BCI Competition IV - Graz Dataset 2b

EEG Brain-Computer Interface

The dataset

- The goal is to allow subjects to navigate a virtual environment by using their thoughts only
- The dataset is created by fitting 10 subjects with non-intrusive EEG sensors and measuring the signals generated
- Subjects were seated 1m from an LCD screen
- The data from 3 EEG (measure brain activity), 3 EOG (measure eye activity) and four EMG (measure muscle activity) sensors were recorded with a sampling frequency of 250 Hz

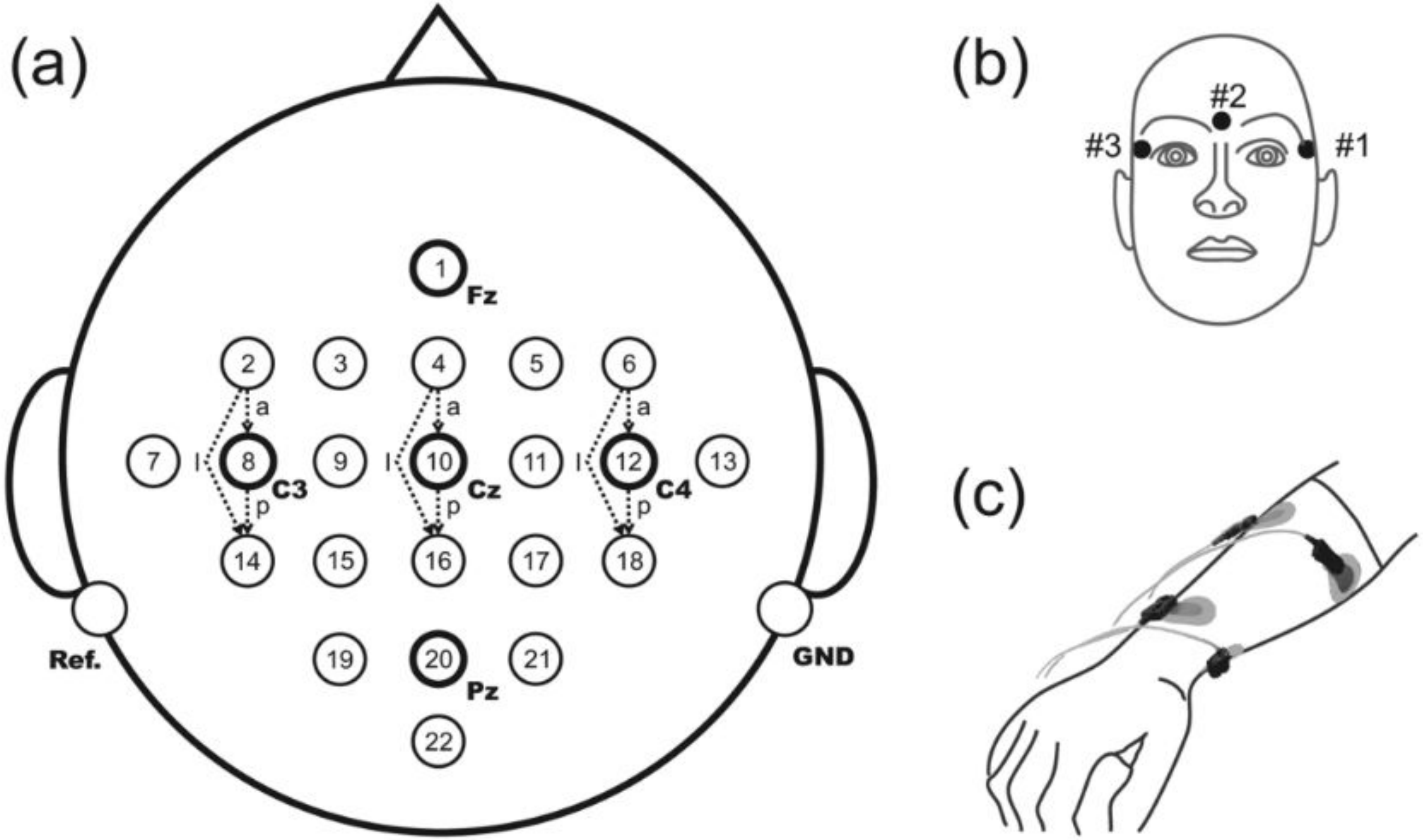


Fig. 1. (a) Placement of the 22 EEG and (b) the 3 EOG electrodes. For both, the reference electrode was placed at the left and the ground electrode at the right mastoid. (c) Location of the EMG electrodes on the right hand placed over the *musculus extensor digitorum* and the *musculus flexor digitorum superficialis*; for the left hand the same positions have been used. The arrows between the EEG electrodes show the analyzed bipolar derivations ($\oplus \rightarrow \ominus$), "a" stands for anterior-central, "p" for central–posterior, and "l" for the large distance between the bipolar electrodes (anterior–posterior).

Experimental Procedure

- Each naive subject had the task to perform kinesthetic motor imagery (MI) indicated by the visual cue on the monitor.
- Prior to the first motor imagery training, the subject executed and imagined different movements for each body part and selected the one which they could imagine best (e.g., squeezing a ball or pulling a brake).
- The cue-based experimental paradigm consisted of two imagery classes: motor imagery of left hand and right hand. Each subject participated in two sessions recorded on two separated days within two weeks. Each session consisted of six runs with ten trials each and two classes of imagery.
- This resulted in 20 trials per run and 120 trials per session. Data of 120 repetitions of each MI class were available for each person in total.

Experimental Procedure – Part 2

- Each trial started with a fixation cross and an additional short acoustic tone (1 kHz, 70 ms).
- Some seconds later a visual cue (an arrow pointing either to the left or right, according to the requested class) was presented for 1.25 s.
- Afterwards the subjects had to imagine the corresponding hand movement over a period of 4 s.
- Each trial was followed by a short break of at least 1.5 s. A randomized time of up to 1 s was added to the break to avoid adaptation.
- Twenty-two monopolar EEG channels (reference left mastoid, ground right mastoid) were recorded.

