Principal Component Analysis on a time series of Micro Electrode-Array recordings of Dopaminergic neurons

A study of energy and spiking activity

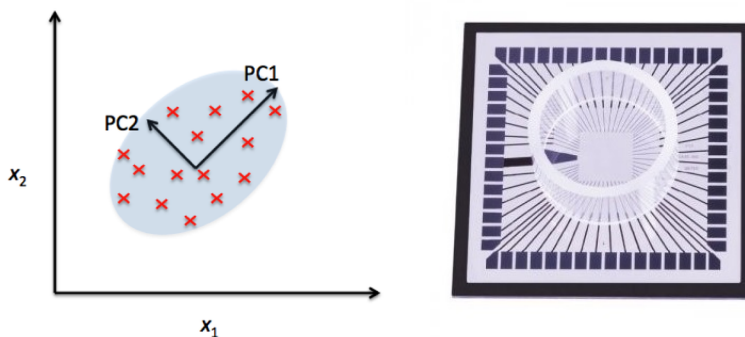## Ivar Thokle Hovden

Supervisors:

Frank Ove Westad

Co-supervisors:

Beatriz Galindo-Prieto

Martinius Knudsen

Ola Huse Ramstad

Vibeke Devold Valderhaug

Project report in cooperation with NTNU Cyborg and St. Olav University Hospital

December 2017

Summary

Functional analysis on voltage recordings from electrodes from a developing dopaminergic neuronal microelectrode array (MEA) culture using Principal Component Analysis on Power spectral densities and interspike histograms revealed a spatial difference in power variation and spike activity in the MEA. Additionally, high frequency power increase for $f \in [500, 2000]$ Hz as well as decreasing power trends at low frequencies $2 - 4$ Hz and at the beta-band $f \in [13, 30\rangle$ Hz along with the age of the culture were observed from four experiments spanning five months. Existing research report that synchronized beta-band oscillations are observed in patients with Parkinson's disease (PD), and that this is related to dopamine depletion. It is hypothesized that the observed power decrease in the beta-band along with the likely increase of number of dopaminergic neurons is possibly related to PD where the cells have developed into mature "less Parkinson's like" cultures. Evidence of a transition from in-phase activity (oscillations) to anti-phase activity on low frequencies along with the age of the culture is also presented with video and sound examples. Understanding mechanisms behind this transition (possibly temporal and rate neural coding), as well as mehanisms behind the observed spatial activity difference and power changes with respect to varying stimulation, could help in building better brain-machine interfaces, implementing biological neural network closed-loop control of the robot in the NTNU Cyborg project as well as help in the development of cell therapy for PD.

Table of Contents

List of Tables

List of Figures

Abbreviations

PD	=	Parkinson's disease
PCA	=	Principal Component Analysis
PC	=	Principal Component
PSD	=	Power Spectral Density
ISI	=	Interspike interval
SVD	=	Singular Value Decomposition
DFT	=	Discrete Fourier Transform
FFT	=	Fast Fourier Transform
STFT	=	Short-Time Fourier Transform
in-vitro	=	"in the glass"
in-vivo	=	"within the living"
DA	=	Dopamine
DAergic	=	Dopaminergic
mDA	=	Midbrain dopaminergic
iPSC	=	Induced pluripotent stem cell
hiPSC or hiPSC	=	Human pluripotent stem cell
MEA	=	Microelectrode Array
EEG	=	Electroencephalography
MEG	=	Magnetoencephalography
ECoG	=	Electrocorticography
LFP	=	Local Field Potential
SUA	=	Single-unit activity
MUA	=	Multi-unit avtivity
EAP	=	Extracellular action potential
IAP	=	Intracellular action potential
SNc or A9	=	Substantia nigra pars compacta
SNR	=	Signal-to-noise ratio
HHT	=	Hilbert–Huang transform
IMF	=	Intrinsic mode function
EMD	=	Empirical Mode Decomposition
MEMD	=	Multivariate Empirical Mode Decomposition
EEMD	=	Ensemble Empirical Mode Decomposition
CWT	=	Continuous wavelet transform
PLS or PLSR	=	Partial least squares regression
N-PLS	=	Multilinear Partial least squares regression
MVA	=	Multivariate analysis
AI	=	Artificial Intelligence
AGI	=	Artificial General Intelligence
ANN	=	Artificial Neural Network
GPU	=	Graphics processing unit
ML	=	Machine learning
DL	=	Deep learning
RL	=	Reinforcement learning
DRL	=	Deep reinforcement learning
BCI	=	Brain–computer interface
HMM	=	Hidden Markov Model
DBN	=	Dynamic Bayesian Networks
MDP	=	Markov decision process
VIP for OPLS	=	Variable influence on projection (VIP) for orthogonal projections to latent structures
TE	=	Transfer entropy
DCM	=	Dynamic casual modeling
GC	=	Granger casuality
fMRI	=	Functional magnetic resonance imaging
SCP	=	Slow Cortical Potential
ICA	=	Independent component analysis
SVM	=	Support Vector Machine
LDA	=	Linear Discriminant Analysis
SMC	=	Sequential Monte Carlo / Particle filter
LQE	=	Linear quadratic estimation / Kalman filter
HCI	=	Human-computer interaction
AR	=	Augmented Reality

Chapter 1

Introduction

A study of variance of various types of electrode signals from an in-vitro neuronal culture (Arenas, 2015; MultiChannelSystems, 2017a) was done using Principal Component Analysis (PCA), a dimensional reduction method (Wold, 1987). A motivation for selecting PCA was to have a linear, transparent method of investigating the complex non-linear and non-stationary nature of neuronal signaling. This explorative analysis could find the most clear signaling properties in the data with respect to the variance. In this setting, PCA is an analysis of functional connectivity of electrodes (Friston, 2011).

The overall goal is to better understand signaling, dynamics, change and learning in biological neural networks. This provides an intermediate goal of improving the signaling in a closed-loop feedback system for movement control of a robot for the Cyborg project at NTNU (Aaser, 2017b; Potter, 2006). In this analysis, the signaling came from an in-vitro Microelectrode Array (MEA) hosting human induced pluripotent stem cell (iPSC)-derived midbrain dopaminergic (mDA) neurons.

The mDA neurons were engineered at the Sandvig laboratory, Department of Neuromedicine and Movement Science, NTNU, as part of ongoing research on Parkinson´s disease (PD). DA neurons are particularly vulnerable in Parkinson´s disease and get severely depleted with disease progression, resulting in functional deficits, including impaired motor control. Hence, studying the behavior of DA neurons could be interesting also for the Cyborg project.

1.1 Original problem description

The original problem description was written in Norwegian. An English summary follows

1.1.1 Norwegian

NTNU Cyborg er et tverrfaglig prosjekt som skal undersøke mulighetene for inkorporering av levende nerveceller og robot i et kognitivt system. Det forskes på hvordan to-veis lærende kommunikasjon mellom nerveceller og robot kan oppnås, med den hensikt å få frem intelligent oppførsel av roboten og dens tilhørende sanntids visualisering av aktivitet i nervecellene.

Det tas utgangspunkt i data fra nåværende lukket sløyfe-kommunikasjon mellom in-vitro nevrostamcellekultur i multielektrodearray (MEA) og en simulering av roboten. Dataen er 10 Hz spen-

ningsavlesinger målt fra 60 elektroder jevnt spredt over stamcellekulturen. I den lukkede sloeyfen tolkes avlesingene til bestemte aksjoner for den simulerte roboten. For eksisterende avlest data har stimulering av nervecellene vært en periodisk puls fordelt over 1 electrode, med økende puls frekvens desto nærmere en vegg den simulerte roboten befinner seg. Heuristikken har altså hittil dreiet seg om å unngå kollisjon med vegger for den simulerte roboten.

Utkast til arbeidsoppgaver er:

1. Multivariat analyse (MVA) på eksisterende data for undersøkelse av sammenhenger/korrelasjon mellom spenningsavlesinger fra elektroder. Kartlegging av andre metoder som kan være nyttige. Er noen elektroder mer interessante enn andre for tolkning av aksjoner? Gir dette belegg for videre utvikling med eksisterende heuristikk?

2. Bruke resultater fra kartleggingen til å forbedre heuristikk med tanke på utvikling i kulturen. Det kan dreie seg om bedre avlesing-aksjon-mapping, annen type målfunksjon/stimulering, andre typer stimuleringer. Blackbox-metoder for juksing frem til man oppnår whitebox. Det skal være mulig å kjøre systemet for ny avlesing av data.

3. Sette opp system for sanntidsprosessering av data. Fra batch-basert databehandling til sanntid databehandling. Mulig samarbeid med Idletechs AS.

Det kan tenkes at repetisjon av stegene 1 og 2 er nyttig for å forstå nevrokulturen. Oppgave 1 er mest egnet utkast for prosjektoppgave.

1.1.2   English summary of realized tasks

An investigation of intercorrelations of electrode signals described in 1., was done in this report using PCA on Power Spectral Densities (PSDs) and Interspike intervals (ISI), as well as some additional time-domain analysis.

Using of multivariate analysis (MVA) results to improve heuristics for controlling a moving robot in a simulated environment to not go into walls described in 2., is discussed in result-section 5.0.1 and is a natural extension of the work into a master's thesis.

3. discussing a batch-based real-time processing system is now regarded as part of realizing 2.

1. was suggested as primary the task in project work.

Chapter 2

Background

2.1 Overview of recording extracellular activity

The consensus regarding signals and recording techniques today when looking at extracellular brain activity is summarized in Figure 2.1. In short, the recording approaches are Electroencephalography (EEG), Magnetoencephalography (MEG), Electrocorticography (ECoG) and various kinds of Intracortical Neuron Recordings (Nicolas-Alonso, 2012). Although MEG is an intracellular recording technique, it is mentioned because its signals are comparable to EEG signals (Waldert, 2009).
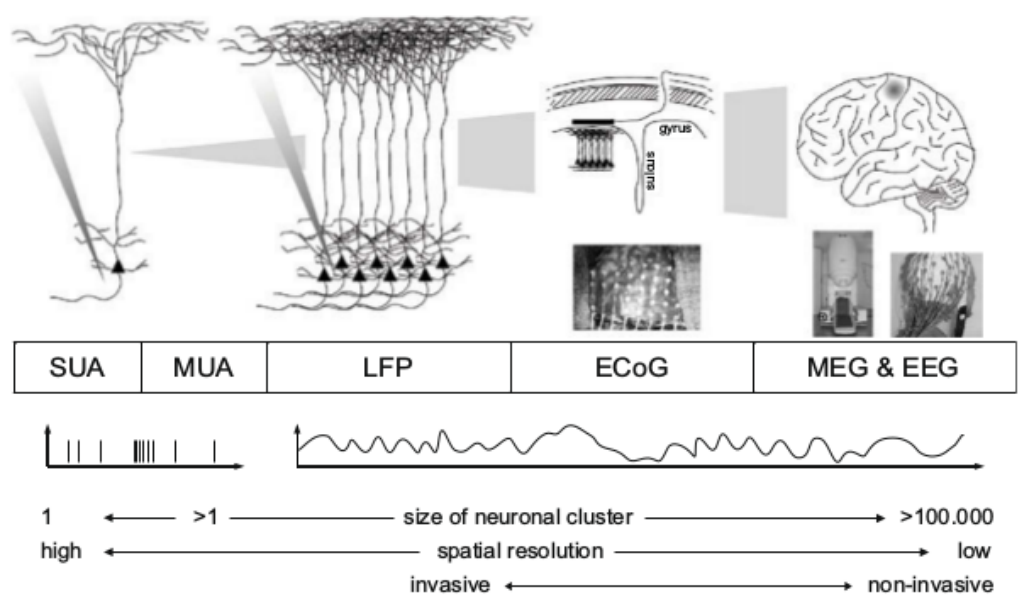
Figure 2.1: Various recording techniques and signals to measure brain activity. Taken from (Waldert, 2009). An in-vitro MEA such as MEA2100 used in this report, measures electrical currents on its electrodes. MEA measurements are a combination of LFPs and analog forms of SUAs and MUAs (or EAPs) (Obien, 2015; Waldert, 2009)

For intracortical recordings such as MEAs placed onto the brain after surgery and variants (in-

---

vivo, Viventi (2012) (ECoG), Maynard (1996)), Local Field Potentials (LFPs) are extracted by low-pass filtering ($f < 300$ Hz) of the neuron activity recorded from electrodes. Single-unit activity (SUA) signals are obtained by high-pass filtering ($f > 300$ Hz) signals of a single neuron, then spike detection and spike sorting. Multi-unit avtivities (MUAs) are obtained in the same way, but the signals may come from multiple neurons (Nicolas-Alonso, 2012; Waldert, 2009; Changkyun, 2016).

Extracellular Action Potentials (EAPs) are analog spike signals roughly in the frequency range $f \in [300, 3000]$ Hz caused by firing actions of single neurons (Obien, 2015; Pine, 1980). LFPs and EAPs are analog signals while SUA and MUA measure the spiking activity of single neurons and can be reduced to discrete events in time (Nicolas-Alonso, 2012). (Obien, 2015) shows by experiments a correlation of intracellular action potentials (IAPs) and EAPs. Assuming that IAPs and SUAs are closely related, EAPs could be seen as an analog version of SUAs.

(Obien, 2015) proposes that data measured from in-vivo and in-vitro MEAs are similar. For this reason, it is assumed that in-vivo brain signaling such as LFPs can also be measured from in-vitro MEAs containing comparable neuronal cells. MEA recordings is therefore considered as electrophysiological extracellular recordings of LFPs, EAPs and summations of EAPs (Obien, 2015; Waldert, 2009).

Measuring EAPs from in-vitro MEAs is well described in (Pine, 1980). In (Gold, 2006), an attempt to make a mathematical model of EAPs in order to draw another connection between intra- and extracellular recordings is made. (Bai, 2001) discusses a technique for recording SUAs using both in-vitro and in-vivo active MEAs.

EEG is the most common extracellular recording technique and consists of placing electrodes onto the scull (Buzsáki, 2012b), hence non-invasive. EEG and MEG signals are said to be smoothed versions of the local field potential (LFP) (Buzsáki, 2012b; Waldert, 2009).

EEG, MEG and ECoG measured signals are simply called EEG, MEG and ECoG signals, while Intracortical Neuron Recorded signals are Local Field Potentials (LFP), Multi-Unit activity (MUA) or Single-Unit activity (SUA) (Changkyun, 2016; Nicolas-Alonso, 2012). Opposed to the other signals mentioned, MEG reflects intracellular currents flowing through dendrites which produce magnetic fields measurable outside the head (Waldert, 2009).

2.2 Functions of dopaminergic neurons in animals and humans, and Parkinson's disease (PD)

In both animals and humans, the primary functions of DA neurons is in general related to motivation, and regulate approach behavior. The basic functionality is activation and inactivation of DA neurons, thereby regulating mood by inducing reward and aversion, respectively (Ikemoto, 2015). In the mammalian brain, activation of DA neurons such as the ones originating from the A9 group analysed in this report, leads to release of 3-hydroxytyramine (named dopamine), a metabolite of the amino acid tyrosine, into either of the four major dopaminergic (DAergic) pathways known in the mammalian brain (Beaulieu, 2011). Release of DA from midbrain DA neurons influences voluntary movement, cognition, and motivational state (Deignan, 2012). There is a diversity in types of DA neurons in humans and animals, and some subtypes even develop postnataly.

It is well known that lack of DA in the brain in humans is related to PD (Bernheimer, 1973). Other diseases related to inappropriate DA levels are Schizophrenia, Alzheimer's and Huntington's diseases (Seeman, 1987). Levels of DA have shown to have a role in both locomotor behavior and learning in humans (Beninger, 1983). Perhaps, in the simplistic biological view of the author, dopamine and dopaminergic neurons could be seen as the major functional parts in the biological neural system equivalent of reinforcement learning (see Kaelbling (1996)). (Holroyd, 2002) describes such a biological equivalent. Recent research has made it possible to generate human pluripotent stem cell (hPSC)-derived DA neurons, capable of inducing behavioral recovery in animal models of PD. This could help in development of cell replacement therapy as a future treatment in PD (La Manno, 2016). (Jo, 2016) even developed a method to differentiate hPSCs into a large multicellular organoid-like structure that contains distinct layers of neuronal cells expressing characteristic markers of human midbrain. They detected electrically active and functionally mature mDA neurons as well as DA production.

DAergic signaling in has been related to various kinds of cognitive tasks in the mammalian brain, which makes analysis of DA neurons interesting in the Cyborg perspective. Some examples are:

- (Barter, 2015) reports from in-vivo experiments on movement of rats identification of DA SNc neurons representing vector components of velocity or acceleration in the directions up, down, left and right, hence showing DA neurons belonging to a velocity control circuit. They hypothesized that DA signaling implements gain adjustment for adaptive transition control (Barter, 2015). This is discussed in result section 4.2.

- (Howe, 2016) reports a new optical method of in-vivo recording rapid signaling in distinct dopaminergic axons during locomotion and reward on mice, suggesting that DA neuro-modulation can differentially impact motor control and reward learning with sub-second precision, and that the results indicate that both precise signal timing and neuronal subtype are important parameters to consider in the treatment of DA-related disorders.

- (Jay, 2015) conducted an in-vivo study of endogenous activity using paired supraspinal DA neuron and motoneuron recordings on awake, paralyzed zebrafish, reporting that show that supraspinal DAergic neurons generate two forms of output: tonic spiking and phasic bursting. Tonic spiking being related to inactivity, bursting correlating to locomotor output.

- (Gadagkar, 2015) even reports on the task of DA neurons in the basal ganglia in the zebra finch (a bird) on giving auditory feedback on the magnitude of distortion in its bird song, as well as relating to movement.

- (Wagenaar, 2005) investigates stimulation of in-vitro rat DAergic neuronal MEA cultures in a closed-loop feedback system in order to control spontaneous bursting behavior. The work is discussed in result section 4.2.

2.3 Brain frequency bands

The frequencies analysed have been systematized into conventional EEG brain bands, in an attempt to relate findings of MEA signals to traditional EEG signals and other recording approached described in section 2.1 and previous work (Hovden, 2017). A table presenting the EEG bands (MultiChannelSystems, 2017b) as well as an unspecified band is shown in Table 2.1.

Hz	Band
$[0.5, 4\rangle$	delta
$[4, 8\rangle$	theta
$[8, 13\rangle$	alpha
$[13, 30\rangle$	beta
$[30, 120\rangle$	gamma
$[0, 0.5\rangle$ and $[120, \infty\rangle$	unspecified

Table 2.1: The typical EEG frequency bands with an additional "unspecified" band.Taken from (Hovden, 2017)

(Nácher, 2013) makes it clear that coherent oscillations on theta, alpha, beta and gamma frequency range have been proposed as mechanisms that coordinate neural activity in large-scale cortical networks in sensory, motor, and cognitive tasks. They propose in this work that coherent delta-band oscillations between cortical areas correlate with decision making in monkeys. All in all, oscillations in any of the EEG bands have been found to relate to different tasks in both humans and animals.

Alpha and gamma oscillations have previously been characterized as feedback and feedforward processing, respectively, in the monkey visual cortex, making a connection between oscillations in these bands regarding visual processing (Kerkoerle, 2014).

For further review on associated functions of EEG-band oscillations, see (Nicolas-Alonso (2012), p. 1216 in section 2.1. "Electroencephalography (EEG)").

