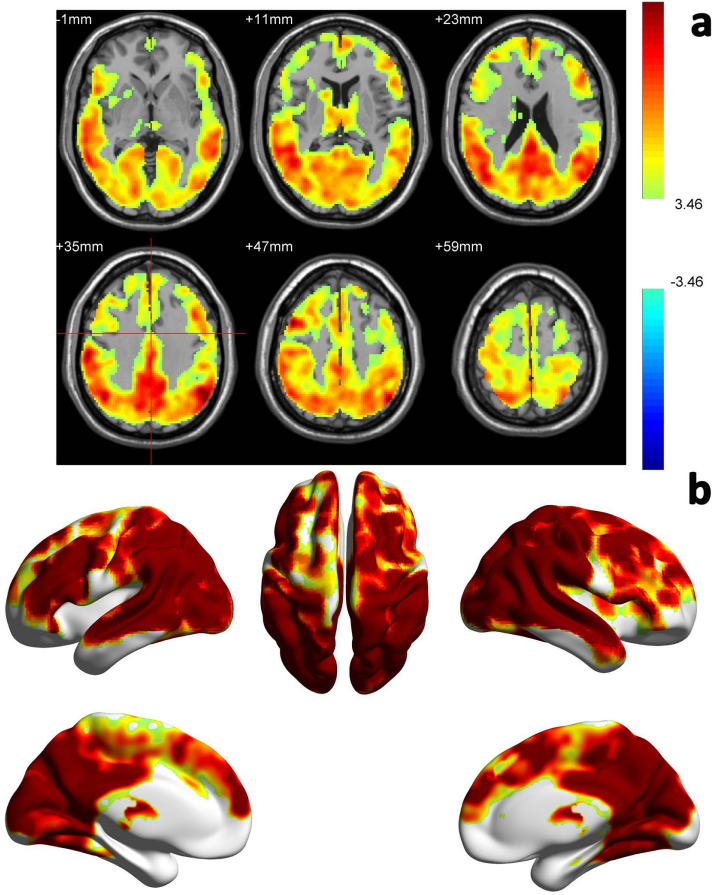
- mean classification accuracy: awake 65.6%, non-REM 73.4%, REM 61.2%
- awake
  - SpecPermEn is most important feature
  - theta and sigma power also show more importance than other features
- non-REM
  - SpecEn most important feature, SpecSampEn and SpecEn also very important
  - delta is only power band showing increased feature importance
- REM
  - theta powe and SpecPermEn have highest feature importance
  - sigma power, SpecEn and PermEn show higher importance as well

Discussion - comparison of the approaches

- high importance of SpecEn, SampEn and SpecSampEn in NREM could be observed throughout all methods
- importance/change in theta and sigma power during AWA and REM visible across apporaches
- SpecPermEn has strongest importance in the multi-feature random forest apporach during AWA and REM but was also visible in single-feature ML
    - spatial distribution very similar in the two approaches
- importance of delta power in NREM more significant in ML approaches than statistics
- change in beta power in NREM not visible in multi-feature random forest approach but visible in single-feature ML and statistics

Discussion - fMRI brain entropy

- fMRI study showing an increase in sample entropy in resting brain induced by a 200 mg caffeine dose
- brain entropy increase in large portion of the cerebral cortex, DMN, visual cortex and motor network
- the fMRI results closely match the increase of sample entropy observed in AWA, NREM and REM
    - only statistically significant difference in NREM, might be due to a larger amount of samples in NREM stage leading to less noise after averaging
- increased brain entropy could also be observed in electrophysiological recordings, not only in the BOLD signal



Discussion - SpecSampEn and SpecPermEn

- SpecSampEn and SpecPermEn are new methods to measure brain entropy using spectral analysis
- important features in the multi-feature, multi-electrode random forest approach
  - SpecSampEn in NREM
  - SpecPermEn in AWA and REM
- SpecSampEn closely matches the results from SpecEn while SpecPermEn shows unique patterns
- might be able to capture more complex patterns in spectral power
  - sample entropy and permutation entropy are more complex measures than Shannon entropy

Discussion - 400 mg group

- in the same study, another group of subjects were recorded with a 400 mg dose of caffeine before sleep
  - only 20 subjects, decreased robustness of analyses
- overall similar results as 200 mg group, difficult to distinguish differences due to low subject count
  - stronger decrease in delta band, visible across statistical and machine learning analyses
  - SampEn statistics and machine learning performances show a weaker entropy increase for 400 mg than for 200 mg

Future work

- analysing difference between 200 mg and 400 mg doses
  - transfer learning, testing on 400 mg data
  - t-SNE showing placebo, 200 mg, 400 mg classes for features
- connectivity analysis (comparison to MEG paper on caffeine)
- comparing PSD vs entropy features in sleep stages (class separation, t-SNE, clustering)
- spectral entropies on frequency intervals (instead of complete spectrum)
- deep learning on raw EEG signal (CNN, LSTM, domain adaptation network)
- removing eye movement artifacts from EEG in REM stage
  - using PCA or ICA to remove artifact components in combined EEG and EOG
  - maybe convolutional autoencoder with modified loss function to exclude eye movements
- further analyses of SpecSampEn and SpecPermEn

Thank you for your attention!