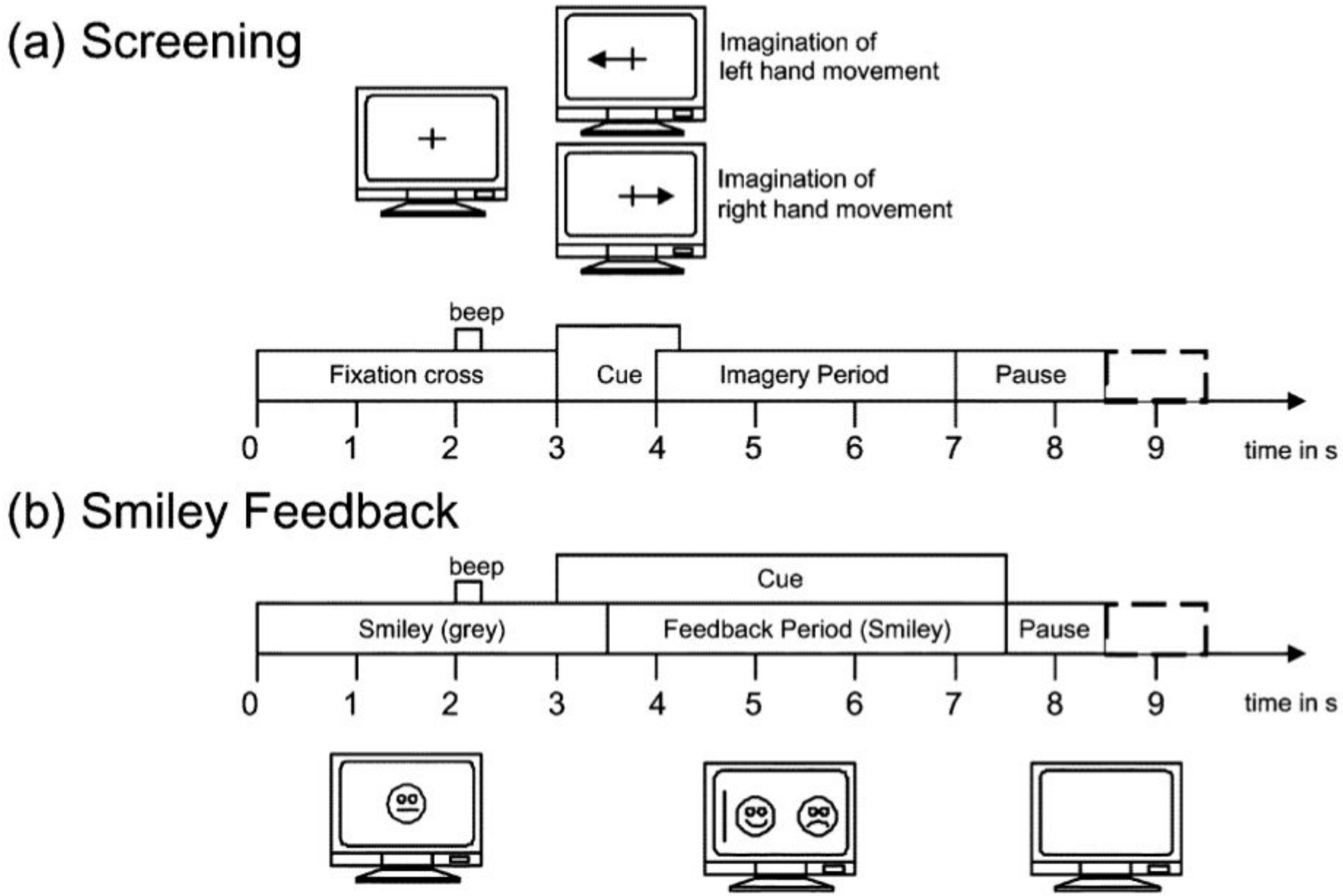


Fig. 2. (a) Timing of the movement imagery task (screening). Cue stimulus between second 3 and 4.25 in form of an arrow (either pointing to the left or right) instructs the participant to imagine the desired movement. (b) Timing of the cue-based feedback experiments and principle of the smiley paradigm.

The dataset *(work in progress)*

Each of the 324 recordings is an 8-second recording of 3 EEG channels (C3, Cz, C4) produced at a sampling frequency of 250Hz. 250 Hz * 8 sec = 2000 data rows in each recording.