Chapter 3

Method

3.1   Making the neurons - neuronal cell differentiation

The in-vitro midbrain DAergic neuronal culture was made following a procedure of neuronal cell differentiation from human induced pluripotent stem cells (hiPSCs) at Axel and Ioanna Sandvigs' laboratory of Translational and Regenerative Neuroscience at the Department of Neuromedicine and Movement Science (INB), Faculty of Medicine and Health Sciences, NTNU. The protocol for differentiating iPSCs into dopaminergic neurons was an adaption from several other protocols (Kirkeby, 2013, 2017; Doi, 2014).

The seeding, lab recordings and maintaining of the culture was conducted by PhD students Vibeke Devold Valderhaug, Ola Huse Ramstad and Rosanne van de Wijdeven. The seeding was a 15 day long specific protocol of adding various neuronal property enhancing ingredients. From day 16 and on, it has received a weekly feeding with neuronal maintenance media.

The resulting population was a mixed cell culture containing a specific type of dissociated human mDA neurons (naturally existing in substantia nigra pars compacta (SNc, A9 group)), which are known to have a selective vulnerability in PD (Arenas, 2015; Barker, 2015). To study their electrophysiological properties, the cells were maintained on a MEA2100 MultiChannel Systems MEA (MultiChannelSystems, 2017a) as seen in Figure 3.1c. When the activity was not recorded, the MEA was placed in an incubator with appropriate living conditions.

The neuronal culture was seeded on 11. of October 2016, and is still alive at the delivery of this report (December, 2017). Two experiments relating to Parkinson's research was done November 11th and 16th 2016, where DA was added. It was nicknamed "Dopey" by the PhD students from its characteristics and long-living property, and was registered as MEA 2.

3.2   Recording and stimulating MEA 2

Online raw and spike recordings were obtained using MultiChannel Experimeter (MultiChannelSystems, 2017d). The raw and spike data for selected experiment #2 (explained later) is shown offline in MultiChannel Analyzer (MultiChannelSystems, 2017b) in Figure 3.2

MEA 2 was recorded in total 86 times over the duration of nine months as seen in Figure 3.3, each experiment typically resulting in 7-10 minute long analog voltage recordings of 60 electrodes in
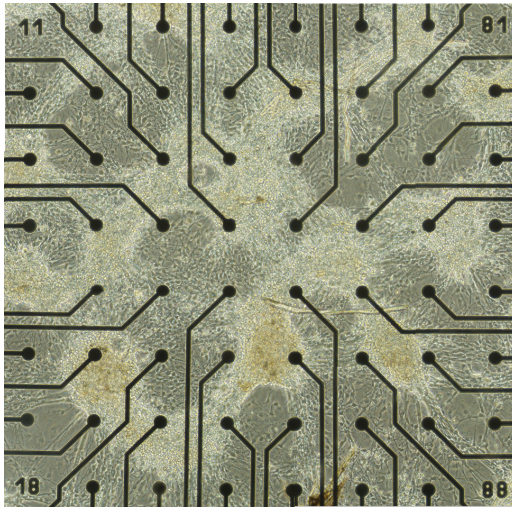
(a) MEA 2 day 46 (2016.11.25)
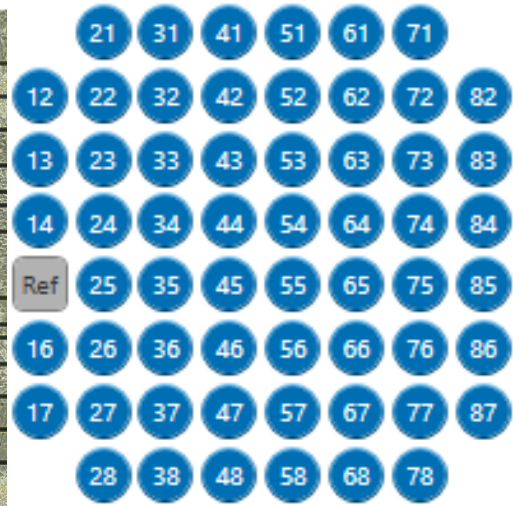
(b) The electrode numbering
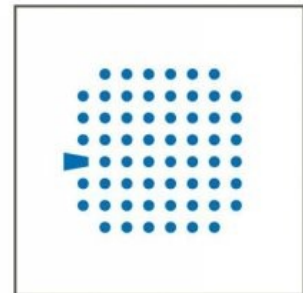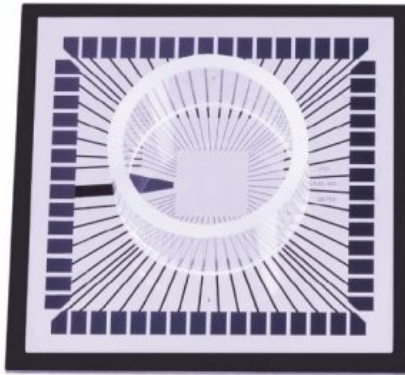
(c) The in-vitro MEA is illustrated in the leftmost figure. It is a small device that stores the cells inside a glass container and connects the neuronal culture to an electric circuit. The 60 electrodes are illustrated in the rightmost figure. The larger mark on the left side is the reference point, measuring a reference voltage

Figure 3.1: MEA2100
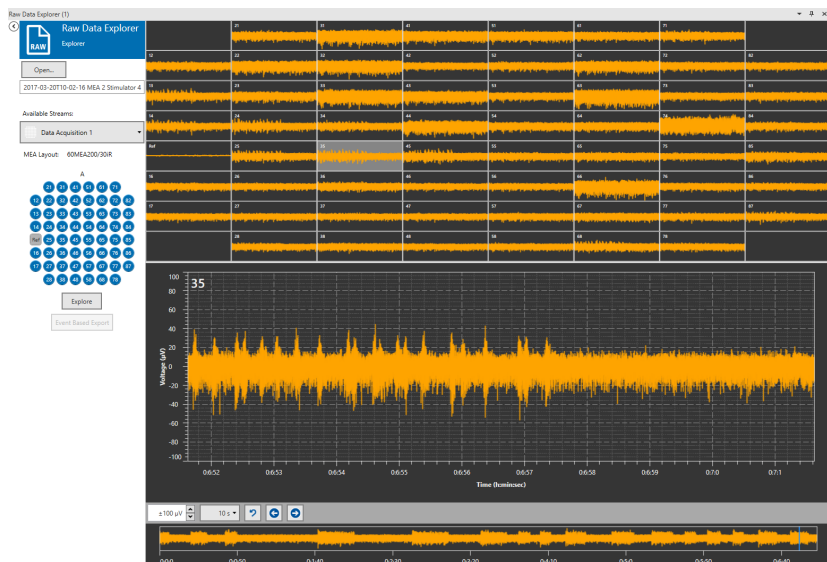
(a) Raw recording of #2 with electrode 35 marked. 35 shows some interesting busts behavior. Stimulation electrode 74 shows high amplitudes
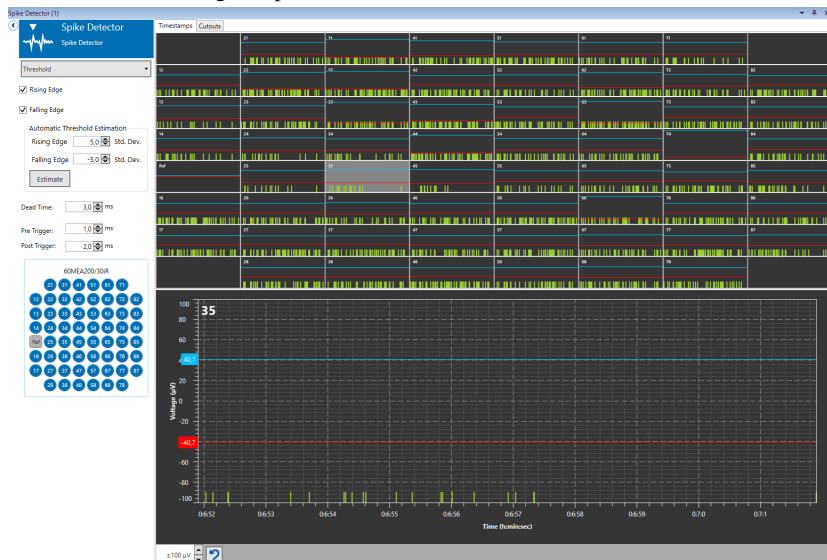
(b) Treshold spike detection of #2 with 35 marked. Spike detection as in 6.30 with 10 second window for estimation of standard deviation. The calculated spike tresholds are the blue and red lines

Figure 3.2: 10 second windows of the end of #2 from MCS Analyzer (MultiChannelSystems, 2017b). A weakness of the treshold-based spike detection can be seen. Although electrode 35 has a lot of distinct behavior during the end of a burst that is seen in this window, it actually gets fewer spikes detected opposed to non-bursting electrodes, f. ex. electrode 37. This is because spike detection tresholds are much larger in 35 because of larger standard deviation (40.7 upper, -40.7 lower treshold) than for 37 (28.4, -28.4). However, as seen later in Figure 3.16a, 35 has more variance in interspike intervals and also a more negative correlation to the low frequency high energy estimates 2 and 4 Hz, than 37

the piko voltage (pV) range. The sampling rate was 10000 Hz. The available recordings were from 17. of October 2016 to 16. of June 2017.

In most of the analyzed recordings (#1-#3), only electrode 74 received an external voltage stimulation by stimulation protocol 4. Stimulation protocol 4 was an electrical bi-phasic pulse of $\pm 500\,\mu V$, each phase lasting $100\,\mu s$, applied on electrode 74 every $30\,s$ of reading.

Naturally, the recordings had unknown Signal-to-Noise Ratio (SNR). A MEA recording is denoted as an experiment from now on.

3.3   Experiment selection

Four experiments were selected for analysis as depicted in Figure 3.3 with numbering #1-#4
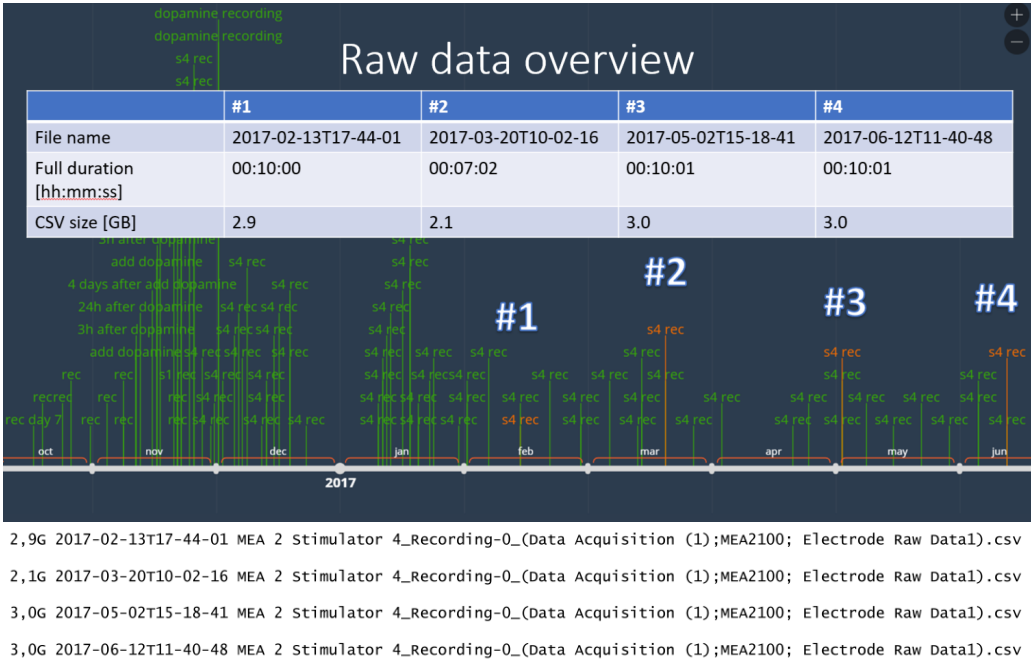
Figure 3.3: All available 86 experiments, as well as the four selected experiments. Each vertical line in the figure correspond to one experiment with 60 measurements. The four selected experiments marked in orange, containing raw and spike data. Experiment date, time duration and raw data size is displayed for the selected experiments. The height of the lines is purely there for readability. An s in the naming of an experiment should mean that a stimulation protocol was used during the recording. s4 is stimulation protocol 4. However, it has later become evident that #4 did not receive any stimulation. The four chosen experiments are marked in orange. The culture went trough two dopamine experiments in November, 2016, where dopamine was added. These experiments were not investigated

#1-#3 received stimulation by stimulation protocol 4, #4 did not receive any external stimulation (although the figure shows the experiment name "s4" for #4). After selection, the four experi-

ments were converted from their original raw files to csv format using Multi Channel DataManager, a tool from MultiChannel Systems (MultiChannelSystems, 2017c).

3.3.1 Simple statistics in time domain

An average min, max and mean for all electrodes in the selected experiments were calculated with Python and plotted in The Unscrambler X as seen in Figure 3.4. The python script is described in section 3.5.

(a) The average min and max of all electrodes

(b) The average of the average of all electrodes
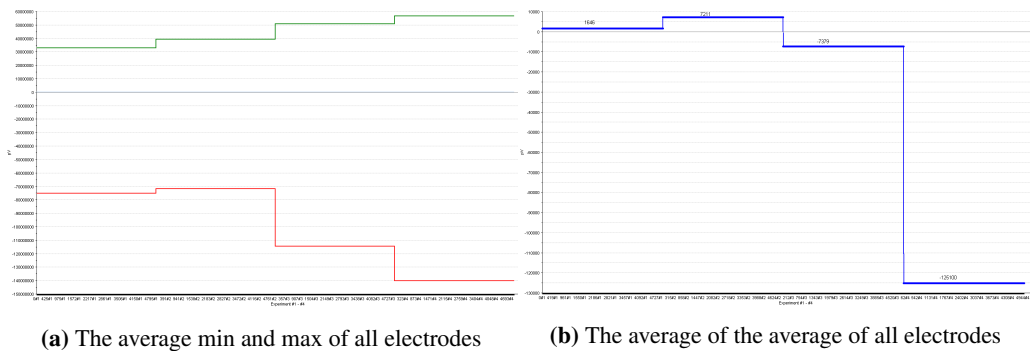
Figure 3.4: Average min, max, and average average of the time series data in #1-#4. y axis is pV, x axis is experiment. There was an increase in max and decrease in min. The average drastically decreased in #4. Note that #4 did not receive stimulation, unlike #1-#3

3.4 Principal Component Analysis on experiment #4 raw data

PhD student Martinius Knudsen did a PCA on similar MEA raw data, where a few electrodes were projected apart from the others in the loadings plot. For comparison, the same was done on a 30 second experiment taken just before #4 (the same day) by treating the analog raw data as the matrix 6.11 with $N = 30 \times 10000$ objects and $M = 60$ electrodes.

The results, seen in Figure 3.5 are very similar to Martinius' results. One PC expresses almost all the variance in this data. The reference electrode and electrode 16 are projected far away from the rest of the electrodes. 66 is also a bit different from the other electrodes. A quick look back at the raw data in MCS Analyzer reveals that electrode 16 and reference electrode have significant different amplitudes from the rest of the electrodes. A little more low frequencies might be seen on electrode 66.

This method of doing PCA was concluded to not give any significant results, simply because of the high sampling rate of the recordings. 30 seconds resulted in 300 000 objects in the score plot.
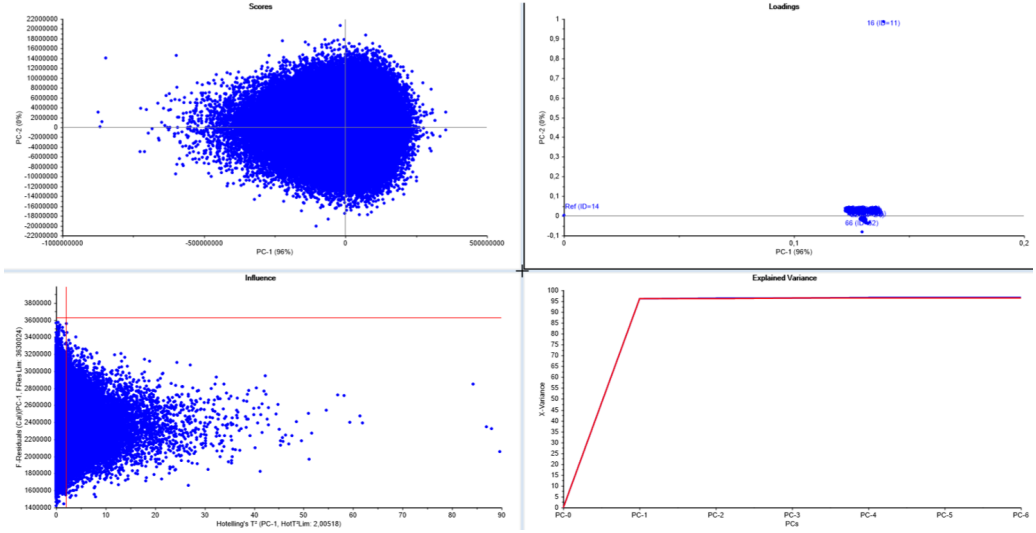
Figure 3.5: PCA on raw data of the 30 second experiment from 2017.06.12

3.5 Computation of Power Spectral Densities on experiments #1-#4

In order to have a more appropriate data format, Power Spectral Densities (PSDs) were calculated from the raw csv files for each of the experiments. This resulted in PSD data with dimensions $N \in [0, 5000]$ Hz and $M \in 60$ electrodes. The Python script is in the Appendix section 6.14. All the PSDs except in the frequency range $f \in [0, 10]$ Hz are visualized in Figure 3.6. The reason for only plotting $f \in [11, 5000]$ Hz was simply because of the impression that the low frequencies were outliers and had a lot of noise (from Figure 3.8). This has turned out to not necessarily be true (see Figures 3.16, 3.19).

Average PSDs over the electrodes for the four experiments were also calculated as shown in Figure 3.7.

3.6 PCA on PSD #1-#4

3.6.1 1-5000 Hz band

Initially, a PCA on a combined PSD data set from the four experiments was done on the entire frequency range $f \in [1 - 5000]$ Hz. Power on 0 Hz was discarded because if its high power. A zero Hz amplitude in the PSD simply shows the power of the offset in the signal, which was a lot higher than the other powers.

The computed PSDs 6.11 had the dimensions $N = [0, 5000] \times 4$ Hz (0 Hz is discarded in The

Figure 3.6: PSDs from the four experiments with 0-10 Hz removed. The left figure shows PSDs for each electrode on top of each other, for each of the four experiments. The y axis pV^2/Hz, x axis $11 - 5000$ Hz four times after another. The right figure is the same data, with Hz as objects and electrodes as variables. This is the input to combined PCA described in 3.6.2. There were low frequency energy peaks ($11 - 16$ Hz) in stimulation electrode 74 in #1-#3. 63 had peaks in #1-#3, but in decreasing frequencies (50, 30, 16 Hz). 54 had a peak at 894 Hz in #4. PSDs in the total frequency range ($0 - 5000$ Hz) were also calculated for the first PSD PCA in described in section 3.6.1

Figure 3.7: Average PSD for all 60 electrodes for the four experiments. The four sections in the x axis are frequencies $11 - 5000$ Hz. The y axis is pV^2/Hz. This shows clearly that there were significant energy increases in high frequencies

Figure 3.8: Scores from PCA on the four experiments selected as in Figure 3.3. This plot tries to summarize in two dimensions (Principal components (PCs)) what happened in terms of variation in energy (PSD) in the 60 dimensions (electrodes, see appendix 6.1.3). The x axis is the projection of energy at individual Hz onto the first principal component (PC1), which expressed 76 % of the variation in the PSDs. The y axis is the same but for principal component two (PC2). PC2 only expressed 6 % of the variation that were not expressed by PC1. In the object labeling, the first number is the frequency from PSD that have a power projecting it to its position. The other number is the experiment number from the experiment selection. Objects from the four experiments are colored. #1 in blue, #2 in red, #3 in green and #4 in light blue. It shows that very high frequencies (around 1000 Hz) in #4 contribute to a lot of the variance in the combined data set. Additionally, increasingly higher frequencies contribute to the variance over the four experiments. Low frequency activity seem to contribute less to the variance of PC1 over the experiments

Unscrambler X) objects and $M = 60$ electrodes. Weighting of the electrodes was done by dividing by standard deviation of their PSDs. Categorical cross-validation using experiment number as category. The reference electrode was not included in the PCA.

3.6.2   11-5000 Hz band

Information about low frequencies in figure 3.8 were combined with insights from empty MEA recordings recorded by Peter Aaser to propose that there were more noise apparent in the range $0 - 10$ Hz than in the rest of the frequency range. Recordings of empty MEAs (no cells inside) showed high amplitude variations in this range. This was the motivation for removing $0 - 10$ Hz and conducting the PCA on the resulting matrix 6.11 with dimensions $N = [11, 5000] \times 4$ Hz and $M = 60$ electrodes. The results are displayed in Figure 3.9.

However, important variation might have been missed when removing these frequencies. Fig-

ure 3.16 shows later that low spike frequencies were dominant in the variation of spike counts. Perhaps could the same spiking frequencies be related to the envelope frequencies of the actions displayed in Figure 3.19. In other words, the trend of decreasing energies in low frequencies from Figures 3.8, 3.13, 3.16, 3.17 and 4.2 could be related to the apparent de-synchronization of low frequencies in Figure 3.19b.

This time the electrodes were not weighted, from the insight that high energy variation electrodes should contribute to most of the variance (the impression of noise in low frequencies + a lot of activity in wide frequency band around 1000 Hz as shown in Figure 3.7). Again, categorical cross-validation with experiments as category was used.

In order to see the individual electrode contributions in each experiment separately, PCA was then conducted separately on each of the experiments. Again, no scaling of electrode weights was used. 10-segment random cross-validation was used on the individual experiment PCAs. Figure 3.10 shows line plots of PC1 loadings from the four individual PCAs.
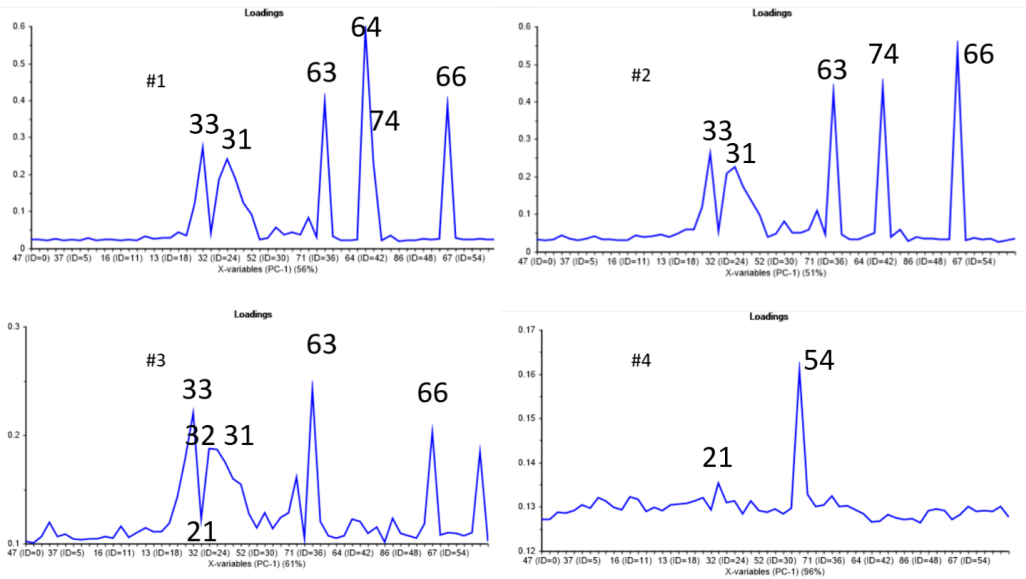
Figure 3.10: PC1 loadings from individual PCAs on #1-#4. Unlike the loadings plot in Figure 3.9, the line loadings plot displays the electrode contributions to energy variation expressed in a single PC. #4 is significantly different with electrode 54 as the main contributor to energy variance in its own PCA model. The single frequency causing an energy peak in 54 is later shown to be 894 Hz, see Figure 3.18b

Going back to the combined PCA model, average power object grouping was then used on score and influence plots for comparison with Figure 3.11.
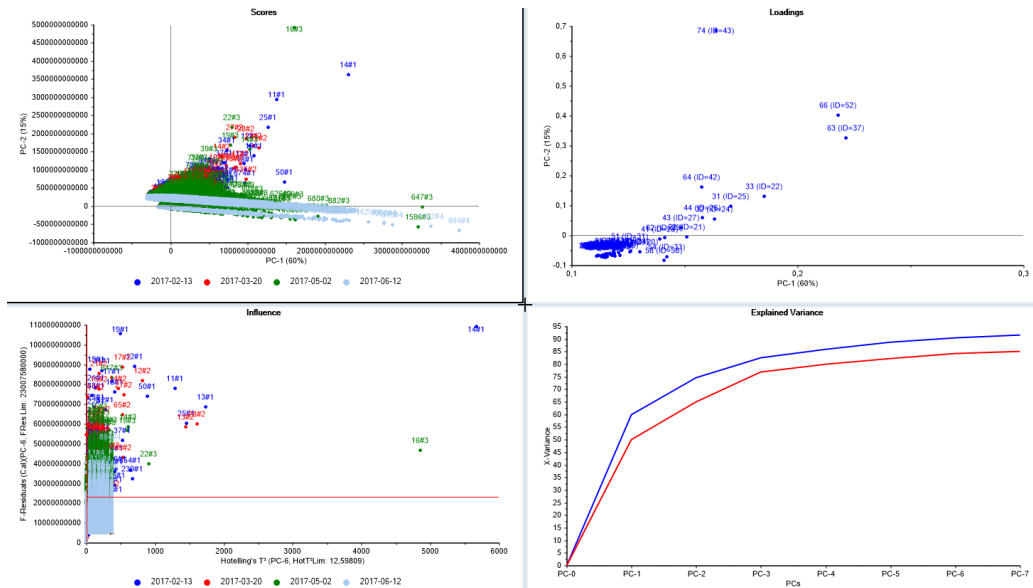
Figure 3.9: PCA on #1-#4, scores and influence plot grouped according to experiments. Upper left is the score plot explained in Figure 3.8 and in the Appendix section 6.1.3. The color grouping are the dates of the experiments. Upper right is the corresponding loadings plot for the same PCs, which displays the contribution of electrodes to the PSD variations along PCs. Electrode 66 and 63 have a large contribution to both variation in PC1 (60 %) and PC2 (15 %). Note that PC1 expressed less of the total variance when 0-10 Hz objects were removed. The lower left is the influence plot, which displays information of how good individual objects fit to the PCA model with up to a number of PCs included (here PC 1-6 included). Here, the x axis is Hoteling's T^2 metric (how well the object is described by the model. A scaled leverage or accumulated score), the y axis is the F-residual (object distance to the model). F-residuals are calculated from the calibration as well as the validated residual x-variance, see Appendix section 6.1.5). The red lines are critical limits from Hoteling's statistic (x axis) and F-residuals (y axis). Both critical limits are based on an F-test with a preset confidence interval (in the plots, always 95 %, or p-value 5 %) which gives a confidence in variance for the objects and can be used to detect outlier objects. Most of the interesting objects seem to be outliers according to these metrics. Most object within the confidence limits are from #4. Extensive information about interpreting PCA plots can be found in (CAMO (2017c), p. 553-581). 14 and 16 Hz from #1 and #3, respectively, are the two frequencies to singularly have the largest influence on the variation in the overall PCA model. Both have higher F-residuals than the F-test critical limit. However, 16 Hz fits better to the model than 14 Hz, since it has a smaller F-residual. As a confirmation 14 and 16 Hz are the two highest peaks in Figure 3.6 (left) from electrode 66 and 74 respectively. 14 has higher energy than 16, which is described in the model with a higher F-residual. Perhaps would frequencies in 0-10 Hz have large influence on the model as well, if they were not removed. The lower right is the explained variance plot described in the Appendix section 6.1.5, for the entire model (calibration, blue) and during model training (categorical cross-validation, red). Too large differences between blue and red are signs of overfitting of the PCA model, which is unwanted in statistical model building (especially when the model shall be used for prediction from new data). The x axis is number of PCs used in the model, the y axis is resulting explained variance in the data
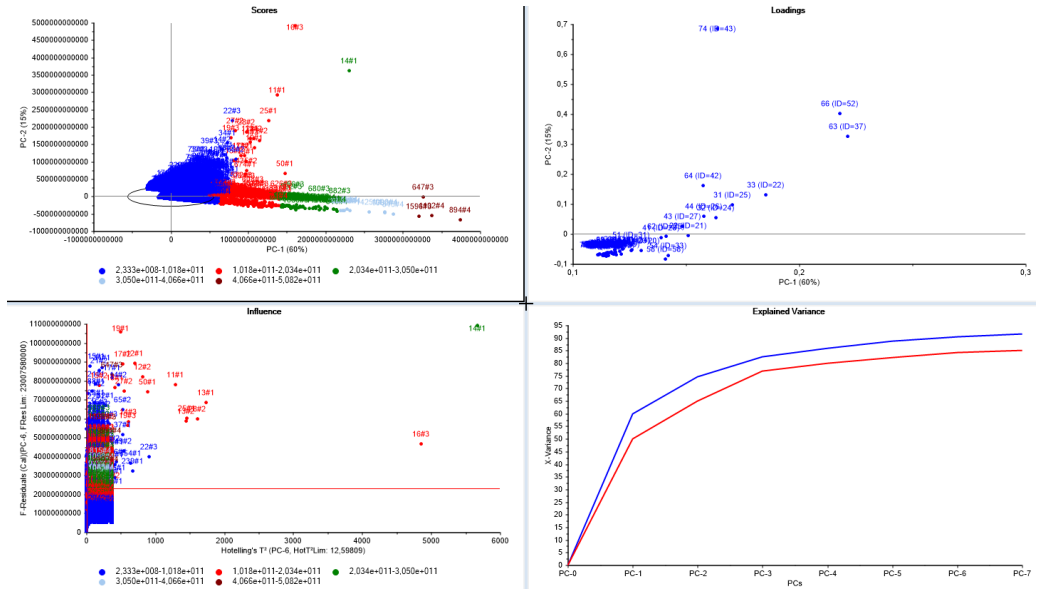
Figure 3.11: The same PCA as in Figure 3.9, but with five intervals of powers from the average PSD used to group the objects. Blue are the frequencies with the lowest average power over electrodes, while brown are the highest powers. Interestingly, objects (Hz) group similarly independently of the experiments. When compared with Figure 3.9 one can see that #4 had the most energy variation in terms of frequency

The individual PCA results on the most interesting experiments, #2 and #4, are shown in Figure 3.12.

3.6.3 Beta-band ($13 - 30$ Hz)

Previous work on MEA 2 shows decreasing beta-band $f \in [13, 30\rangle$ power, as well as increasing band power on $f \in [120, 5000\rangle$, from two second segments of #1-#4 (Hovden, 2017). A combined PCA on $f \in [13, 30]$ Hz from PSDs was done do further investigate beta-band energy behavior. Results can be seen in Figure 3.13.

3.7 Spike counts

In addition to raw data, threshold-based discrete spike data explained in the Appendix section 6.30 was available. The spikes were detected and recorded using the "Spike Detector" module in MultiChannel Experimeter (MultiChannelSystems, 2017d) by the neurobiology PhD students at Sandvigs' lab at the same time of analog raw recordings. The window length used for estimation of standard deviation is unknown, but not longer than 10 seconds since 10 seconds is the maximum data preview length in Spike Detector.

For each experiments, the number of discrete spike timestamps were counted for each electrode using Microsoft Excel. The resulting plots of accumulated spikes are shown in Figure 3.14.
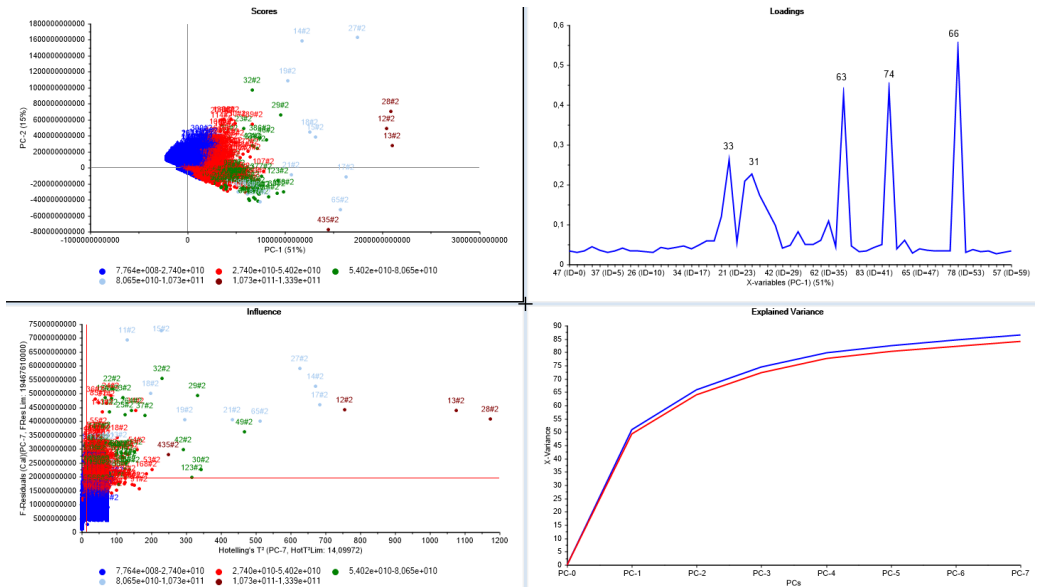
(a) #2. 51 % or the variance is expressed in PC1
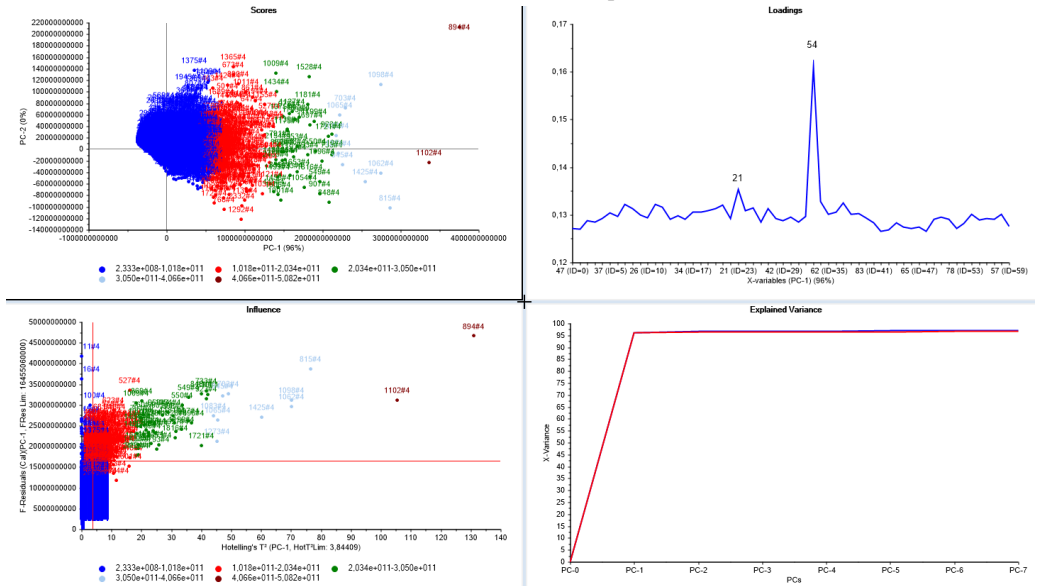
(b) #4. 96 % of the variance is expressed with PC1. The color grouping shows that PC1 describes the average power over electrodes of the frequencies

Figure 3.12: Comparison of PCA on #2 and #4 with average power object grouping as in Figure 3.11, with line loadings plots for PC1. It seems that the average power over PSD of electrodes are essentially PC1 in (b) and to some degree in also PC1 in (a)
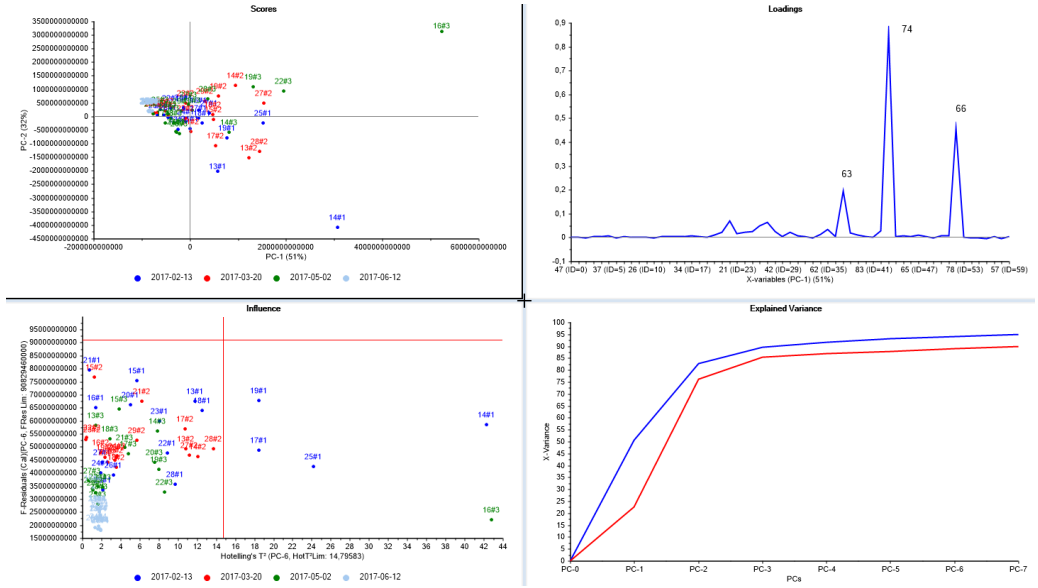
(a) With experiment object grouping and PC1 loading plot
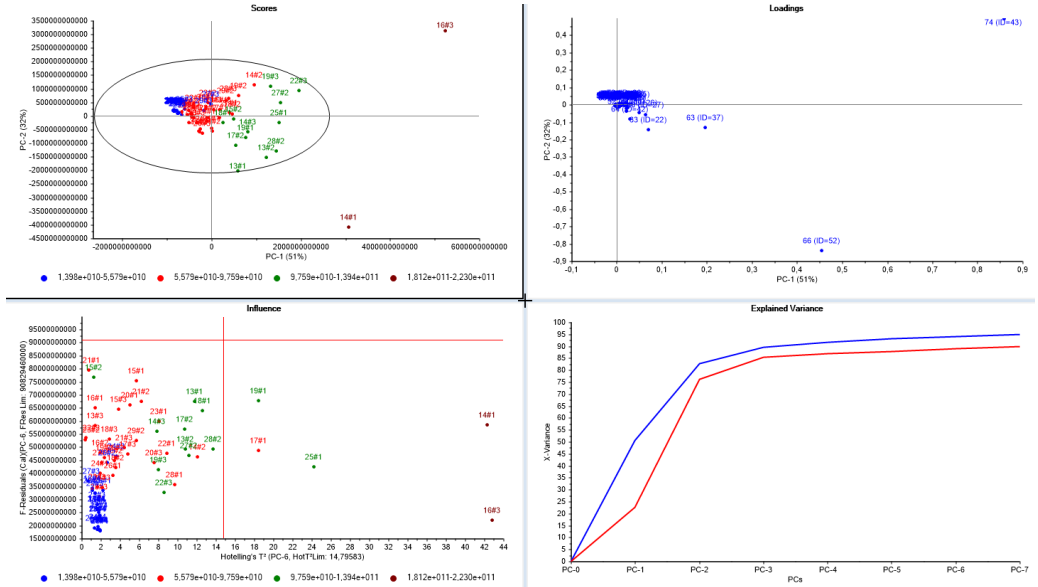
(b) With average power object grouping, PC1-PC2 loadings, and the Hoteling's T^2 ellipse on the score plot. Hoteling's T^2 ellipse is a multivariate version of Hoteling's T^2 statistic, and displays in this case which objects were within the 95 % confidence limit for Hoteling's T^2 with PC1 and PC2 selected