Band pass filter applied between 2.0 Hz and 60.0 Hz

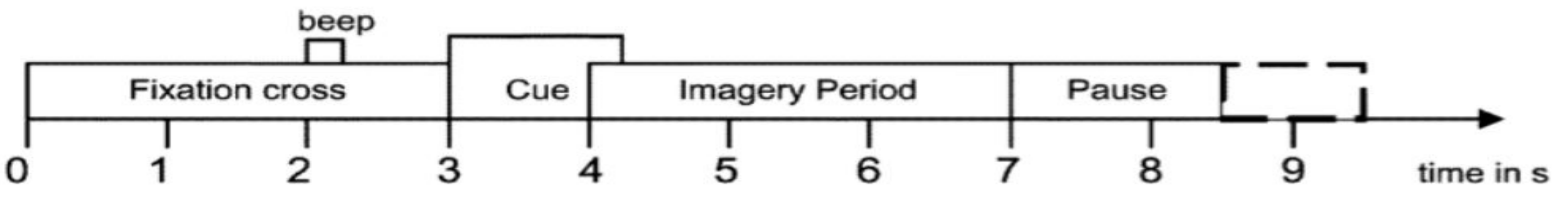
Notch filter applied at 50Hz

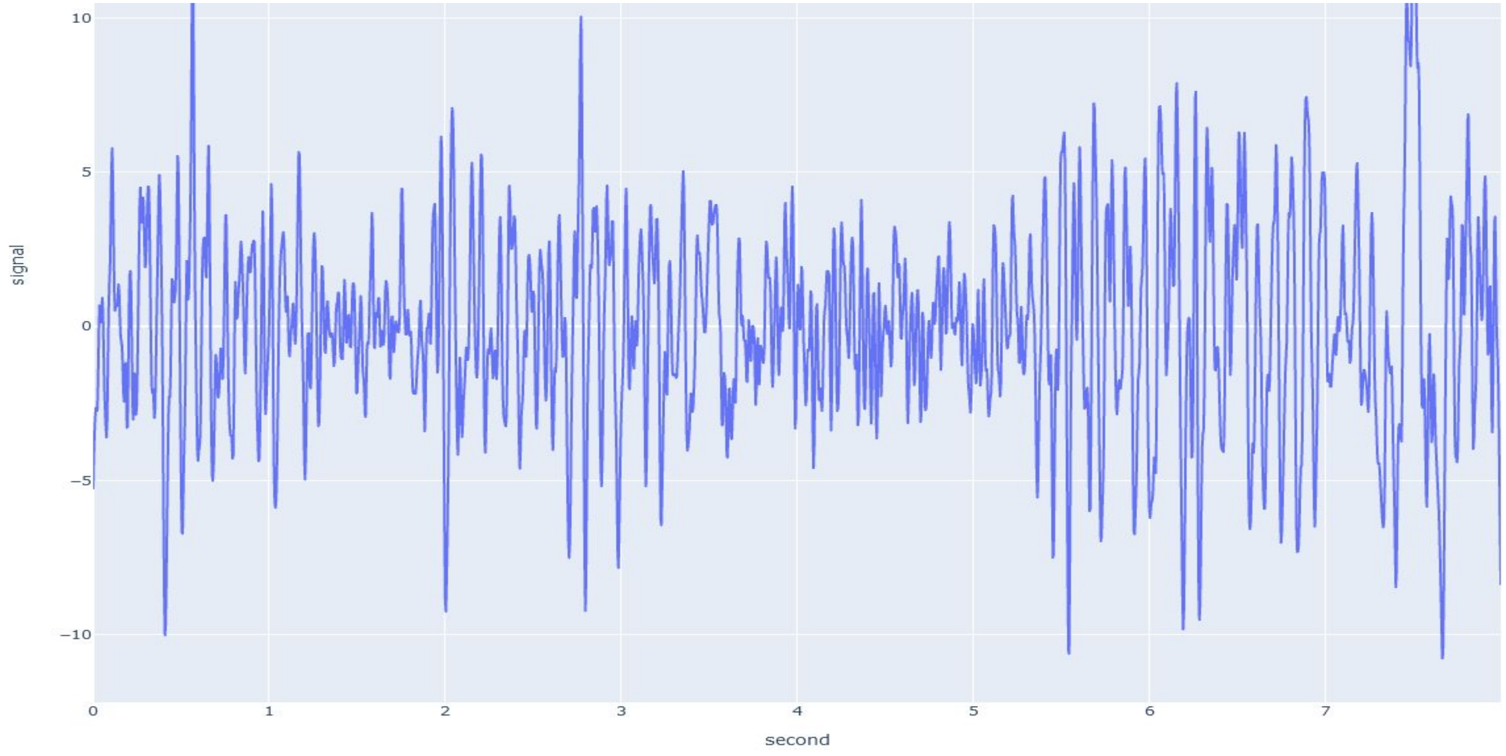
	C3	Cz	C4
0	-5.24511	-1.82425	-3.23970
1	-4.18287	-3.42862	-1.96952
2	-3.41794	-4.17030	-0.35956
3	-2.89989	-3.77997	0.98580
4	-2.63874	-3.17990	1.82998
5	-2.74359	-2.98616	2.46849
6	-2.68974	-2.69485	3.09208
7	-1.75103	-1.82494	3.41837
8	-0.22834	-0.94594	3.29316
9	0.68729	-0.88505	3.14259
10	0.50816	-1.48062	3.22686
11	0.13007	-1.90799	2.89296
12	0.43936	-1.98659	1.46716
13	0.92552	-2.31579	-0.36009
14	0.49707	-3.13650	-1.14277
15	-0.81805	-3.92324	-0.63780
16	-2.02061	-4.23979	0.02626
17	-2.69504	-4.15818	0.15727
18	-3.24304	-3.61127	0.15335
19	-3.59652	-2.18442	0.29360
20	-2.96299	0.05612	0.29204

x_train[0] Format: %.5f

A single recording with dimensions (2000, 3)