Figure 3.13: Combined PCA beta-band $(13 - 30$ Hz$)$ on #1-#4. By looking at the influence plots ((a), (b) lower left) it is clearly seen that #4 had the lowest energy in this band. The same 14 and 16 Hz as displayed in Figure 3.9 and 3.11 as highly influential objects, are above critical limits in the scores and influence plots. However, unlike the in the comparable models, they have not critical F-residuals and are therefore a valuable part of the model
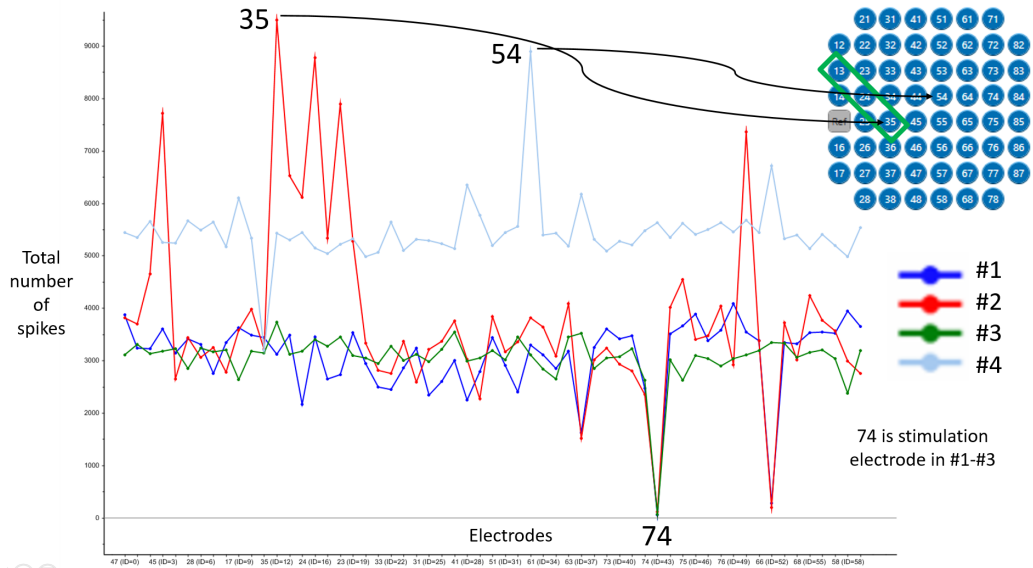
Figure 3.14: Spikes detected in the experiments from discrete timestamp data. #4 had generally the most spikes detected and a peak in electrode 54. 54 had also the most variance in power over frequencies from PCA analysis on #4 (see Figure 3.12b). #2 had high spiking activity on three electrodes close to the reference point in the MEA

3.8 PCA on interspike interval histograms

Due to the removal of energies in the 0-10 Hz range in the PSD analysis, it would be interesting to try PCA on lower frequencies using another type of data and then try to group these frequencies by average electrode power computed earlier from average PSDs.

The "Spike Analyzer" module in MultiChannel Analyzer (MultiChannelSystems, 2017b) can further analyse the detected spikes from the "Spike Detector" module (both MCS Analyzer and MCS Experimeter). The module has the ability to compute ISI histograms on each electrode from the raw data of the (7-10 min) experiments.

interspike interval (ISI) histograms show the distribution of ISI in fixed 10 ms bins for each channel. The maximum ISI displayed is 1000 ms (MultiChannelSystems, 2017b). Therefore, the histograms show spike counts over a distribution of 100 bins.

The histograms had to be generated from the raw voltage data. This required setting new tresholds for spike detection. The standard deviations were now calculated from 10 second window segments. Note that this might have lead to different total number of spikes detected than presented in Figure 3.14. The significant spiking frequencies were hopefully preserved. The decision of trying PCA on ISI histograms came from the impression that better temporal resolution on low frequencies might be derived from spike counts.

3.8.1 Estimation of spike frequency from ISI bins

An estimation of spiking frequency was calculated for each 10 ms bin in Microsoft Excel using the formula

$$BinFrequency_i[Hz] = \frac{1}{Avg(BinStartTimeMs_i, BinEndTimeMs_i) \times 10^{-3}} \qquad (3.1)$$

where $BinEndTimeMs_i - BinStartTimeMs_i = 10$ ms and $i \in 100$. For example, for the last bin, $BinFrequency_{100} = \frac{1}{((990+1000)/2) \times 10^{-3}} = 1.005$ Hz. The resulting frequency band was $f \in [1.005 - 200]$ Hz with uneven frequency spacing.

3.8.2 Histogram interpolation

The frequency spacing needed to be even in order to group the frequencies using the average PSD powers from earlier. This was achieved by importing the modified histograms (conversion to Hz as row header) into The Unscrambler X and using the "Interpolate..." transform with a step size of 2 on the imported data. The procedure was done on ISI histograms for each of the experiments. The results for #4 is seen in Figure 3.15.

3.8.3 Result

Individual PCAs were done on the interpolated histograms of #1-#4. For each experiment, the input matrix had dimensions $N \in [2, 200]$ Hz and $M \in 60$ electrodes. The results for #2 and #4 are shown in Figure 3.16. They can be compared to the high passed PSD PCA and beta-band PSD PCA in Figure 3.12 and 3.13. The objects were grouped by the the sections of average PSD shown in Figure 3.17.

PCA directly on ISI histograms as in Figure 3.15a would prove a better frequency resolution on especially $f \in [1, 10]$ Hz. However, because of the grouping of object according to energy and thereby connection to PSD analysis, the interpolated versions were used.
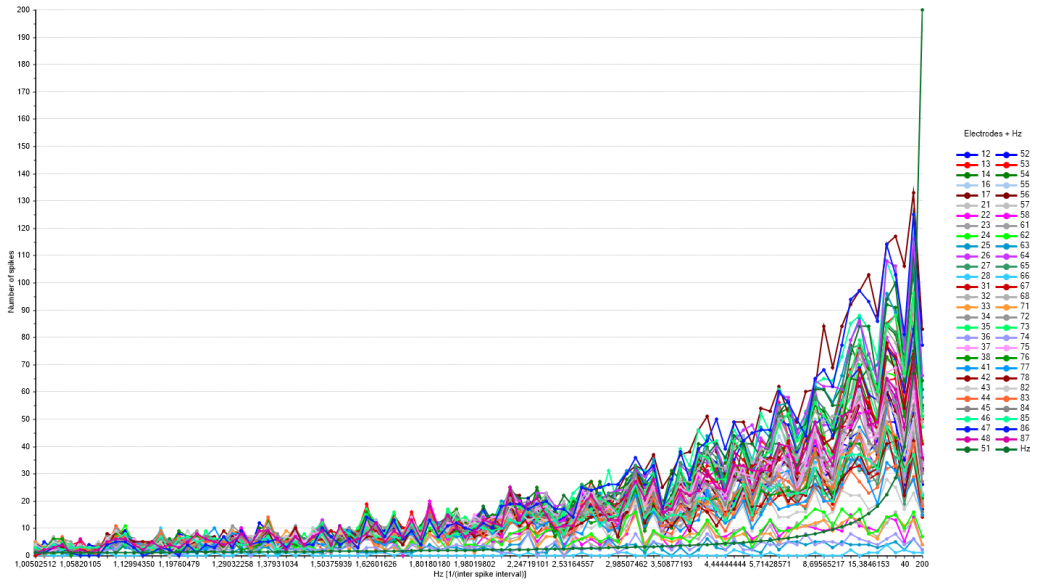
(a) ISI histograms with estimated bin frequencies. Frequency range $f \in [1.005 - 200]$ Hz with uneven step size. The estimated frequencies are plotted for reference to (b) (the green exponential line)
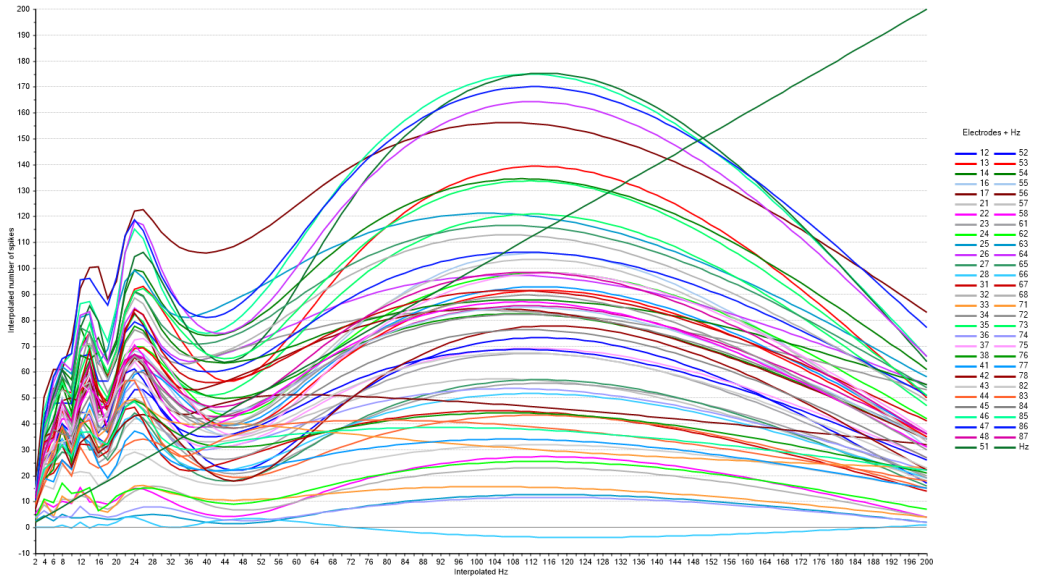
(b) The interpolated histograms. Frequency range $f \in [2 - 200]$ Hz with 2 as step size. By comparision with (a), most information lies in $f \in [2-32]$ Hz. Note that the height of the histograms are no longer spike counts, but stretched-out interpolations for integer frequencies. This was done in order to group objects using average electrode power from PSD in score and influence plot after PCA on the interpolated ISI histograms. A step size of 2 was the solution for having correct interpolation in The Unscrambler X software. The straight line of the plotted Hz verifies that the interpolation was done correctly

Figure 3.15: The original and interpolated interspike interval histogram of #4

(a) #2

(b) #4

Figure 3.16: PCA on interspike interval histograms of #2 and #4. Objects (Hz) are grouped in intervals based on their average electrode power (average PSD computed earlier). In the object labeling, the number is now the interpolated, estimated frequency from the interpolated ISI histograms that have a interpolated total spike count projecting it to its position. The objects that occur in smooth waves in the score plots are likely artifacts from the interpolation, and should be denoted as the least interesting frequencies. 2 and 4 Hz have outlier behavior in both experiments. 2 Hz also is the spike frequency with the highest power in both experiments (brown color). Blue objects have the lowest power. High power and influence of 2 and 4 Hz could indicate that these spike frequencies have high amplitudes and cause a major part of the detected spikes. Perhaps they are further caused by many simultaneous synchronized action potentials. The most spiking activity seem to not occur at the same electrodes that have the highest energy variation (Figure 3.12). 84 % and 96 % of the total variance are explained by PC1 in #2 and #4, respectively. #2 needs more PCs to explain the total variance

(a) #2

(b) #4

Figure 3.17: The average PSDs used to group the objects in Figure 3.16. $f \in [2, 200]$ sections of the total average PSDs calculated. 2 and 4 Hz have in total higher energies in #2 than in #4. However, 2 and 4 Hz are still dominant in the average energies in #4. The scale in the y axis in (a) is a lot larger in (b), so energies might seem to have decreased in this range. There is a peak in 100 Hz in #4

3.9 Looking back at data

From Figures 3.14, 3.13, 3.11 and 3.7, it has become evident that #2 and #4 contained the most interesting recordings. #2 in spike activity, #4 in energy variation. This has been investigated further in Figures 3.12 and 3.16, showing that PC1 from #4 is essentially expressing the average PSD and that low-frequent spiking behavior was still apparent in #4 although with lower power.

3.9.1 PSD object plots of #2 and #4

Figure 3.18 displays the full PSD data ($f \in [0, 5000]$ Hz) in order to look for clues on variation in power over electrodes.

3.9.2 Short-Time Fourier Transform on spiking electrodes of #2 and #4

From Figures 3.16, 3.14 and 3.10, it is clear that some electrodes had more spiking activity than others. An attempt to investigate spiking activity was done with Short-Time Fourier Transform (STFT) on noise reduced time versions raw signals of electrode 35 and 87 in 46 and 86 in #4.

First, the raw voltage data from the selected electrodes were converted to waveforms (sound) with equal (10000 Hz) sampling rate. Then, they were noise reduced using a selected noise profile from electrode 35 of #2. This was done in Audacity using the "Noise Reduction" plugin. The noise profile used was approximately the section of the recording of 35 between the first and second bursts. The first burst can be seen in Figure 3.2a. The parameters of the noise reduction method is described in Table 3.1.

The filtered sound was then inputted to a spectrogram analyzer (Baudline, 6.13.4) in Ubuntu using the command:

```
parec --format=s16le --channels=1 -d alsa_output.pci-0000_00_05.0.analog-stereo.monitor | baudline -stdin
```
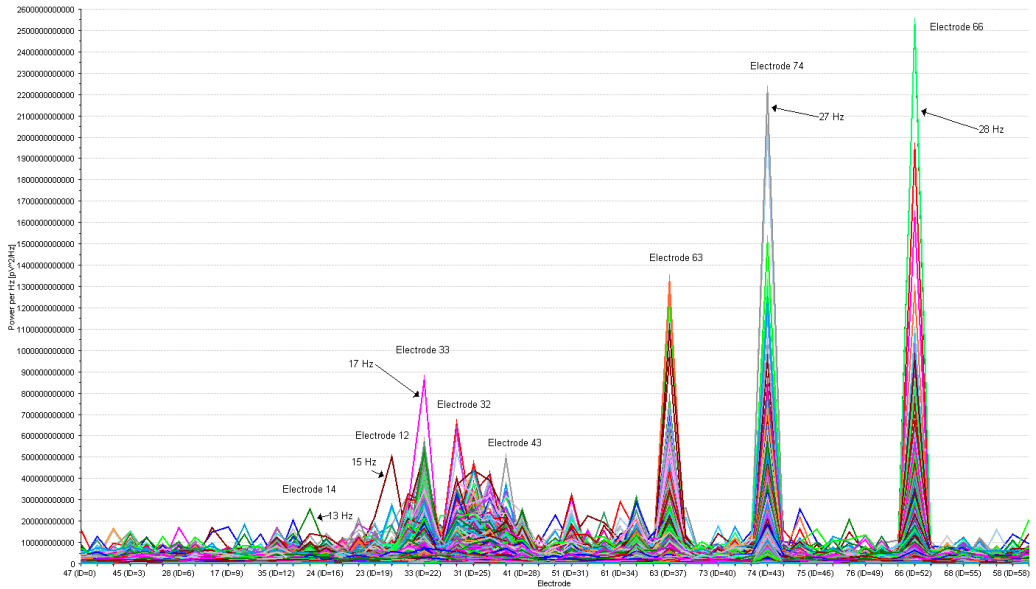
(a) #2. Shows that much of the energy peaks were dominated by low frequencies at a few electrodes
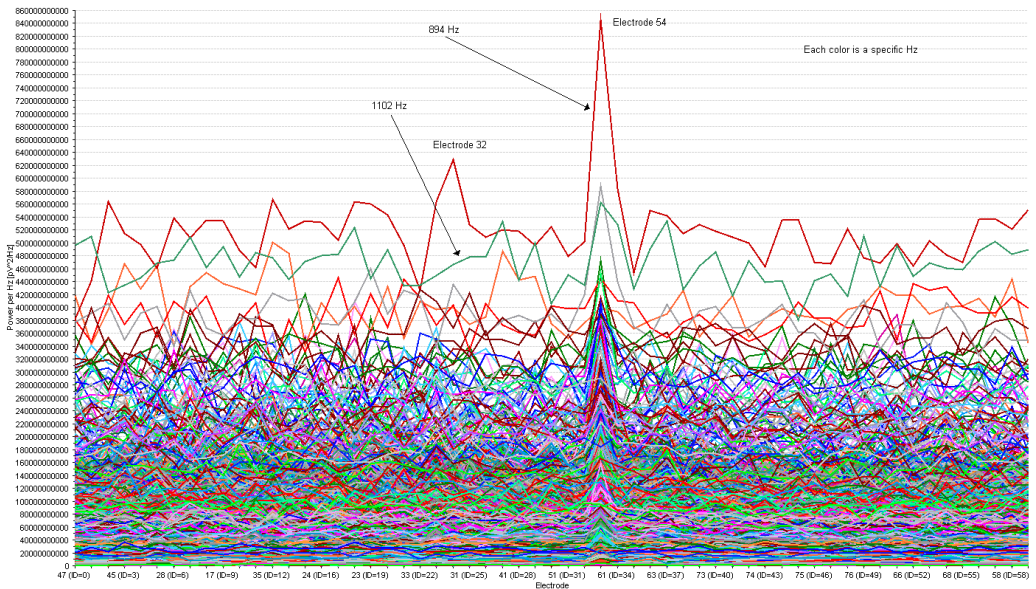
(b) #4. Electrode 54 has in overall higher energy than any other electrode, for a lot of frequencies. 894 Hz was the single frequency with the highest energy, in almost all electrodes. Perhaps 54 simply was a broken electrode?