Visualization of a training sample in time domain

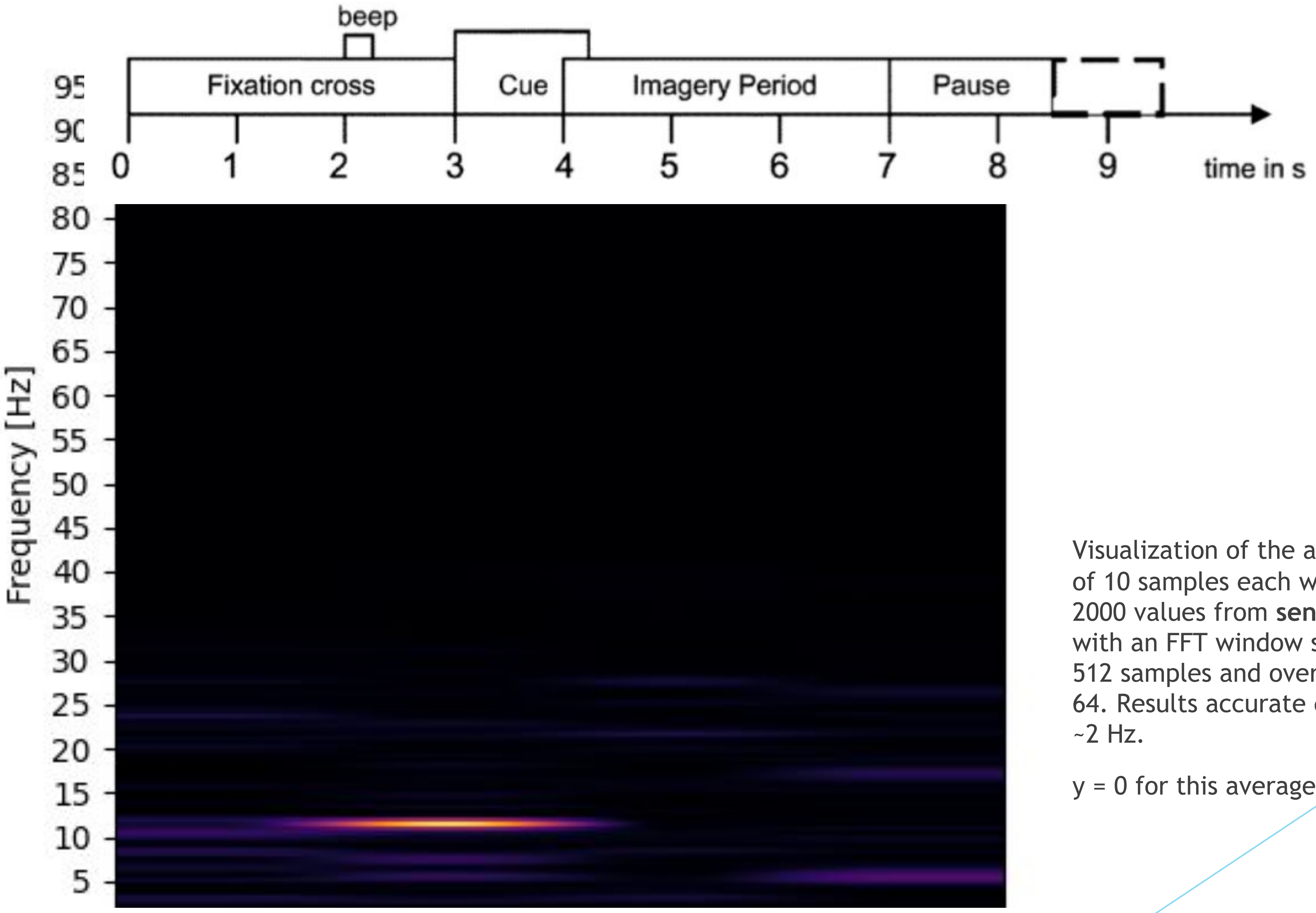




Visualization of 2000 samples from sensor C3.

y = 0 for this sample.

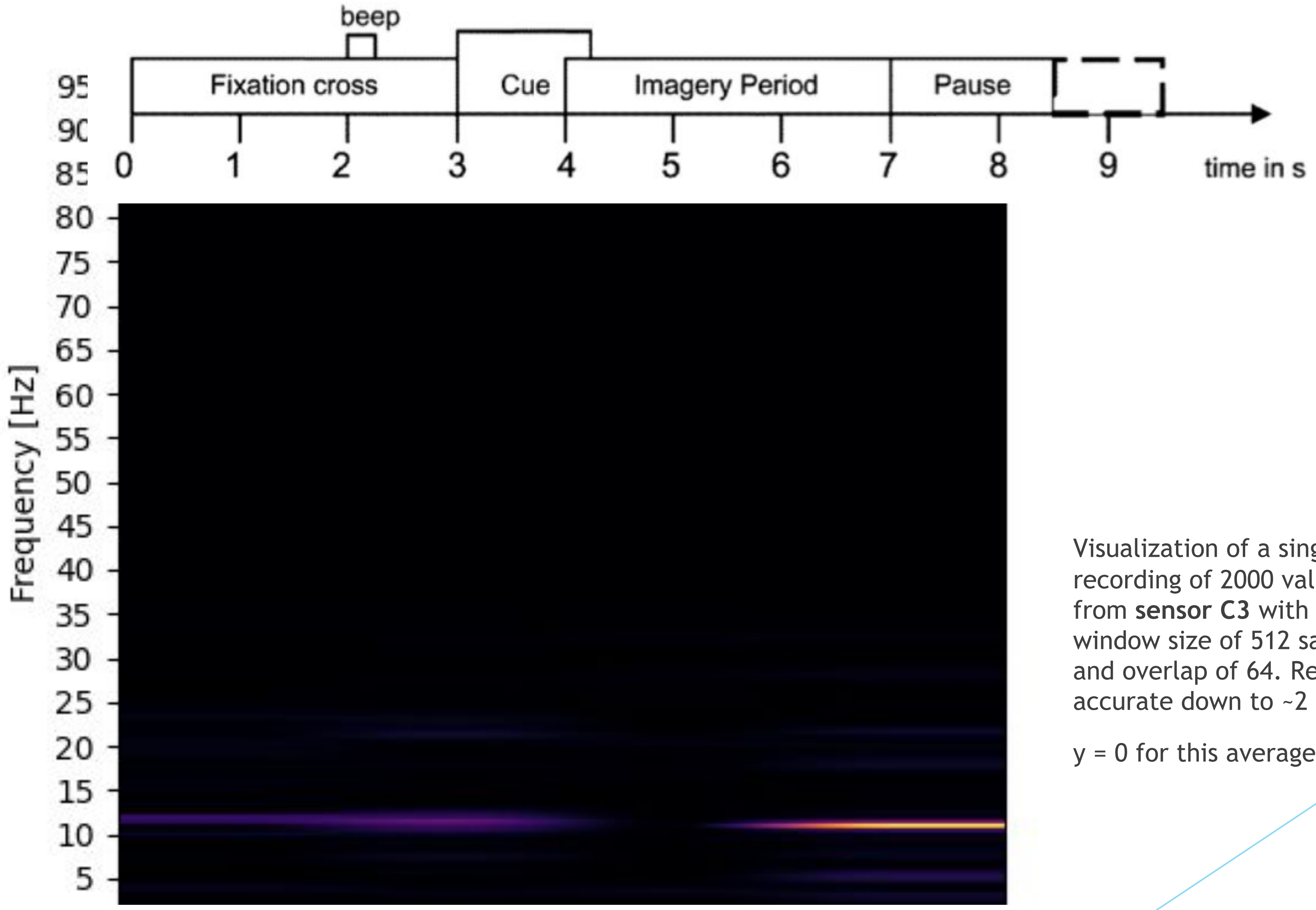
Spectrogram - Average for class 1 (y=0)



Visualization of the average of 10 samples each with 2000 values from **sensor C3** with an FFT window size of 512 samples and overlap of 64. Results accurate down to ~2 Hz.

y = 0 for this average.

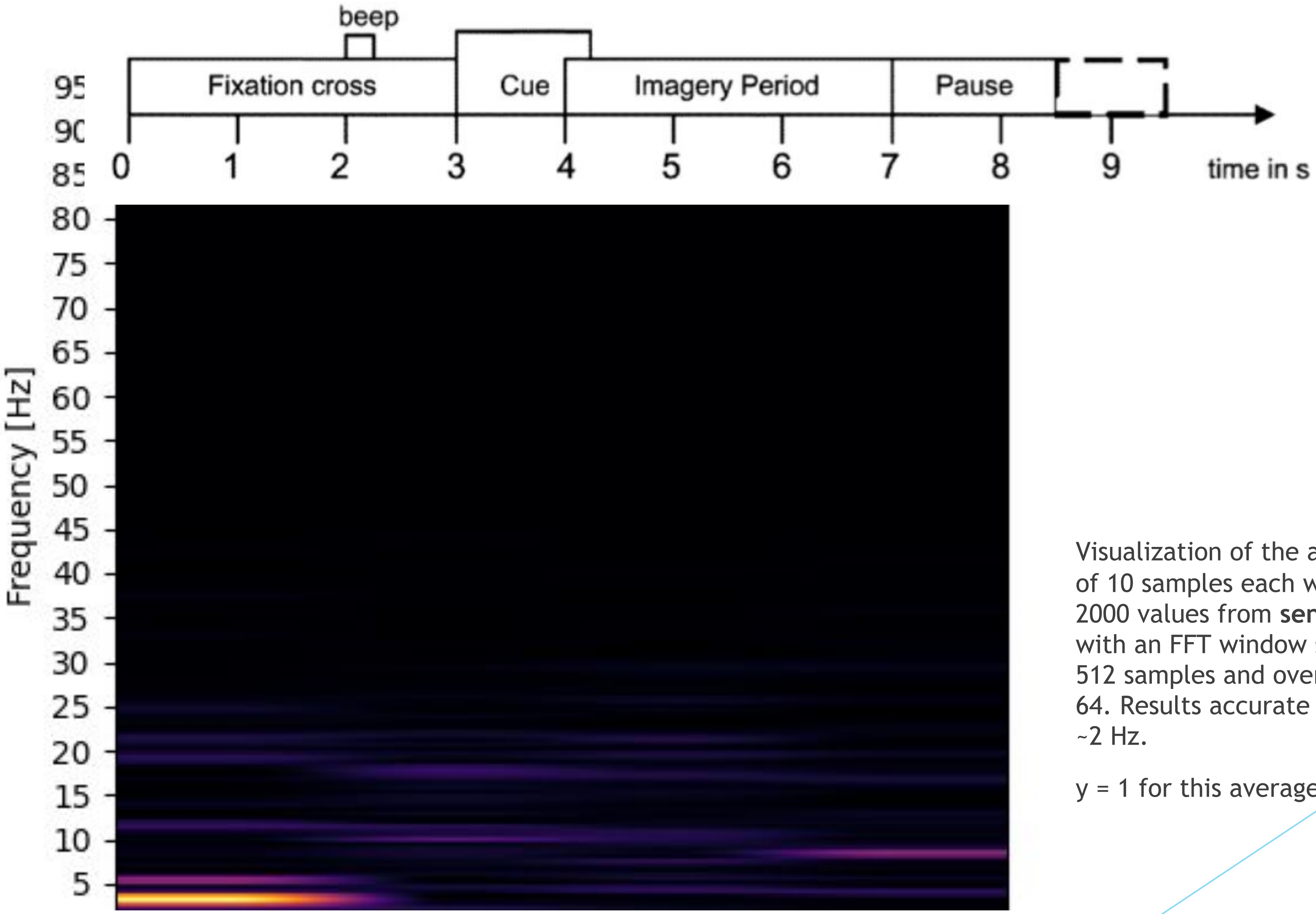
Spectrogram - A recording of class 1 (y=0)



Visualization of a single recording of 2000 values from **sensor C3** with an FFT window size of 512 samples and overlap of 64. Results accurate down to ~2 Hz.

y = 0 for this average.

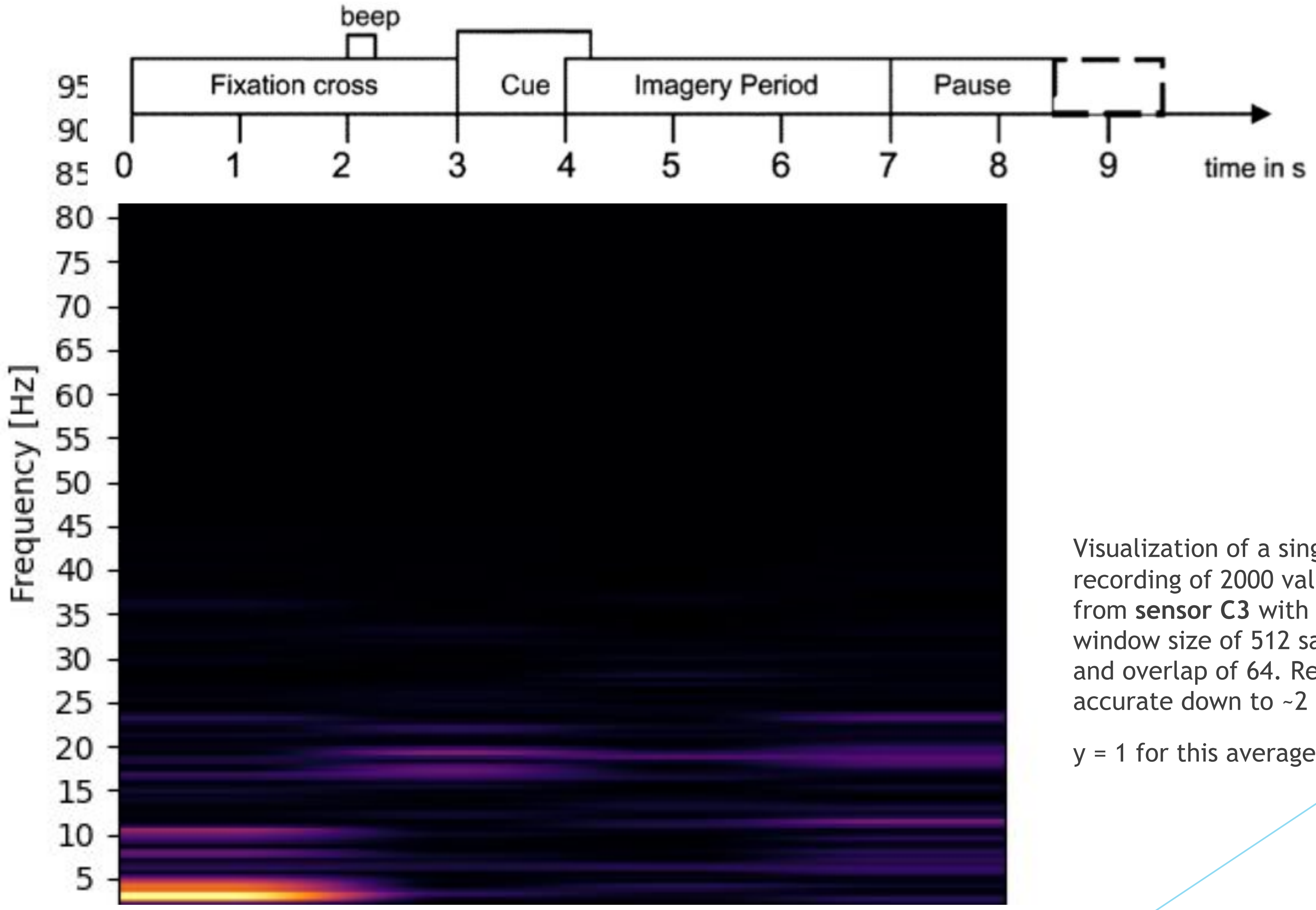
Spectrogram - Average for class 2 (y=1)