Figure 3.18: Further investigation on energy variation. Sample plot of PSD data from #2 and #4. Each color is a single frequency. The plots are as if you look into the data from the variable side of the 3D plot in Figure 3.6, but with individual 3D plots for the individual experiments. #4 had clearly more variance in power over the frequencies than in #2. The other major difference is a much larger y scale in (a) because of few electrodes with very high energy at some low frequencies. The fact that #4 had the most variance in power was confirmed in Figure 3.11

Parameter	Setting
Noise reduction [dB]	48
Sensitivity	4
Frequency smoothing	0

Table 3.1: Parameters in the Noise Reduction plugin in Audacity, referring to **this documentation**

The spectrogram used a Hamming window (Harris, 1978), and it was fine-tuned to preview the bursting activity visible on raw signals of 35 in Figure 3.2a. The noise reduction resulted in many clear frequency components occurring during the bursts in #2 and sporadically in #4. A video of #2 of the spectrogram on stereo sound with electrode 35 (left) and 87 (right) can be seen following **this link**. A screenshot of STFT on the selected electrodes of #2 and #4 are in Figure 3.19, and the noise reduced stereo signal from selected electrodes of #2 are shown in Figure 3.20.
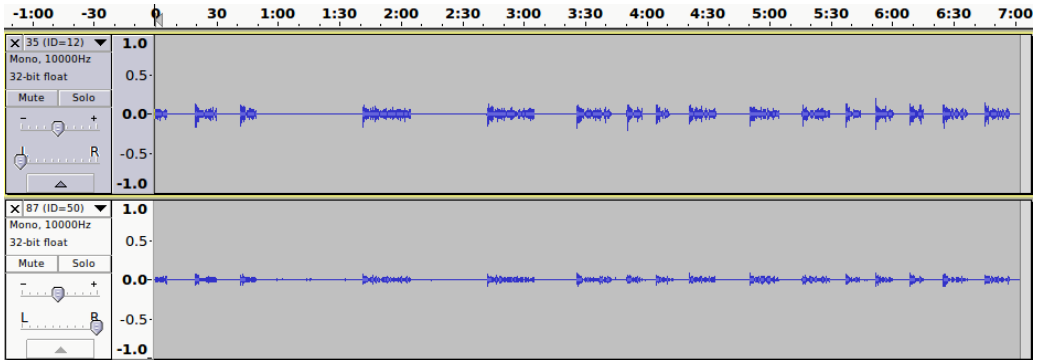
Figure 3.20: The filtered audio signals from #2 in Audacity used as audio input to the STFT in the video described in section 3.9.2. The video had some minor parametric differences in the settings of STFT (other color aperture tresholds) and noise reduction parameters (a little lower noise reduction (40 dB) and sensitivity (3)). Links to the actual sound used in the STFT in Figure 3.19 are available in the subfigure descriptions of that figure
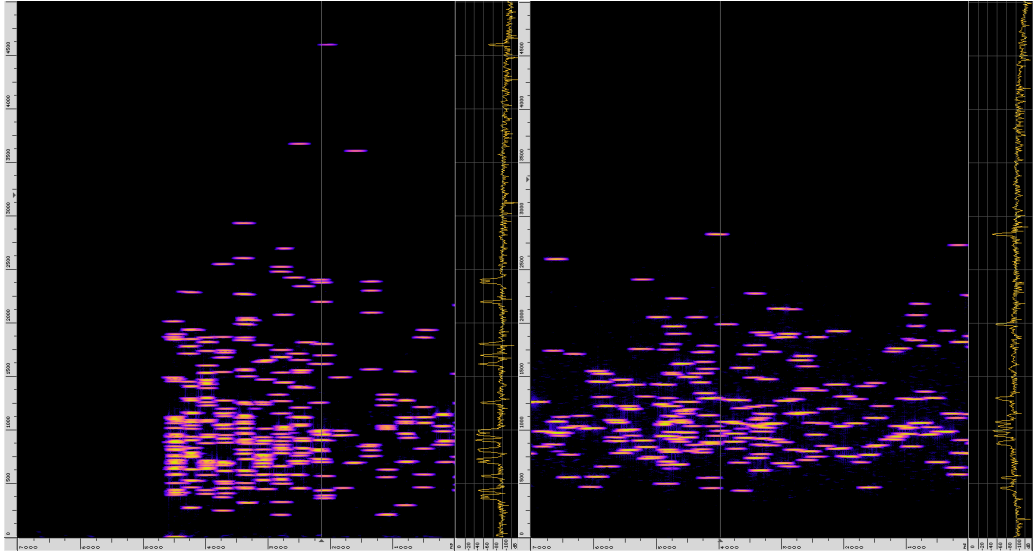
(a) #2 on 35 and 87. In-phase events. The sound can be heard **here**

(b) #4 on 46 and 86. Anti-phase events. The sound **here**

Figure 3.19: Further investigation of spiking activity. Short-Time Fourier Transform (STFT) on a noise reduced selection of the most spiking-active electrodes from #2 and #4. The x axis is a time window of 7 seconds, while y is frequency $f \in \langle 0, 5000]$ Hz. #2 shows the start of a 8 second long burst. In experiment #2, bursts seem to occur with an average length of 10 s, in a frequency 10 s (rough estimates from Figure 3.20). In #4, comparable activity (bursts...) seem to happen almost continuously. In #4, most of the frequency components occur in a wide frequency range around 1000 Hz. Perhaps one can say that a synchonization of the frequency components are what makes the spikes in (a) and seen and detected in Figure 3.2. It is interesting that no continuously changing frequencies occur, and that the length of each individual frequency instance in (a) and (b) seem to be the similar with this noise reduction. Also, most of each individual frequency instance happen at both electrodes. However, the frequency of the instances happening simultaneously can be different. They can be apart from each other in height in the figure. Sometimes, this makes a stereo effect on the sound

Chapter 4

Results and Discussion

4.1 Spatial difference in power and spiking activity

There is a spatial difference in the electrode activity in terms of variance in power (green electrodes and electrode 54 and 64), variance in number of counted spikes (brown and yellow electrodes, electrode 64), as well the number of counted spikes (yellow electrodes, electrode 54). The variance was measured over frequencies (the objects) using PCA. This is further described in Figure 4.1.

Assuming denser brown areas in Figure 4.1b resembles higher density of neuronal cells, it is possible to connect many of the most spiking electrodes (yellow and brown) in Figure 4.1a to the dense areas in this figure.

One could expect that high variance in power means high spiking activity. This was not observed. Two possible reasons are listed below.

- **Stimulation could prevent natural spiking activity**: The neurons get too stressed by the strong periodic electrical voltage to act naturally as they get electrocuted and try to recover from the shock periodically (stimulation period was 30 seconds).

- **Stimulation could increase spike detection treshold**: From PCA, it is evident that stimulation drives the variation in energy. This leads to higher standard deviation, which is used in the formula for spike detection 6.30.

Relating to existing research

Perhaps the spatial difference has something to do with the fact that power and synchronized oscillations and spikes do not necessarily represent the same information if the same power results from de-syncrhonized oscillations (Buzsáki (2012a), gamma-band).

In a study where 58 dissociated cultures of cortical cells, each with different cell densities, were grown and recorded in MEAs, richer cell population bursting patterns (and more detected spikes in total) were observed in denser cultures. The number of spike counts during bursts also increased with ages of cultures in general (Wagenaar, 2006).

(Wagenaar, 2005) did also an attempt on controlling bursts in a similar in-vitro cortical culture with closed-loop control with stimulation, proposing that externally applied electrical stimulation

(a) #1 (day 126), #4 (day 161), #3 (day 204) and #4 (day 245) summary

(b) Day 160. #2 is day 161

(c) Day 46

(d) Day 403
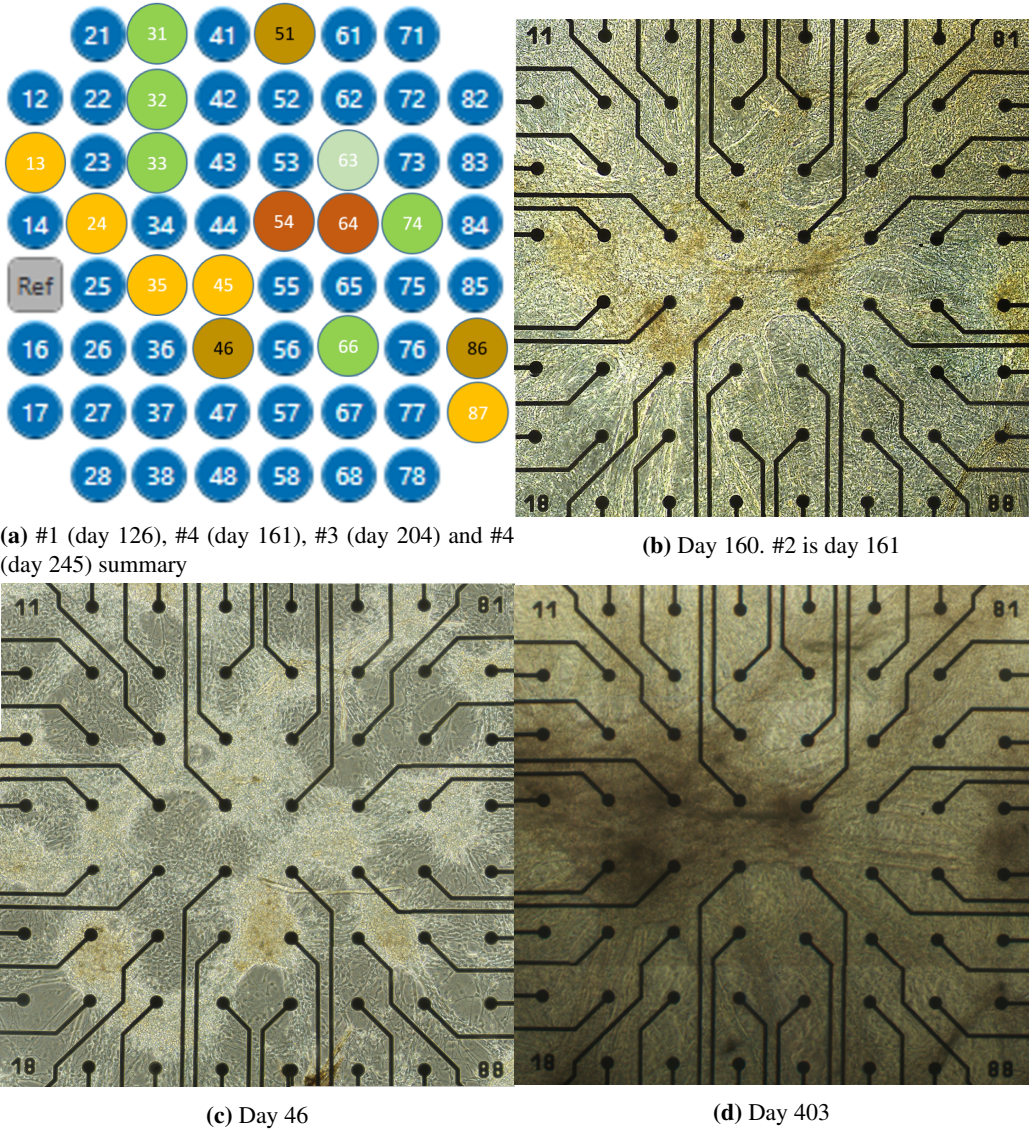
Figure 4.1: Summary of active electrodes in terms of variation in energy and number of detected spikes, as well as MEA pictures before, during and after selected experiments. Green electrodes had a lot of variation in their PDSs, which was expressed by PCA on PSDs. Yellow and brown electrodes had a lot of spike activity, either in terms of detected spikes (yellow, 54) or variation of spike frequency of detected spikes (yellow and brown). 54 had a lot of variation in energy, as well as many detected spikes. 64 had a lot of variation in energy and in detected spikes. (a) and (b) might show that the dense brown areas of neurons in (b) are connected to the spiking activity (yellow, brown and red), and/or that stimulation electrode 74 are connected to the varying energies (green, red)

can substitute for natural inputs to cortical neuronal ensembles in transforming burst-dominated activity to dispersed spiking. Dispersed spiking is more reminiscent of the awake cortex in-vivo. Of 10 MultiChannel Systems MEA cultures containing 60 electrodes, they managed to suppress spontaneous bursting completely in $> 50\%$ of the cultures by randomly selecting 10 electrodes for sequential bi-phasic stimulation with a fixed rate of 10 Hz and with varying voltage amplitude between $100 - 900$ mV. The varying voltage of the stimulation was the feedback loop, and it was determined as $5\times$ spontaneous firing rate (intra-burst firing rate). If the culture in this report were to react similarly as those described, then closed-loop multi-electrode stimulation on varying electrodes and with varying pulse voltage could probably be used to further analyse and control the spatial difference in power and spiking activity.

4.2 De-synchronization on low frequencies (2-4 Hz) could be related to high frequency power increase in $f \in [500, 2000]$ Hz

From Figure 3.16, estimations of the spiking frequencies with the highest powers were 2 and 4 Hz in both #2 and #4. Loading plots from this result were used to select interesting electrodes for de-noising and Short-Term Fourier as seen in figure 3.19. All electrodes were noise reduced with the same parameters and noise profile. It is proposed that activity seen in the STFT is caused by MUAs or mutli-EAPs.

Electrodes with high loading along PC1 were selected. These were not the electrodes with the most spikes detected. Figure 3.16 show that they were in fact among the electrodes with the highest negative correlation to variation in spike frequency. These electrodes seem to be the ones with the most visible bursting activity on raw recordings (at least in #2, which can be verified in Figure 3.2a). The PCA also shows that the selected electrodes are among the ones that negatively correlate the most with the highest power from PSD, and that the estimated spike frequencies associated to the highest powers are 2 and 4 Hz.

The suggested interpretation of the negative correlation is that increasing number of spikes detected correlates with decreasing spike frequency. This makes sense from looking at the original exponential histogram before interpolation in Figure 3.15a. Most of the original spike counts arrive are at very low frequencies. What's more, the selected visible bursting electrodes seem to also have fewer spikes detected in general because of higher spike thresholds. This is further described in Figure 3.2. Perhaps this implies that they are the most interesting electrodes.

By combining the the observation of evident low spiking frequencies in both #2 and #4 with decreasing power from PSD (see Figure 3.13), as well as the frequency behavior from STFT in Figure 3.19, it is observed that the power decrease could be related to de-synchronization of oscillations at these low frequencies. Another observation is that the time duration of a short occurrence of energy at a high frequency (in a wide band around 894 Hz, roughly $f \in [500, 2000]$ Hz), or "stripe", in Figure 3.19a seem to be of similar length in Figure 3.19b. In Figure 3.19a, the frequency if envelopes of most of these "stripes" seem to be around $2 - 4$ Hz, too.

Note that the PSD calculation is not dependent on the spike threshold for noticing the power decrease in the low frequencies (around 2 and 4 Hz). The behavior in Figure 3.19 could also be

related to the fact that the noise profile used for noise reduction on both experiments were taken from experiment #2. The noise profile was the recording of #2 between first and second burst. Hence, #4 was noise reduced as if this recording clip in #2 was noise in #4.

It must also be mentioned that the STFT behavior in #4 happened almost all the time in the recording, while #2 happened in bursts. Such dispersed occurrence of "stripes" could just as well be what causes dispersed spiking observed in #4. Remember, #4 had in total a lot more total spikes detected on all electrodes, than all other experiments (Figure 3.14), and the highest power at 894 Hz (see Figures 3.7, 3.18b). Dispersed spiking as in Figure 3.19b happening almost all the time explains the fact that #4 has the highest power at these very high frequencies (see Figure 3.7). Perhaps is the activity seen in Figure 3.19b de-synchronized oscillations of which would have caused non-stop bursting such as in Figure 3.19a if they were synchronized oscillations.

From Figure 3.7, 3.8, 3.9 and 3.11, it is clear that high frequencies around 894 Hz contributed to most of the variance in the combined PCAs, and most of this energy variance was in #4.

Figure 3.12b shows that PC1 of this PCA expresses the average PSD very well. This is almost the same for #2 in Figure 3.12a and for all experiments in Figure 3.11.

Relating to existing research

There exist research on in-vitro midbrain dopaminergic neuronal cultures proposing that the persistence of global bursting in mature cultures is evidence that such cultures are in a state of arrested development as a result of lack of sensory input. (Wagenaar, 2005) proposes and confirms this view. (Paladini, 2014) proposes that future in-vivo intracellular studies of DA neurons are needed to improve our mechanistic understanding of burst and pause firing of defined DA neurons in awake behaving animals.

(Barter, 2015) hypothesizes from in-vivo recordings on rats that rate neural coding in DA signaling implements gain adjustment for adaptive transition control. They found the firing rate of neurons to be positively correlated to vector components of velocity and acceleration during movements in four directions—up, down, left, right.

Using calcium imaging (Kirwan, 2015) describes similar low frequency oscillatory activity in two long-term externally stimulated in-vitro cultures of hPSC-derived cerebral cortex neurons. Synchronised oscillatory networks develop in a highly stereotyped pattern over several weeks (~5) in culture. An initial phase of increasing frequency of oscillations is followed by a phase of decreasing frequency, before giving rise to non-synchronous, ordered activity patterns. They demonstrated that the formation of hPSC-derived cortical networks mimics in-vivo cortical network development and function as well. Some other works reporting robust long-term electrical activity recorded from hiPSC-derived stem cells are (Odawara, 2016; Lam, 2016).

Using sequential stimulation randomly over all electrodes in a 60 electrode in-vitro MEA containing rat contical neurons, (Hales, 2012) demonstrated the ability to induce pathological high frequency oscillations $(250 - 600\,\mathrm{Hz})$, which they then propose are present in the brains of elipeptic animals and humans. Oscillations were more apparent with higher stimulus voltages.

A journal reviewing principles of cortical rhythms in cognition, presents that rhythms produced by an interneuronal mechanism in cortical circuits are ultrafast $(> 100\,\mathrm{Hz})$ when single neurons fire irregularly in a sparsely synchronized oscillation, but at 40 Hz when single neurons

behave like regular oscillators. Another cortical mechanism described, the excitatory-inhibitory loop mechanism, gives rise to gamma- and beta-band oscillations during highly irregular firing of single neurons (Wang, 2010). It is unclear if any of these cortical mechanisms could exist in the MEA, or if they can be related to midbrain dopaminergic neurons. MEAs that do extra-cellular recordings are anyways "blind" to subthreshold synaptic potentials generated by single cells (Spira, 2013). If the dopaminergic neurons acts anything similar to these cortical neurons, irregular firing of single neurons could be the cause of the lack of synchronization, and hence the increase in high frequency power.

(Stephens, 2012) discusses from in-vitro experiments on adult rat hippocampal neural progenitor cell (NPC)-derived neurons that high-frequency network-wide bursting or "superbursting" is a hallmark of early plasticity that is ultimately refined into mature stable neural network activity. The superbursts are clusters of bursts, and is regarded a transient phase of the culture before it is stabilizing into short high-frequency bursts of activity that persist for the lifetime of the culture. The transient superbusting has also been observed in-vivo.

In an in-vivo study of CA1 pyramidal cells in the hippocampus of rats where sharp wave-associated high-frequency oscillations were measured (200 Hz), (Ylinen, 1995) hypothesize that these fast field oscillations the CA1 region reflect summed Inhibitory postsynaptic potentials (IPSPs) in pyramidal cells as a result of high-frequency barrage of interneurons. They relate these findings to transiently stored memory.

(Wang, 2016) found from in-vitro recordings of human, monkey and mouse neocortex slices that the maximum frequency of APs from cortical fast-spiking neurons were different between species and cortical areas. The maximal mean AP frequency in humans were found to be 338 Hz.

4.3   Decrease of Beta-band power

In Figure 3.13, it is observed that #4 contributes to the least of the variance in the beta-power, following a stable decrease throughout the experiments. Consequently, #4 is the experiment with the least energy in the beta-band. It is assumed that the type of dopaminergic neurons analyzed in the report would have had capabilities of releasing dopamine in an in-vivo scenario (Arenas, 2015)

Relating to Parkinson's disease

As mentioned earlier in section 2.2, Release of dopamine (DA) from midbrain dopaminergic neurons influences voluntary movement, cognition, and motivational state. If the highly likely assumption that the amount of dopaminergic neurons have increased over the experiments is true, a relation to Parkinson's research might have been observed in the MEA culture. The author has had an impression that the data analysed have become "less Parkinson's-like" over the months, when looking at beta-band energy and spiking activity.