Visualization of the average of 10 samples each with 2000 values from **sensor C3** with an FFT window size of 512 samples and overlap of 64. Results accurate down to ~2 Hz.

y = 1 for this average.

Spectrogram - A recording of class 2 (y=1)



Visualization of a single recording of 2000 values from **sensor C3** with an FFT window size of 512 samples and overlap of 64. Results accurate down to ~2 Hz.

y = 1 for this average.

Competition Results

The performance measure is kappa value. The first column shows the average across all subjects, columns 2 to 10 show the results for the individual subjects.
-> Note: The expected kappa value, if classification is made by chance, is 0. <-

#.	contributor	kappa	1	2	3	4	5	6	7	8	9	research lab	co-contributors
1.	Zheng Yang Chin	**0.60**	0.40	0.21	0.22	0.95	0.86	0.61	0.56	0.85	0.74	Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore	Kai Keng Ang, Chuanchu Wang, Cuntai Guan, Haihong Zhang, Kok Soon Phua, Brahim Hamadicharef, Keng Peng Tee
2.	Huang Gan	**0.58**	0.42	0.21	0.14	0.94	0.71	0.62	0.61	0.84	0.78	School of Mechanical Engineeing, Shanghai Jiao Tong University, China	Liu Guangquan, Zhu Xiangyang
3.	Damien Coyle	**0.46**	0.19	0.12	0.12	0.77	0.57	0.49	0.38	0.85	0.61	Intelligent Systems Research Centre, School of Computing and Intelligent Systems, Faculty of Computing and Engineering, Magee Campus, University of Ulster, UK	Abdul Satti, Martin McGinnity
4.	Shaun Lodder	**0.43**	0.23	0.31	0.07	0.91	0.24	0.42	0.41	0.74	0.53	E+E Engineering, University of Stellenbosch, South Africa	Johan du Preez
5.	Jaime Fernando Delgado Saa	**0.37**	0.20	0.16	0.16	0.73	0.21	0.19	0.39	0.86	0.44	Robotica y Sistemas Inteligentes, Universidad del Norte, Colombia	
6.	Yang Ping	**0.25**	0.02	0.09	0.07	0.43	0.25	0.00	0.14	0.76	0.47	Perception-Motor Interaction Lab, School of Life Science and Technology, University of Electronic Science and Technology, China	Xu Lei, Yao Dezhong

Taken from: http://bbci.de/competition/iv/results/

Kappa Coefficient (Cohen’s kappa)

Measures “inter-rater reliability” for categorical items.

Though to be more robust than % accuracy because it takes into account the possibility of the agreement occuring by chance.

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

where p0 is the accuracy and pe is the hypothetical probability of chance agreement

For k categories, N observations to categorize and n_{ki} the number of times rater i predicted category k:

$$p_e = \frac{1}{N^2} \sum_k n_{k1} n_{k2}$$

Kappa ranges

- kappa == 0: predicted classes show no correlation with the actual classes
- kappa == 1: perfect classification
- kappa < 0: classifier suggests a different assignment between the output and the true classes

Schlögl A., Kronegg J., Huggins J., Mason S. (2007b). “Evaluation criteria in BCI research,” in Toward Brain-Computer Interfacing, Chapt. 19, eds Dornhege G., Millán J., Hinterberger T., McFarland D. J., Müller K.-R. (Cambridge: MIT Press;), 327–342

https://books.google.com/books?id=V88swGX83ecC&lpg=PA327&ots=mY2JAx5J8Y&dq=Evaluation%20criteria%20in%20BCI%20research%20pdf&lr&pg=PA328#v=onepage&q&f=false

Winning Approach

- The authors removed the EOG with a bandpass filter and extracted their features via a Filter Bank CSP (FBCSP) using mutual information rough set reduction (MIRSR). Classification of selected CSP features was performed using the Naïve Bayes Parzen Window classifier. A more detailed explanation of the winning algorithm is given in a separate paper (Ang et al., 2012). [5]

Second best

- The EEG was bandpass filtered in different frequency bands and the EOG artifacts were removed afterwards. Common spatial subspace decomposition (CSSD) were extracted from the preprocessed signals with optimized window sizes and an LDA discriminant function was made for each time point. [5]

Third Best

- CSP on spectrally filtered neural time series prediction preprocessing (NTSPP) signals was applied to all signals all subjects using the self-organizing fuzzy neural network (SOFNN). Furthermore the log variance of each filtered channel was calculated with a 1-s sliding window. The best classifier among 3 variants of LDA and 2 variants of SVM was chosen for each subject individually. [5]

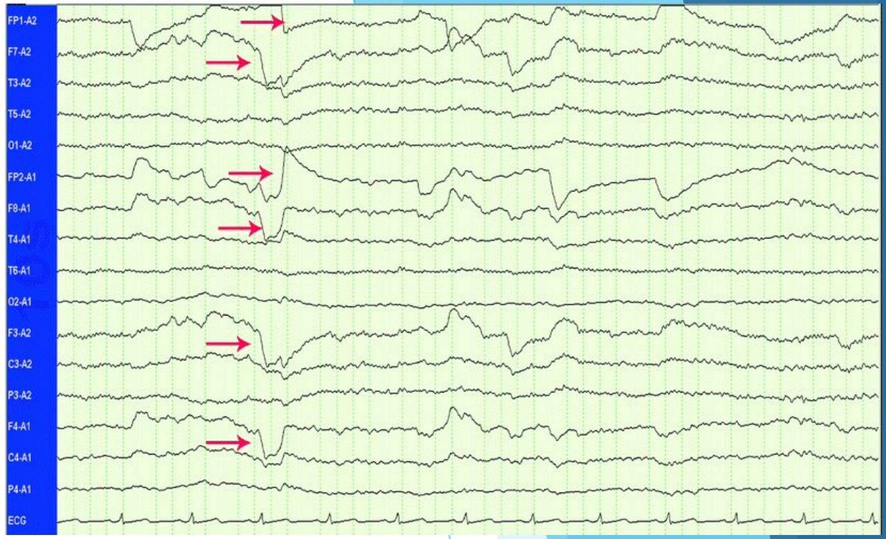
Fourth Best

- Wavelet packet transform was applied only on electrodes C3 and C4 (Cz was ignored). Selected frequency bands were extracted and concatenated to form a multidimensional vector and classified with LDA. [5]

Fifth Best

- EOG was removed with linear regression and the signals high pass filtered with 4 Hz. The algorithm used spectral features in the mu and beta bands (from electrode C3 and C4) as inputs for a neural network classifier. [5]

Discussion

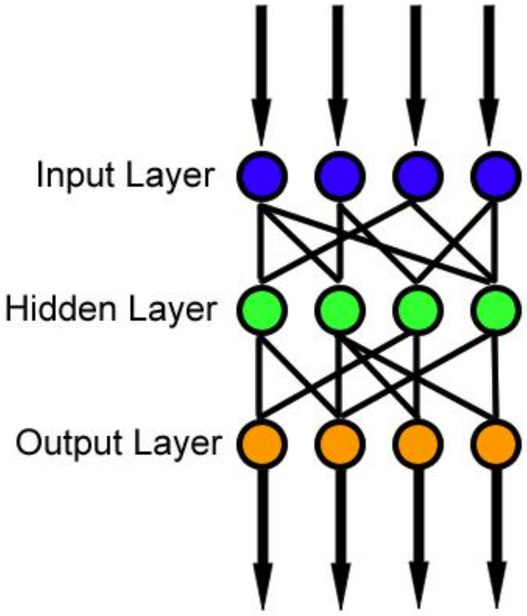


- The dataset presented two major challenges:
  1. The influence of eye movement artifacts on the EEG
  2. The generalization of the selected features to be successful on the session-to-session transfer

No method achieved good results on all subjects. Especially the session-to-session transfer could have been a source of problems. Although we provided training data sets from 3 different days, 2 training data sets were recorded without feedback and just 1 dataset with feedback was given. Of course we wanted to see the performance on online data sets with feedback recorded on different days. The winning algorithms could address this problem best, but their method had a 2-s inference delay to achieve its maximum performance. This approach is very useful for offline classification but cannot be used for real-time control. [5]

Approach 3: Feed-forward Neural Network (FFNN)

- A feed-forward neural network works by propagating a set of inputs forward one layer at a time
- At each node, the input gets modified by a linear function $y = ax + b$ and then an activation function is applied to introduce non-linearities
- As a result, with enough neurons, the FFNN can interpret complex non-linear features
- However, it cannot interpret time-series data and needs a spectrogram layer like the CNN



Abstract FFNN architecture.

Image Credits: https://en.wikipedia.org/wiki/File:Feed_forward_neural_net.gif#/media/File:Feed_forward_neural_net.gif

My model - Architecture and Hyperparameters

MLP with RELU activation. There were two versions of the model, one with a spectrogram layer and one without.

Hyperparameters: Dropout rate, number of hidden neurons per layer, number of hidden layers. Tuned using hyperopot.

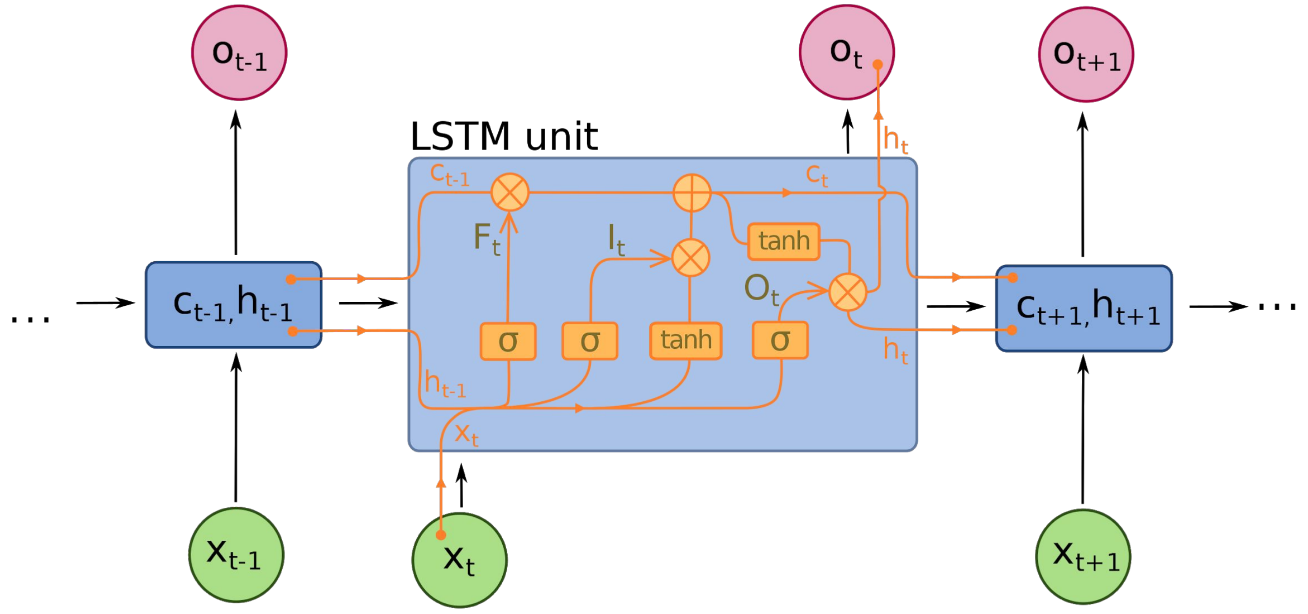
Hyperparameter values: {'dropout': 0.48519681298668427, 'n_hidden_layers': 0, 'neurons_per_layer': 500.0}

Accuracy: 55.16% with kappa of 0.07.