In awake rats, dopamine depletion has been found to increase the power and coherence of beta-oscillations (Sharott, 2005). In humans, Parkinson's disease has been found to relate to depletion of dopaminergic neurons such as the type analyzed in this report (Arenas, 2015), leading to motor

symptoms (ex. rigidity, bradykinesia, tremor and postural instability) (Bellucci, 2016; Arenas, 2015).

Previous Time-Frequency analysis conducted on the same experiments by the same author presents similar trends in band power. However, only 2 second segments of electrode 47 from the same four experiments were analysed. Plots of calculated power from band-filtered versions of the last two seconds of electrode 47 of the four experiments are shown in Figure 4.2. Band powers on (a) and (b) were computed from PSDs, while the band powers in (c) and (d) were computed from reconstructed EEG band signals computed from a self-developed adaptive energy based reconstruction algorithm. Bands signals were reconstructed by adding or subtracting Intrinsic Mode Functions (IMFs) the from the original signal, after getting the IMFs from Epirical Mode Decomposition (EMD) of the original signal. A band signals was only reconstructed if EMD managed to decompose the original signal into IMFs where at least some of the IMFs had a certain amount of its power within the band. All figures show similar trends.

Results in Figures 3.13, 3.11 and 3.7 conforms to the results in Figure 4.2 regarding decreasing beta-band activity, which seems to be related to the increasing amount of DAergic neurons.

In the journal paper, multiple references were included reporting on synchronized beta-band oscillations in PD patients, as well as on less beta-band synchronization or de-synchronization after dopamine medications on the patients (Hovden (2017), references [13],[14],[15],[18]. Download link available in Bibliography).

What's more, a substantial increase in power at high frequencies is also observed in Figure 4.2a, which conforms to power results in this report (f. example, see Figure 3.7). Attempts to relate the high frequency power increase this to DAergic neurons, PD or neuronal firing activity in general, was previosly done in the last sections 4.1 and 4.2.

About the peak in gamma power

The apparent peak in gamma-band power in #3 in Figure 4.2 is harder connect to PCA on dopaminergic neurons. On a cellular level, transient cell assemblies may be organized into gamma-wave cycles (Buzsáki, 2012a). Hence, synchronous gamma activity could be related to the making of cells.

(Leondopulos, 2012) reports increased power in theta ($4 - 11$ Hz) and slow gamma ($30 - 55$ Hz) bands in chronic stimulated hippocampal rat neurons.

In a more general perspective, increased gamma activity has been related to sensory responses from in-vivo recordings on mice (Cardin, 2009). Gamma-band activities are well known during sensory-cognitive events in EEG signals of many different species (Bagar-Eroglu, 1996). (Herrmann, 2004) connects gamma activity to memory match and utilization in humans and animals.
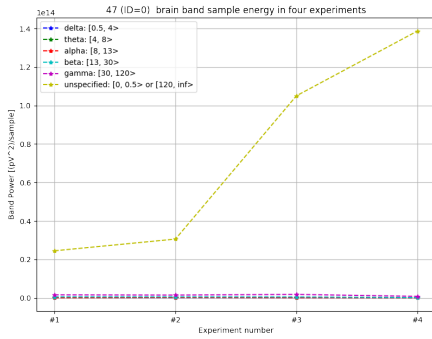
(a) Total band powers from PSD
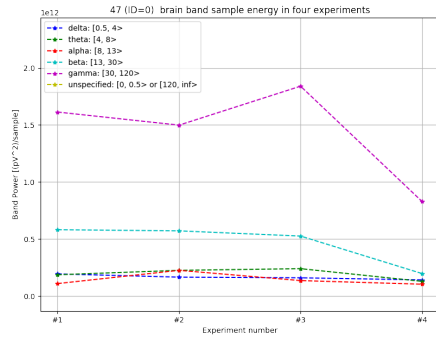
(b) (a) zoomed in on other bands
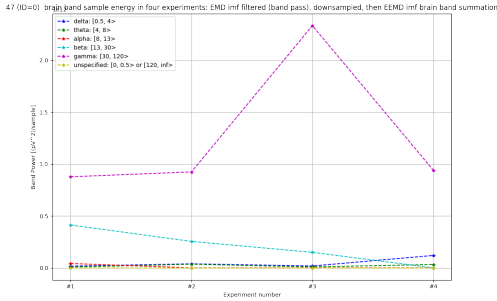
(c) Total band powers from energy based Intrinsic Mode Function (IMF) filtering
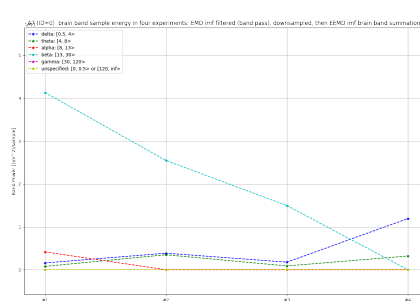
(d) (c) zoomed in on other bands

Figure 4.2: Comparison of power in and outside EEG bands from the last 2 second recordings of electrode 47 in the four experiments. (a) shows a power increase outside EEG bands (unspecified). The energy increase was caused majorly by high frequencies $f \in [120, inf\rangle$. (b) shows energy decrease in beta band (blue) and energy peak in gamma-band in #3. Using an additional method, (c) and (d) shows the same as (b). From (Hovden, 2017)

(a) #1

(b) #2

(c) #3

(d) #4
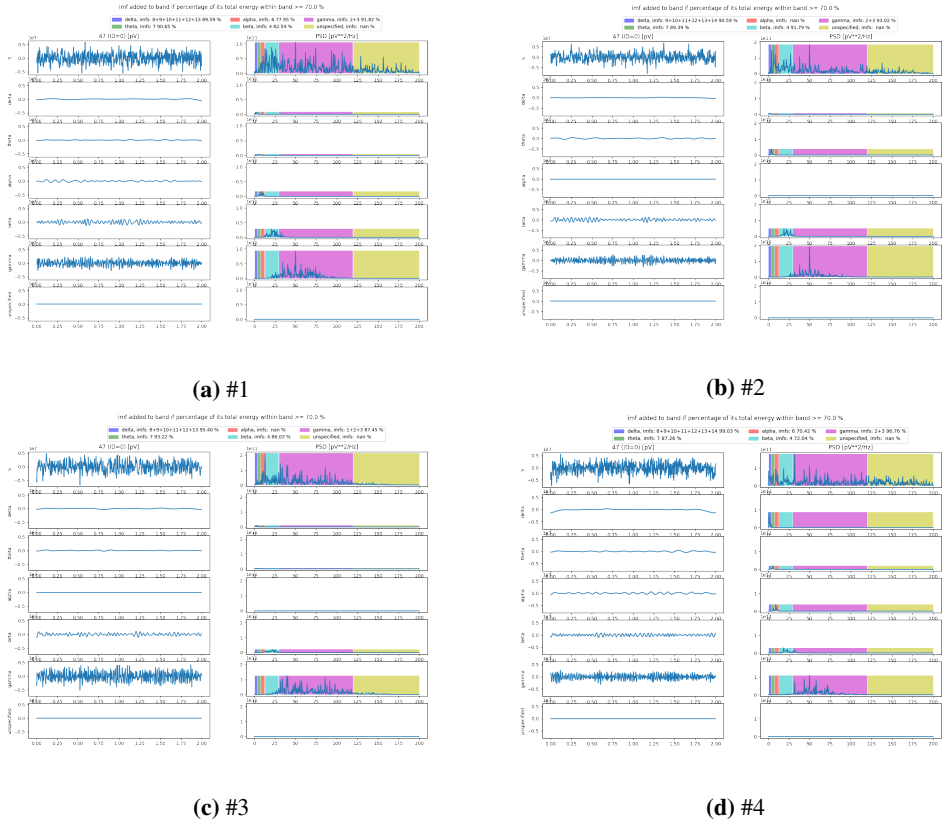
Figure 4.3: Time and frequency domain (PSD) of the reconstructed band signals from last 2 seconds of electrode 47, as well as the original raw signal, from #1, #2, #3 and #4. The height of the marked colors in PSDs are determined from the max peak in the PSD. The gamma signal has the highest oscillations in #3. Decreasing beta oscillations might bee seen. The decreasing power in the beta signal is seen Figure 4.2 (c) and (d). Each band signal was reconstructed by summation of IMFs from Ensemble EMD, based a criteria that minimum 70 % energy criteria were within the band in the PSD. Which IMFs that were made to reconstruct the band signals are shown in the legend. Both time and PSD amplitudes are made comparable within the experiments. All the time signals were used to calculate the powers in Figure 4.2 (c). The reconstructed delta signal should be mostly noise because S was highpass filtered $f > 4$ Hz of same reason for removing $0 - 10$ Hz in PSDs for PCA in this report (assumed low SNR). From (Hovden, 2017)

4.4 Caveats and problems

- **Fault of discarding 0-10 Hz in PSD PCA**: The information in this range is probably not just noise, although empty MEA recordings results in recording much noise in this range (described earlier). Neurons seem to have an insulating effect on the recordings as well.

- **Bad data quality**: The metadata of the recordings available have multiple human errors, such as incorrect stimulation protocol and wrong MEA numbe. It has been reported that configuration files from MCS Experimeter were re-used many times, where the students forgot to configure it correctly to the experiment. This has resulted in recordings with metadata reporting that stimulation occurred, while in reality stimulation did not occur. The MEA electrode layout can also be changed in configuration, hopefully different MEA layouts were not apparent in the files analysed. It is important that human errors do not occur in the systematic work of recording MEAs under various conditions, stimulations and configuration files.

- **Cells' ability to move**: The neuronal cells can move a bit around in the MEA during its development. This is challenging for analysis of development of the culture.

- **Floating offsets**: It might be that the electrode voltage measurements are based on an offset in reference electrode/plate that can happen to change. This is challenging for comparing different experiments. However, variance analysed in this report are not affected by an offset by mean centering or removal of 0 Hz components.

- **Broken or miscalibrated electrodes**: Perhaps is electrode 54 in #4 just a broken electrode? This was the single electrode with the most variance. Perhaps is it just more sensitive than the others from calibration.

- **Impossible to detect neural information propagation**: Helge-André Langåker calculated the minimum information transfer speed in the electrodes to be 11 m/s. This was based on an insight that almost all spike detections from action potentials on MEA 2 occurred at a single and same timestamp for all electrodes that detected the spike. This means that the used sampling rate of 10000 Hz is not fast enough to detect propagation of information in the neural network.

- **Compatibility problems: Use of , in decimals instead of .**: Actually, The Unscrambler X uses comma (,) in the representation of decimals instead of punctuation (.) which is a lot more normal these days. This resulted in manually programming in string manipulation to substitute . for , in the Python script used for PSD computation 6.14 so that The Unscrambler X recognized the decimals as numerical variables.

Chapter 5

Conclusion and further work

In general, two different kinds electrode activity seem to be spatially spread in the MEA. This seems to be caused by the placement of the one-electrode stimulation at electrode 74 as well as the placement of the reference electrode. The first kind of activity was variation in power over frequencies. The second kind was variation in number of detected spikes over spike frequencies, as well as detected spike activity. Activity by means of total detected spikes is regarded similar to activity by means of variation in number of detected spikes over spike frequencies, hence both being referred to as the second kind of activity (however, they are colored differently in Figure 4.1a. Description in figure text). The first kind occurred in electrodes closer to the stimulation electrode opposed to the second kind that occurred closer to the reference point of voltage measurement (with some exceptions). When stimulation was removed, less of the first kind was observed, while some new electrodes recorded more of the second kind. In this experiment (#4), activity seemed to obey less of the spatial difference in activity as was observed in experiments with stimulation (#1-#3). There exist research that has managed to control the occurrence of bursts, hence controlling spatial and rate coding.

A general trend of increasing power at high frequencies ($f \in [500, 2000]$ Hz) as well as decreasing power at low frequencies ($2 - 4$ Hz)) was observed along with the age of the culture, including the experiment that received no external stimulation. The experiment with the most developed properties in this trend was in fact the non-stimulated experiment #4. It is discussed whether this has a connection with the oscillations (in-phase activity visible in #2 at the low frequencies during bursts) de-synchronizing into dispersed anti-phase activity (visible in #4 most of the time) as a result of development in the culture. This behavior can not singularly be caused by external stimulation, but also by development of intrinsic properties in the culture along with age. In other words, the neural coding seem to have gone from rate to temporal (and perhaps still rate) coding. Various kinds of external and internal/intrinsic stimulation has been reported to invoke high-frequency oscillations in neuronal cells in existing research.

Power was also shown to decrease in the beta-band $f \in [13, 30\rangle$ Hz (Figure 3.13), a band in which oscillations have a connection to DA depletion and PD progression from extensive research. A course work on 2 second segments of electrode 47 from the same experiments also verifies decreasing power in beta-band trend, as well as an interesting peak in gamma-band power in #3. (Figures 4.2 and 4.3). Previous research reports that gamma oscillations might be related to transient cell assemblies.

5.0.1 Further work

The ongoing goal is to better understand signaling, dynamics, change and learning in biological neural networks. In this report, analysis of PD-related neuronal data was done to resemble to those aspects as well as to cell replacement therapy for PD.

There is a lot of approaches in further analysis and control

- Sliding window implementation of FFT, CWT, or HHT (Huang, 1998) on noise reduced data streams, for example Short-Time Fourier Transform (STFT) or real-time Continuous wavelet transform (Bömers, 2000), into the existing near real-time reservoir closed-lopp computing Scala framework **SHODAN** (Aaser, 2017a).

- Specialized mother wavelet functions of Continuous wavelet transform (CWT) for using on sliding window. Morlet Wavelet gives good frequency resolution when analyzing gamma activity. Bi-scale wavelet for 1-4 Hz asynchronous BCIs during imaginary movement. Daubechies wavelet for Slow Cortical Potentials (SCPs) $f < 1$ Hz Nicolas-Alonso (2012).

- Analysis of firing rate, phase delay or temporal coding using histograms techniques.

- Analysis of frequency-temporal information of spikes by dividing STFT spectrum into bins and counting spikes separately in each bin. Then creating ISI histograms of each of the bins.

- Evolving PCA (Vitale, 2017; Martens, 2015) on FFT windows (STFT+MVA).

- Multivariate Empirical Mode Decomposition (MEMD) for functional analysis of ex. multivariate beta-oscillations. Empirical Mode decomposition based time frequency analysis of multivariate signals. (Rehman, 2009; Mandic, 2013). (Bueno-López, 2016) (Bueno-López, 2017)

- Effective connectivity analysis such as Dynamic Casual Modeling (DCM, generative model, Friston (2003)), Granger Casuality (GC, Roebroeck (2005)) and Transfer entropy (TE, Vicente (2011); Orlandi (2014); Zuo (2013); Li (2015)).

- Functional (predictive model) vs effective (generative model) connectivity analysis comparison (Friston, 2011).

- Computational models (Bassett, 2017; Kuznetsova, 2010; Cohen, 2016).

- iPSC derived astrocyte recordings as noise profiles (Kuijlaars, 2016)

- Other types of cells: Motor neurons (Shimojo, 2015) (Moser, 2014; Bjerknes, 2014)

- More MVA: Sparse PCA (Zou, 2012). Variable influence on projection (VIP) for orthogonal projections to latent structures (VIP for OPLS) (Galindo-Prieto, 2014). Independent component analysis (ICA). L-PLS as a way to build in more preferences of interpretation of the data. Multilinear PLS (N-PLS).

- Data visualization such as with the Python library **Glumpy**.

- Self-organizing maps and similar.

- Feedback stimulation: Dynamic Clamp DA bursting simulations (Paladini, 2014). Intraburst-based stimulation (Wagenaar, 2005).

- Implementation of existing data intensive processing systems: Apache Spark framework for mapping, example: two-photon imaging brain activity mapping of zebrafish (Freeman, 2014). Spike sorting framework (Rossant, 2016). Examples of specific hardware implementations (Imfeld, 2008; Ballini, 2014).

- Further investigation of dopamine system for Bio Reinforcement learning in the culture (Holroyd, 2002). Support Vector Machines (SVMs), Naive Bayes, Decision trees, Linear Discriminant Analysis (LDA) or various regression methods for feature classification.

- Hidden Markov Models (HMM) or Dynamic Bayesian Networks (DBN) for modeling dynamics in the neural network.

- Particle filter (Sequential Monte Carlo (SMC)) or Kalman filter (linear quadratic estimation (LQE)) as time-domain prediction-update estimators of imagined position of the robot.

The next section tries to describe a possible complete set-up of appliance of combined theory for near real-time bio-robotic control. It is a particle filter based closed loop control of a robot with the MEA in the loop. The section was included as a naive attempt to discuss a possible scenario in the future. Have in mind that there might be misconceptions in the interpretations of the theory.

Closed-loop particle filtered casually modelled MEA robot behavior

First, the 60 electrophysiological MEA signals are noise reduced so that as much as possible of the rate and temporal neural coding are visible as seen in Figure 3.19b.

A sliding window spectrum method (based on FFT, CWT or HHT) is then applied on the electrode signals before or after possible removal of non-important electrodes using MVA approaces such as Variable influence on projection (VIP) for orthogonal projections to latent structures (VIP for OPLS) (which is used either on electrode signals in time, or frequency domain from spectrum of the sliding window) (Galindo-Prieto, 2014, 2015).

The time window is 5 seconds. In any time window, temporal and rate dynamics is recorded. Using STFT or similar, the frequency range in the spectrum is divided into B number of bins. For each of the bins, total number of action potentials in the bin within the time window is recorded. Also for action potentials in each of the bins, ISI histograms are constructed. The result is that for each time window of 5 seconds, maximum $60 \times B$ total spike counts and ISI histograms are recorded.

Then, effective connectivity analysis using transfer entropy (TE) connectivity graphs (Vicente, 2011; Zuo, 2013) as illustrated in Figure 5.1 is applied from discrete spike data of treshold based spike detection on the windowed time signals of maximum 60. A specific discrete type of TE connectivity analysis is used (Pastore, 2017). For each electrode in the connectivity graph, rate and temporal neural coding are maintained trough number of inputs and edge weights, respectively.

TE is an alternative approach to Dynamic Casual Modeling (DCM) (Friston, 2003) for measuring effective connectivity. DCM is essentially based on a Bayesian framework (Friston, 2003) and uses bilinear approximation to describe interactions in nonlinear dynamics, while TE introduces entropy measures from information theory and hence a way to describe nonlinear dynamics without interaction models. TE bases its casual measures on the Markov condition (Vicente, 2011). Another more used casual model is Granger Casuality (GC) and has been used on Functional

magnetic resonance imaging (fMRI) (Roebroeck, 2005). GC, DCE and TE are all casual models and associated with Directed Acyclic Graphs (Janzing, 2013).

(Orlandi, 2014) did a TE study on calcium traces from in-vitro neuronal cultures, while (Li, 2015) did a similar study on calcium traces using fluorescence indicators on mice under anesthetics. Hence, fluorescence signals from individual neurons are evaluated as alternatives to electrodes signals in TE connectivity analysis, if available.