Spectrogram version: 2 hidden layers, 100 neurons per layer and 0.2 dropout. Accuracy of 83.44% and kappa of 0.64.

Approach 2: Recurrent Neural Network (RNN)

- A recurrent neural network can interpret time-series data because it includes a “memory” abstraction in its architecture
- As a result, it should be able to interpret the raw audio signal



An LSTM, a type of recurrent neural network. Notice that it takes into account different time slots, denoted by C_{t-1} and C_t.

Image credit: https://commons.wikimedia.org/wiki/File:Long_Short-Term_Memory.svg#/media/File:Long_Short-Term_Memory.svg

My model - Architecture and Hyperparameters

LSTM arbitrary number of hidden layers.

Hyperparameters: recurrent dropout rate, number of hidden neurons per layer and the number of hidden layers. Ttuned using hyperopot.

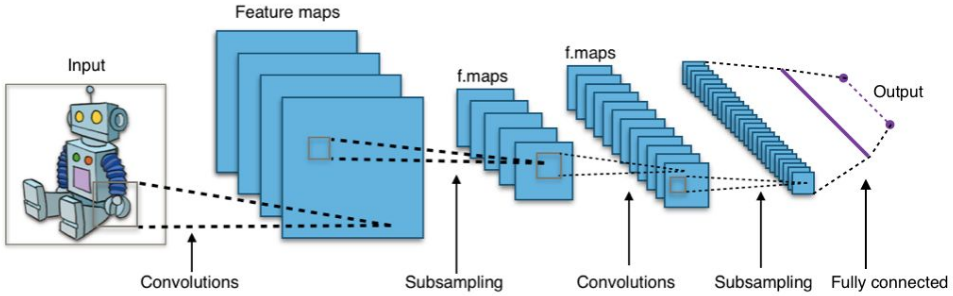
Hyperparameter values: {'dropout': 0.5650525229186382, 'n_hidden_layers': 0, 'neurons_per_layer': 140.0}

Which produced accuracy of 57% and kappa of 0.2

This model was not constructed with a spectrogram layer.

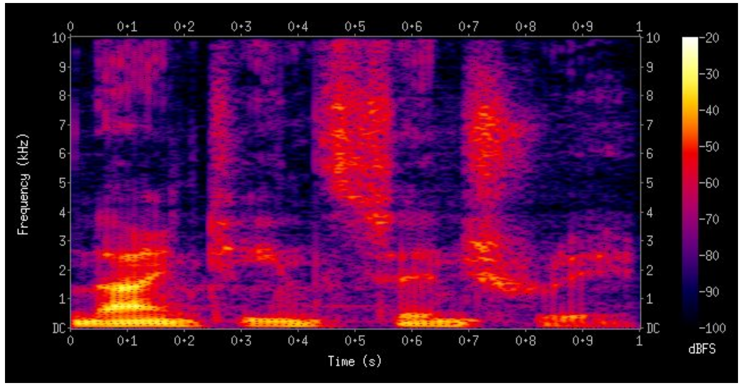
Approach 1 – Convolutional Neural Network

- A convolutional neural network can learn complex attributes and is most applied to image recognition
- However, it is not designed to interpret time series data
  - Workaround: convert the input to a spectrogram



Abstract CNN architecture.

Image Credits:
https://commons.wikimedia.org/wiki/File:Typical_cnn.png#/media/File:Typical_cnn.png



Spectrogram:

https://commons.wikimedia.org/wiki/File:Spectrogram-19thC.png#/media/File:Spectrogram-19thC.png

My model - Architecture

CNN with three blocks. Each block has:

1. Convolutional layer
2. Batch normalization layer
3. Max pooling layer

There were two versions of the CNN, one with a spectrogram layer and one without.

Simplifying Actions:

- Kernel size will remain the same across blocks

Hyperparameters considered:

- Number of filters for each block
- Kernel size across blocks

My Model - Hyperparameters and Results

Hyperparameter tuning: *hyperopot* with Bayesian optimization. A range of candidate values was defined for each parameter. 5-fold cross validation was used per trial. Hyperopot evaluation metric: average validation accuracy.

Hyperparameter Values: {'filter1_s': 90.0, 'filter2_s': 40.0, 'filter3_s': 60.0, 'kernel_size': 6.0}

Produced an average validation accuracy of: 82% and kappa of 0.62

With the spectrogram, the accuracy remains within margin of error but the kappa improved to 0.81.

The spectrogram CNN network was tuned by Mohammad Reza Razaei.

Results Summary - Average Cross-validation Accuracy

Go To Summary

Alternative:
https://docs.google.com/spreadsheets/d/1xp5YxuAmCuwWML72FYkpUKHyVBL0qAUsKWlMgg
u4S_4/edit?usp=sharing

Resources

Code:

https://github.com/panargirakis/ML_Extra_Credit

Competition (make sure took under dataset Graz 2b):

Intro: http://www.bbci.de/competition/iv/

Dataset description: http://www.bbci.de/competition/iv/desc_2b.pdf

Results: http://www.bbci.de/competition/iv/results/index.html

Results overview paper: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3396284/

References

1. R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, G. Pfurtscheller. Brain-computer communication: motivation, aim, and impact of ex- ploring a virtual apartment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 15, 473–482, 2007.
2. M. Fatourechi, A. Bashashati, R. K. Ward, G. E. Birch. EMG and EOG artifacts in brain computer interface systems: a survey. Clinical Neurophysiology 118, 480–494, 2007.
3. A. Schl ̈ogl, J. Kronegg, J. E. Huggins, S. G. Mason. Evaluation criteria in BCI research. In: G. Dornhege, J. d. R. Mill ́an, T. Hinterberger, D. J. McFarland, K.-R. Mu ̈ller (Eds.). Toward brain-computer interfacing, MIT Press, 327–342, 2007.
4. A. Schl ̈ogl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. Clin.Neurophys. 2007 Jan;118(1):98-104.

References – Cont'd.

5. Tangermann M, Müller KR, Aertsen A, et al. Review of the BCI Competition IV. *Front Neurosci*. 2012;6:55. Published 2012 Jul 13. doi:10.3389/fnins.2012.00055