The dynamics of change in the TE connectivity graphs is modelled with a Hidden Markov Model (HMM) or Dynamic Bayesian Network (DBN). Results of functional analysis on each time window are used to update the parameters of the dynamical model. The $60 \times B$ total spike counts and ISI histograms are used for this purpose. MVA methods such as evolving bilinear subspace modeling / evolving PCA (Vitale, 2017) complement in the further interpretation of what changes to update in the model.

Training is attempted on the biological neural network / MEA using closed-loop varying rate and temporal-changing multi-electrode stimulation as in Wagenaar (2005), where different parameters in the stimulation are tested and improved consecutively using a heuristic. The heuristic is to find an optimal set of stimulation parameters in order to find the most discriminating features in the dynamical model when the robot is approaching and leaving a distance to a wall, respectively. (Barter, 2015) reports that distributions of firing rate in DAergic neurons in-vivo in rats were positively correlated to vector components of velocity and acceleration. Such information is what we are looking for in the ISI histograms. The control of the robot is based on a particle filter in top of the discriminating features of the updated dynamical model.

The last sections sum up some additional thoughts about future work.

Stimulation triggered DA release as the real reward feedback?

The presence of electrical stimulation itself is perhaps not working sufficiently enough for any induced learning in the biological neural network, if it is not trying to mimic some neural coding. Perhaps more concrete rewards in the form of DA releases has to be controlled to induce learning. Given that the intrinsic potential of releasing DA in DAergic neuronal cultures exist, optimal learning-inducing stimuli could be suspected to be of the types of neural coding that leads to DA release.

Use of parametric models

For the cyborg project, it would be important to not use too non-parametric models, but instead rely on a fixed complexity on the interpretations of neural coding. This is because the primary task of controlling the robot to not run into walls should not require never-ending training in the biological network. If somehow this is learned in the network, it the network should be able to repeat what it has learned autonomously in a way that does not require much further learning.
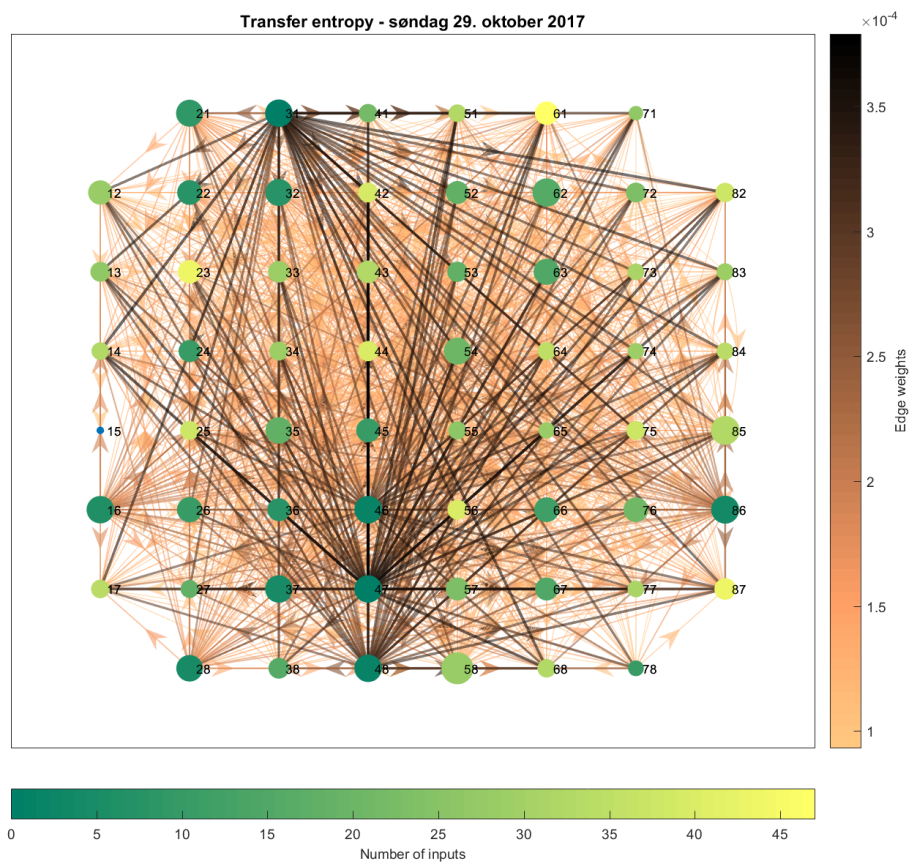
Figure 5.1: TE Connectivity graph from MEA 2 from a recording 29th October 2017. It contains some errors, but it suffices as an illustration. Helge-André Langåker pointed out that it was based on a connectivity matrix that was transposed, and that the arrows point in the wrong direction. Work by Helge based on (Pastore, 2017)

Biology and brain-computer interfaces for taming Artificial General Intelligence

In a learning perspective, making use of biological neural networks in closed-loop circuits (such as brain-computer interfaces (BCIs)) could perhaps be seen as an attempt to counteract the possible threat of Artificial Strong Intelligence (AGI) by making the AI more human-like. AGI is the intelligence of a machine that could successfully perform any intellectual task that a human being can. Many people see AGI as a threat to humanity.

In computer science, research on Artificial Neural Networks (ANNs) have seen significant progress in learning data representations in recent years. ANNs was in fact closely inspired by the biological neural network, leading to simplified models of graphical models of weighted neurons, where the magnitude of a weight represents the ability of information propagation over a synapse. Learning data representations can be effectively accomplished by supervised and unsupervised training in computers using weight updating algorithms such as back-propagation and gradient descent. With the help of the parallel processing abilities of Graphics processing units (GPUs), this algorithm has made ANNs able to learn features in data faster.

The field of ANNs with many hidden layers, called deep learning (DL), has been recently combined with concepts of learning agents using reinforcement learning (RL) techniques, as well as further speeding up the ANN learning processes with more parallelization. Promising learning performances with high-speed learning, resulting in deep reinforcement learning (DRL) techniques such as the system described in (Mnih, 2016) is perhaps the closest research has come in realizing AGI.

Opposed to state-space and action-space modeling in traditional reinforcement learning, deep reinforcement learning discards this modeling in favor of ANNs trained on general rewards. **This video** from (Mnih, 2016) shows the potential of deep reinforcement learning by letting a trained agent optimally drive a race car on a race-track using only raw pixels as input. The reward here was proportional to the agent's velocity along the center of the track at the agent's current position.

The downside of DRL is that the learned data representations are in a black-box in the form of weights and gates, etc. in complex ANNs, hence hiding dynamic relationships of latent variables that would have been described in differential equations in traditional control systems. Such hidden representations is a major reason for skepticism about DL, as it is basically a "blind" machine learning technique. However, it has been very successfull by external metrics, although we don't exactly know how it works.

If such systems get too much impact in important processes such as decision making and control of important infrastructure (ex. self-driving cars, image- or Natural Language Processing (NLP) expert systems used in forensics, judgemental processes, recommender systems on the web for marketing and personalization, or diagnostization in medical sciences), it can have a negative impact on humans if it is wrongly used. Perhaps could future BCIs help humans in controlling and understanding these systems, as well as understanding more about our brain.

Ethics

The neurons analysed in this report were developed from human stem cells. However, rat cells could probably be used instead in the Cyborg project. The reason for using human cells was likely because of the PD research.

Although the general idea of making a cybernetic organism may seem unethical, the benefits from insights and progress along the way will certainly benefit at least the medical sciences.

More controversially, future cyborg-like appliances may be seen as attempts to keep humans in the loop of controlling AGI. Examples can be invasive or non-invasive BCIs for humans that enable high bandwidth communication with computers. Hence it may introduce faster ways of Human-computer interactions (HCIs). Some recent innovations in HCIs today are Virtual reality (VR) headsets and Augmented reality (AR) smartglasses.

HCIs can have the potential to change human behavior and it is important to discuss what kinds of human futuristic behavior we could value. For example, HCIs might challenge the notion of physical presence among other humans in close proximity. What are the most valued types of communications in humans, how can they change and how do we want them to change? From a human perspective, do we want new HCIs what change our ways of communicating?

Chapter 6

Acknowledgements

Bibliography

Aaser, P., December 2017a. Investigating in-vitro neuron cultures as computational reservoir.
   URL https://github.com/PeterAaser/semproj/blob/master/semesterrapport%20peter%20aaser.pdf

Aaser, P., 2017b. Towards making a cyborg: A closed-loop reservoir-neuro system.

Allen, J., 1977. Short term spectral analysis, synthesis, and modification by discrete fourier transform.

Arenas, E., 2015. How to make a midbrain dopaminergic neuron.

Bagar-Eroglu, C., 1996. Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance.

Bai, Q., 2001. Single-unit neural recording with active microelectrode arrays.

Ballini, M., 2014. A 1024-channel cmos microelectrode array with 26,400 electrodes for recording and stimulation of electrogenic cells in vitro.

Barker, R., 2015. Cell-based therapies for parkinson disease —past insights and future potential.

Barter, J., 2015. Beyond reward prediction errors: the role of dopamine in movement kinematics.

Bassett, D., 2017. Small-world brain networks revisited.

Beaulieu, J., 2011. The physiology, signaling, and pharmacology of dopamine receptors.

Bellucci, A., 2016. Review: Parkinson's disease: from synaptic loss to connectome dysfunction.

Beninger, R., 1983. The role of dopamine in locomotor activity and learning.

Bernheimer, H., 1973. Brain dopamine and the syndromes of parkinson and huntington.

Bjerknes, T., 2014. Representation of geometric borders in the developing rat.

Bueno-López, M., 2016. Analysis of neural activity from eeg data based on emd frequency bands, dept. Engineering Cybernetics, NTNU.

Bueno-López, M., 2017. Understanding instantaneous frequency detection: A discussion of hilbert-huang transform versus wavelet transform, dept. Engineering Cybernetics, NTNU.

Buzsáki, G., 2012a. Mechanisms of gamma oscillations.

Buzsáki, G., 2012b. The origin of extracellular fields and currents — eeg, ecog, lfp and spikes.

Bömers, F., 2000. Wavelets in real time digital audio processing: Analysis and sample implementations.

CAMO, December 2017a. The unscrambler x method references.
URL `http://www.camo.com/helpdocs/The_Unscrambler_Method_References.pdf`

CAMO, December 2017b. The unscrambler x product page.
URL `http://www.camo.com/rt/Products/Unscrambler/unscrambler.html`

CAMO, December 2017c. The unscrambler x user manual.
URL `http://www.camo.com/files/TheUnscramblerXv10.3-UserManual.zip`

Cardin, J., 2009. Driving fast-spiking cells induces gamma rhythm and controls sensory responses.

Changkyun, I., 2016. A review of electrodes for the electrical brain signal recording.

Cochran, W., 1967. What is the fast fourier transform?

Cohen, J., 2016. The segregation and integration of distinct brain networks and their relationship to cognition.

Deignan, J., 2012. Sk2 and sk3 expression differentially affect firing frequency and precision in dopamine neurons.

Dellaert, F., 2008. Singular value and eigenvalue decompositions.

Doi, D., 2014. Isolation of human induced pluripotent stem cell-derived dopaminergic progenitors by cell sorting for successful transplantation.

Freeman, J., 2014. Mapping brain activity at scale with cluster computing.

Friston, K., 2003. Dynamic causal modelling.

Friston, K., 2011. Functional and effective connectivity: A review.

Gadagkar, V., 2015. Dopamine neurons encode performance error in singing birds.

Galindo-Prieto, B., 2014. Variable influence on projection (vip) for orthogonal projections to latent structures (opls).

Galindo-Prieto, B., 2015. Variable influence on projection (vip) for opls models and its applicability in multivariate time series analysis.

Gold, C., 2006. On the origin of the extracellular action potential waveform: A modeling study.

Hales, C., 2012. Stimulus-evoked high frequency oscillations are present in neuronal networks on microelectrode arrays.

Harris, F., 1978. On the use of windows for harmonic analysis with the discrete fourier transform.

Herrmann, C., 2004. Cognitive functions of gamma-band activity: memory match and utilization.

Holroyd, C., 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity.

Hovden, I., 2017. Long-term frequency band characteristics of a stimulated stem cell-derived dopaminergicneuronal culture.
URL `http://folk.ntnu.no/ivarthov/Long-term%20frequency%20band%20characteristics%20of%20a%20stimulated%20stem%20cell-derived%20dopaminergic%20neuronal%20culture_Hovden_TTK7_IEEEtran_journal_paper_corrected_20171211.pdf`

Howe, M., 2016. Rapid signalling in distinct dopaminergic axons during locomotion and reward.

Huang, N., 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis.

Ikemoto, S., 2015. Basal ganglia circuit loops, dopamine and motivation: A review and enquiry.

Imfeld, K., 2008. Large-scale, high-resolution data acquisition system for extracellular recording of electrophysiological activity.

Janzing, J., 2013. Quantifying causal influences.

Jay, M., 2015. Firing dynamics and modulatory actions of supraspinal dopaminergic neurons during zebrafish locomotor behavior.

Jo, J., 2016. Midbrain-like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons.

Jolliffe, I., 1986. Principal Component Analysis. Springer Science+Business Media, LLC.

Kaelbling, L., 1996. Firing frequency maxima of fast-spiking neurons in human, monkey, and mouse neocortex.

Kerkoerle, T., 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex.

Kirkeby, A., 2013. Generating regionalized neuronal cells from pluripotency, a step-by-step protocol.

Kirkeby, A., 2017. Predictive markers guide differentiation to improve graft outcome in clinical translation of hesc-based therapy for parkinson's disease.

Kirwan, P., 2015. Development and function of human cerebral cortex neural networks from pluripotent stem cells in vitro.

Kuijlaars, J., 2016. Sustained synchronized neuronal network activity in a human astrocyte co-culture system.

Kuznetsova, A., 2010. Regulation of firing frequency in a computational model of a midbrain dopaminergic neuron.

La Manno, G., 2016. Molecular diversity of midbrain development in mouse, human, and stem cells.

Lam, R., 2016. Functional maturation of human stem cell- derived neurons in long-term cultures.

Leondopulos, S., 2012. Chronic stimulation of cultured neuronal networks boosts low-frequency oscillatory activity at theta and gamma with spikes phase-locked to gamma frequencies.

Li, J., 2015. Reconstructing neuronal connectivity from calcium imaging data using generalized transfer entropy.

Mandic, D., 2013. Empirical mode decomposition-based time-frequency analysis of multivariate signals.

Martens, H., 2015. Quantitative big data: where chemometrics can contribute.

Martens, H., 2016. Three algorithms for principal component analysis, dept. Engineering Cybernetics, NTNU.

Maynard, A., 1996. The utah intracortical electrode array: a recording structure for potential brain-computer interfaces.

Mnih, V., 2016. Asynchronous methods for deep reinforcement learning.

Moser, E., 2014. Grid cells and cortical representation.

MultiChannelSystems, December 2017a. Mea2100-system manual.
URL `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/MEA2100-System_Manual.pdf`

MultiChannelSystems, December 2017b. Multichannel systems analyzer manual.
URL `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/Multi%20Channel%20Analyzer_Manual.pdf`

MultiChannelSystems, December 2017c. Multichannel systems datamanager manual.
URL `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/Multi%20Channel%20DataManager_Manual.pdf`

MultiChannelSystems, December 2017d. Multichannel systems experimeter manual.
URL `http://www.multichannelsystems.com/sites/multichannelsystems.com/files/documents/manuals/Multi%20Channel%20Experimenter_Manual.pdf`

NationalInstruments, December 2017. The fundamentals of fft-based signal analysis and measurement in labview and labwindows/cvi. From National Instruments.
URL `http://www.ni.com/white-paper/4278/en/`

Nicolas-Alonso, L., 2012. Brain computer interfaces, a review.

Nácher, V., 2013. Coherent delta-band oscillations between cortical areas correlate with decision making.

Obien, M., 2015. Revealing neuronal function through microelectrode array recordings.

Odawara, A., 2016. Physiological maturation and drug responses of human induced pluripotent stem cell-derived cortical neuronal networks in long- term culture.

Orlandi, J., 2014. Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging.

Paladini, C., 2014. Generating bursts (and pauses) in the dopamine midbrain neurons.

Pastore, V., 2017. Spicodyn: A toolbox for the analysis of neuronal network dynamics and connectivity from multi-site spike signal recordings.

Pine, J., 1980. Recording action potentials from cultured neurons with extracellular microcircuit electrodes.

Potter, S., 2006. Closing the loop: stimulation feedback systems for embodied mea cultures.

Proakis, J., 2014. Digital Signal Processing. Pearson Education Limited, Ch. 4, 7, pp. 252, 474.

Rehman, N., 2009. Multivariate empirical mode decomposition.

Richardson, M., 2009. Principal component analysis.

Risvik, H., 2007. Principal component analysis (pca) nipals algorithm.

Roebroeck, A., 2005. Mapping directed influence over the brain using granger causality and fmri.

Rossant, C., 2016. Spike sorting for large, dense electrode arrays.

Scipy, December 2017. scipy.signal.periodogram. From the Scipy Community.
URL https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.periodogram.html

Seeman, P., 1987. Human brain d and d dopamine receptors in schizophrenia, alzheimer's, parkinson's, and huntington's diseases.

Sharott, A., 2005. Dopamine depletion increases the power and coherence of b-oscillations in the cerebral cortex and subthalamic nucleus of the awake rat.

Shimojo, D., 2015. Rapid, efficient, and simple motor neuron differentiation from human pluripotent stem cells.

Shlens, J., 2014. A tutorial on principal components analysis.

Siemens, December 2017a. The autopower function... demystified! From Siemens Product Lifecycle Management Community.
URL https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/The-Autopower-Function-Demystified/ta-p/362368

Siemens, December 2017b. Digital signal processing: Sampling rates, bandwidth, spectral lines, and more.... From Siemens Product Lifecycle Management Community.
URL https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Digital-Signal-Processing-Sampling-Rates-Bandwidt ta-p/402991

Siemens, December 2017c. What is a power spectral density (psd)? From Siemens Product Lifecycle Management Community.
URL https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/What-is-a-Power-Spectral-Density-PSD/ta-p/360969

Smith, L., 2002. A tutorial on principal components analysis.

Spira, M., 2013. Multi-electrode array technologies for neuroscience and cardiology.

Stephens, C., 2012. Adult neural progenitor cells reactivate superbursting in mature neural networks.

Vicente, R., 2011. Transfer entropy—a model-free measure of effective connectivity for the neurosciences.

Vitale, R., 2017. On-the-fly processing of continuous high-dimensional data streams.

Viventi, J., 2012. Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo.

Wagenaar, D., 2005. Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation.

Wagenaar, D., 2006. An extremely rich repertoire of bursting patterns during the development of cortical cultures.

Waldert, S., 2009. A review on directional information in neural signals for brain-machine interfaces.

Wang, B., 2016. Firing frequency maxima of fast-spiking neurons in human, monkey, and mouse neocortex.

Wang, X., 2010. Neurophysiological and computational principles of cortical rhythms in cognition.

Weisstein, E., December 2017a. Eigen decomposition theorem. From MathWorld–A Wolfram Web Resource.
URL `http://mathworld.wolfram.com/EigenDecompositionTheorem.html`

Weisstein, E., December 2017b. Singular value decomposition. From MathWorld–A Wolfram Web Resource.
URL `http://mathworld.wolfram.com/SingularValueDecomposition.html`

Wold, S., 1987. Principal components analysis.

Ylinen, A., 1995. Sharp wave-associated high-frequency oscillation (200 hz) in the intact hippocampus: Network and intracellular mechanisms.

Zou, H., 2012. Sparse principal component analysis.

Zuo, K., 2013. Adaptive kernels and transfer entropy for neural connectivity analysis in eeg signals.

Appendix

A top-down collection of background information.

6.1 Principal Component Analysis (PCA)

In this section, the mathematical foundation of PCA is explained using eigen decomposition. The relation of eigen decomposition and Singular Value Decomposition (SVD) for PCA is then explained. The concept of scores and loadings comes from PCA using SVD, which is later explained.

$\mathbf{W}$ from 6.11 is the input matrix used in The Unscrambler X software, which is selected to use SVD for the PCA (CAMO, 2017a). It is the transpose of the following $\mathbf{X}$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1N} \\ x_{21} & x_{22} & \ldots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \ldots & x_{MN} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} \in \mathbb{R}^{M \times N} \tag{6.1}$$

with M variables and N objects.

Each variable sequence are first mean centered as in 6.23. From linear algebra, we know that any object in $\mathbf{X}$ can be expressed by a linear combination of $Rank(\mathbf{X})$ orthonormal (unit length, linearly independent) basis vectors.

PCA on $\mathbf{X}$ finds new variables that form a new basis for the objects, on the criteria that they in decreasing order express the variance in the objects. The new latent variables no longer hide information about covariances between objects, as the original variables did, because they are orthogonal.

6.1.1 Derivation using Eigen Decomposition

The expression of the data in orthonormal basis vectors follows from the linear transformation (Shlens, 2014; Richardson, 2009)

$$\mathbf{Y} = \mathbf{PX} \tag{6.2}$$

where the rows of $\mathbf{P}$ are the latent variables, or the Principal Components (PCs). Hence, the new variables are linear combinations of the original variables.

The most straight-forward method for finding $\mathbf{P}$ is based on eigen decomposition (6.24) of the covariance matrix of $\mathbf{X}$, $\mathbf{C_X}$, with an additional requirement that the eigenvectors of $\mathbf{C_X}$ are principal components and row vectors of $\mathbf{P}$. This results in a diagonal covariance matrix of $\mathbf{Y}$.

The covariance matrix of $\mathbf{Y}$ is

$$\mathbf{C_Y} = \frac{1}{n-1}\mathbf{YY}^T \tag{6.3}$$

Inserting 6.2 into 6.3 and recognizing $\mathbf{C_X}$ results in

$$\mathbf{C_Y} = \mathbf{PC_X P}^T \tag{6.4}$$

Eigen decomposition 6.24 is now done on $\mathbf{C_X}$ and substituted into 6.4.

$$\mathbf{C_Y} = \mathbf{P(EDE^{-1})P}^T \tag{6.5}$$

Finally, since it is required that the principal component row vectors of $\mathbf{P}$ are the eigenvector colums of $\mathbf{E}$, and since the inverse of an orthogonal matrix is its transpose

$$\mathbf{E} \equiv \mathbf{P}^T \tag{6.6}$$

6.6 is then inserted into 6.5 resulting in

$$\mathbf{C_Y} = \mathbf{D} \tag{6.7}$$

Consequently, the covariance matrix of the linear orthonormal transformation $\mathbf{Y}$, $\mathbf{C_Y}$ is a diagonal square matrix with its variance expressed along the diagonal equal to $\mathbf{D}$. The diagonal of $\mathbf{D}$ are the eigenvalues from the eigen decomposition of the covariance matrix of $\mathbf{C_X}$.

6.1.2 Relation of Eigen Decomposition to SVD for PCA

The eigendecomposition of $\mathbf{C_X}$ is related to Singular Value Decomposition (SVD) by defining (Shlens, 2014; Richardson, 2009)

$$\mathbf{Z} = \frac{1}{\sqrt{N-1}}\mathbf{X^T} \in \mathbb{R}^{N \times M} \tag{6.8}$$

$$\mathbf{Z^T Z} = \frac{1}{N-1}\mathbf{XX^T} = \mathbf{C_X} \tag{6.9}$$

and performing SVD (6.25) on $\mathbf{Z}$. This results in non-zero singular values of $\mathbf{Z}$ equal to the square roots of the non-zero eigenvalues of $\mathbf{Z^T Z}$ (Richardson, 2009; Wold, 1987)

$$\sqrt{\lambda_i} = s_i \quad , i \in Rank(\mathbf{Z}) \tag{6.10}$$

However, when solving PCA with SVD, the input matrix is more typically just the transpose of $\mathbf{X}$

$$\mathbf{W} = \mathbf{X^T} \in \mathbb{R}^{N \times M} \tag{6.11}$$

with N objects and M variables (Jolliffe, 1986; Wold, 1987).

The lack of $\frac{1}{\sqrt{N-1}}$ in 6.11 when performing SVD on $\mathbf{W}$ results in scaled principal components (Dellaert, 2008), or a scaled version of non-zero singular values of $\mathbf{W}$ equal to the square roots of the non-zero eigenvalues of $\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$.

$$\sqrt{\lambda_i} = \frac{s_i}{\sqrt{N-1}} \quad , i \in Rank(\mathbf{W}) \tag{6.12}$$

λ_i in 6.12 are the eigenvalues following from PCA using SVD (Jolliffe (1986), p. 38). Naturally, $Rank(\mathbf{Z}) = Rank(\mathbf{W})$, since $\mathbf{Z}$ and $\mathbf{W}$ only differ by a scalar value.

Since in a practical scenario only the first k principal components are selected to display the variance in $\mathbf{X}$, the reconstruction of $\mathbf{W}$ using SVD introduces an error term, or residual $\mathbf{E}_k$

$$\mathbf{W}_k = \mathbf{S}_k\mathbf{L}_k + \mathbf{E}_k = Structure + Noise \tag{6.13}$$

where

$$\mathbf{S}_k = \mathbf{U}_k\mathbf{D}_k \quad , \quad \mathbf{L}_k = \mathbf{V^T}_k \quad and \quad k <= Rank(\mathbf{W}) \tag{6.14}$$

denoted as scores $\mathbf{S}$, loadings $\mathbf{L}$ and number of selected principal components k (Wold, 1987; Martens, 2016; CAMO, 2017a). $\mathbf{S}_k\mathbf{L}_k$ is the $Structure$ part, while $\mathbf{E}_k$ is the $Noise$ part. Naturally, the $Noise$ part includes less of the variance in X when k is increased.

Typically, only the first k principal components of the projection account for most of the variance in the objects. This is because eigenvalues and convectors of the SVD are Rank-ordered, meaning that less variance in $\mathbf{X}$, or less magnitude of eigenvalues, is explained when adding a principal component to the model. For this reason, the other "less important" principal components are discarded in further analysis. According to this procedure, the reconstruction of the original data matrix would not be lossless, but the significant activity in the data regarding variance is hopefully included along the existing principal components.

6.1.3   Scores and map of objects

A map of objects that describes how they relate to each other. The first column of $\mathbf{S}$ is the score of the first PC for all objects. The term "Scores" are the elements in the $\mathbf{S}$ matrix (Risvik, 2007). The scores can be plotted against each other in a scores plot, which is scores plotted against each other.

6.1.4   Loadings and map of variables

A map of variables. $\mathbf{L}$ contains the weights (or influence) of the variables in $\mathbf{X}$ on the scores $\mathbf{S}$. Displays which variables that are responsible for patterns found in scores when using a loadings plot. The loadings plot is the loadings of a PC plotted against the loadings of another PC.

6.1.5 Residuals and explained variance plot

The part of $\mathbf{X}$ that is not explained by the model. The model is the $Structure$ part of $\mathbf{W}_k$ when a k number of PCs are used. If the model expresses 100% of the variance in $\mathbf{X}$, then $\mathbf{E}_k = 0$ and $k = Rank(\mathbf{X})$. This means 100% explained variance and 0% residual variance from the relation

$$ExplainedVariance + ResidualVariance = 100\% \tag{6.15}$$

The residual variance is calculated from residues in $\mathbf{E}_k$. A residue for an object e_i is a distance between the object in variable space and the representation of the object in PC space. It is important that variable segments in $\mathbf{X}$ are mean-centered. Then the residual variance is relative to the average object of $\mathbf{X}$, and is calculated using the formula (Risvik, 2007)

$$ResidualVariance = e_{tot}^2 = \sum_{n=0}^{N-1} e_i^2 \tag{6.16}$$

6.15 and 6.16 is used to make a plot of explained variance, which can be used to select an appropriate number of PCs for the model.

6.2 The covariance matrix $\mathbf{C_X}$

Given the $M \times N$ data matrix with M variables and N objects

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1N} \\ x_{21} & x_{22} & \ldots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \ldots & x_{MN} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} \in \mathbb{R}^{M \times N} \tag{6.17}$$

the covariance matrix is defined to be (Richardson, 2009):

$$\mathbf{C_X} = \frac{1}{N-1}\mathbf{XX^T} = \frac{1}{N-1} \begin{bmatrix} \mathbf{x}_1\mathbf{x}_1^T & \mathbf{x}_1\mathbf{x}_2^T & \ldots & \mathbf{x}_1\mathbf{x}_M^T \\ \mathbf{x}_2\mathbf{x}_1^T & \mathbf{x}_2\mathbf{x}_2^T & \ldots & \mathbf{x}_1\mathbf{x}_M^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_M\mathbf{x}_1^T & \mathbf{x}_M\mathbf{x}_2^T & \ldots & \mathbf{x}_M\mathbf{x}_M^T \end{bmatrix} \in \mathbb{R}^{M \times M} \tag{6.18}$$

which expresses covariances between all the different variables. Hence,

$$\mathbf{C_X} = \frac{1}{N-1} \begin{bmatrix} Cov(\mathbf{x}_1, \mathbf{x}_1) & Cov(\mathbf{x}_1, \mathbf{x}_2) & \ldots & Cov(\mathbf{x}_1, \mathbf{x}_M) \\ Cov(\mathbf{x}_2, \mathbf{x}_1) & Cov(\mathbf{x}_2, \mathbf{x}_2) & \ldots & Cov(\mathbf{x}_1, \mathbf{x}_M) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{x}_M, \mathbf{x}_1) & Cov(\mathbf{x}_M, \mathbf{x}_2) & \ldots & Cov(\mathbf{x}_M, \mathbf{x}_M) \end{bmatrix} \in \mathbb{R}^{m \times m} \tag{6.19}$$

6.3   Covariance

The covariance between the two variable sequences of length N, $\mathbf{x}_1$ and $\mathbf{x}_2$, is defined as

$$Cov(\mathbf{x}_1, \mathbf{x}_2) = E[(\mathbf{x}_1 - E[\mathbf{x}_1])(\mathbf{x}_2 - E[\mathbf{x}_2])] \tag{6.20}$$

Assuming equiprobable outcomes, the covariance formula becomes (Smith, 2002)

$$Cov(\mathbf{x}_1, \mathbf{x}_2) = E[(\mathbf{x}_1 - \bar{\mathbf{x}}_1)(\mathbf{x}_2 - \bar{\mathbf{x}}_2)] = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2) \tag{6.21}$$

6.4   Variance

The variance is a measurement of variation on a single variable sequence $\mathbf{x}$, or the covariance between $\mathbf{x}$ and $\mathbf{x}$

$$Var(\mathbf{x}) = Cov(\mathbf{x}, \mathbf{x}) = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})^2 \tag{6.22}$$

6.5   Mean

The mean $\bar{\mathbf{x}}$ of $\mathbf{x}$ is defined as (Smith, 2002)

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \tag{6.23}$$

6.6   Eigen Decomposition Theorem

Let $\mathbf{E}$ be a matrix of eigenvectors of a given square matrix $\mathbf{A}$ and $\mathbf{D}$ be a diagonal matrix with the corresponding eigenvalues on the diagonal. Then, as long as $\mathbf{E}$ is a square matrix, $\mathbf{A}$ can be written as an eigen decomposition

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1} \tag{6.24}$$

where $\mathbf{D}$ is a diagonal matrix. Furthermore, if $\mathbf{A}$ is symmetric, then the columns of $\mathbf{E}$ are orthogonal vectors.

If $\mathbf{E}$ is not a square matrix, then it cannot have an inverse and eigen decomposition is not possible. However, if $\mathbf{E} \in \mathbb{R}^{M \times N}$ where $\mathbf{M} > \mathbf{N}$, Singular Value Decomposition can be used instead to decompose $\mathbf{A}$.

Source: (Weisstein, 2017a)

6.7 Singular Value Decomposition (SVD)

Let $\mathbf{A} \in \mathbb{R}^{M \times N}$ be a real matrix with $\mathbf{M} > \mathbf{N}$. Then $\mathbf{A}$ can be written on the form

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathbf{T}} \tag{6.25}$$

where $\mathbf{U} \in \mathbb{R}^{M \times M} \quad \mathbf{D} \in \mathbb{R}^{M \times N} \quad \mathbf{V} \in \mathbb{R}^{N \times N}$

and $\mathbf{U}$ and $\mathbf{V}$ have orthogonal columns so that

$$\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} \tag{6.26}$$

and

$$\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I} \tag{6.27}$$

where $\mathbf{I}$ is the identity matrix. The two identity matrices can have different dimensions.

Source: (Weisstein, 2017b)

6.8 Parseval's relation for discrete periodic signals

Parseval's relation is an application of the Plancherel theorem on the Fourier transform. The discrete version yells for the Discrete Fourier Transform (DFT), used in the project, and tells the relation of total energy in a discrete signal in time and frequency domain, where the frequency domain is the DFT of the time domain signal. (Proakis, 2014)

$$E_N = \sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2 \tag{6.28}$$

where $X(k) = DFT(x(n))$.

6.9 Power Spectral Density (PSD)

PSD is a scaled version of the Fast Fourier Transform (FFT), which represents power in the discrete signal independently of the sampling rate. Said differently, PSD is the auto power spectrum of $x(n)$ scaled by the frequency resolution, $\frac{1}{\Delta f} = N \times \Delta t = \frac{N}{F_s}$ (this equation is called the golden equation of digital signal processing).

$$PSD = \frac{N}{Fs}S_{xx} = \frac{1}{Fs \times N}|X(k)|^2 \quad and \quad S_{xx} = \frac{1}{N^2}|X(k)|^2 \tag{6.29}$$

where Fs is the sampling rate, $X(k)$ is the FFT of $x(n)$, S_{xx} is the auto power spectrum and N is the length of the discrete signal, following from 6.28 (Siemens, 2017b,a,c; NationalInstruments, 2017; Scipy, 2017).

6.10 Detecting spikes from raw signals

A spike on sample n on the time domain signal $x(n)$ is detected if

$$x(n) > 5\sqrt{Var(x(n))} \quad or \quad x(n) < -5\sqrt{Var(x(n))} \tag{6.30}$$

where the variance was calculated in MultiChannel Systems Analyzer (MultiChannelSystems, 2017b) using a selected sample segment N. In the PCA spike recordings, $N = TimeDuration[s] \times SampleRate[\frac{1}{s}] = 10[s] \times 10000[\frac{1}{s}] = 100000$.

6.11 DFT, FFT, STFT and window functions

The Discrete Fourier Transform (DFT) is the frequency-amplitude spectrum of discrete periodic time-domain signals (Proakis, 2014).

The Fast Fourier Transform is an efficient software implementation of DFT (Cochran, 1967).

The Short-Time Fourier Transform (STFT) is a time-frequency analysis of time-domain signals. It computes multiple DFTs from a small sliding window, resulting in a frequency-amplitude-time representation (Allen, 1977).

There exist many types of windows that can be used in STFT (Harris, 1978). A typical window is the "Hamming" window. This window is used in the electrode sound STFT analysis.

6.12 Cross-validation

A common way to evaluate the fitness of the model during training (or statistical modeling) by dividing the training data to N segments, then training the model N times on $N - 1$ segments, each time validating the model with the single segment that was not used for training at that time.

6.13 Software used

6.13.1 MultiChannel Systems Lab Software

- MultiChannel Experimeter, on-line MEA recorder and analyzer. Used to record the data at Sandvigs' lab at St. Olav University Hospital.

- MultiChannel Analyzer, off-line analyzer. Used to re-run the recorded data to do spike detection with new standard deviation estimation for spike detection 6.30.

- MultiChannel DataManager, Used to convert the proprietary raw format to comma-separated (csv) format.

6.13.2 The Unscrambler X

Licensed access to a multivariate analysis software called "The Unscrambler X" (CAMO, 2017b) was given for use in a course in the master's studies (TTK19 Big Data Cybernetics) as well as in project works for related students.

It is typically used in the Chemometrics industries, and provides with a fast way to analyse not so heavy but pre-processed multivariate data.

Although neuronal data is a little heavy in this program, it presents all the PCA plots and model calibration and validation metrics in an organized way. Most of the plots in the report were made with this software.

6.13.3 Microsoft Excel

Used for simple pre-processing of spike data from from MultiChannel Analyzer and Experimeter.

6.13.4 Open-source scientific computing and visualization software

- Python 2 and 3 with the following python libraries used to make script to convert raw csv time domain recordings to frequency down-scaled Power Spectral Densities for each electrode.

  - os
  - glob
  - numpy
  - pandas
  - scipy
  - matplotlib
  - pytftb
  - pyhht
  - PyEMD
  - datetime

- Jupyter Notebook, a programming web server. Used during programming to plot data.

- Audacity, an audio editor. Used for noise reduction for converted audio signals from electrodes.

- Baudline, a scientific spectrogram analyzer. Used to analyze the noise reduced audio signals.

- Ubuntu Linux, used with all mentioned software except The Unscrambler X, MultiChannel Software and Microsoft Excel.

6.14 Python script

When placed in the same folder as the existing converted MEA 2 raw electrode csv files (files such as in Figure 3.3), it computes the PSDs of the electrode raw data in the csv files and combines them into a combined experiment format described in section 3.6.1 that The Unscrambler X understands. It also adds some additional metadata colums for each experiments. The header in the unscrambler-friendly PSD data is described in Table 6.1.

Header column	Description
Hz	$1 - 5000$
PSD for Electrode 1-60 $[pV^2/Hz]$	$1 - 60$
Average Power Density per Hz [np.mean(P[Hz])]	decimal
Duration [s]	decimal
Total Energy [sum(Average Power Density per Hz[Hz]*(Fs/df))]	decimal
Experiment average signal mean [pV]	decimal
Experiment average signal min [pV]	decimal
Experiment average signal max [pV]	decimal
Day nr. from seeding date 2016-10-11	integer
Experiment date [yyyy-mm-dd]	string
Stimulation info	string
Frequency band	EEG bands or unspecified

Table 6.1: The header of the combined PSDs for PCA analysis in The Unscrambler X. The Average Power Density per Hz column was used to group objects into sections in Figures 3.11, 3.12, 3.13b and 3.16

The script works on Ubuntu Linux with the imported dependencies installed.

```
1   import os
2   import glob
3   import pandas
4   import numpy as np
5   from scipy import signal
6   from datetime import datetime
7
8   # raw csv or modified (difference in line number start of file, header_row)
9   raw_csv = True
10
11  # the date that MEA2 was seeded
12  date_seeded = "2016-10-11"
13
14  # simple low pass filter f <= c
15  frequency_treshold = 5000
16
17  # get filenames
18  raw_filenames = sorted(glob.glob('*.csv'))
19
20  # sampling info
21  Fs = 10000 # [1/s] # 10 kHz
22  dt = 1/float(Fs) # [s]
23
24  # create output directory PSD if it doesn't exist
25  if not os.path.exists("PSD"):
26      os.makedirs("PSD")
27
28  # define functions
29  def find_index_of_list_containing_string(l, s):
30      for e in l:
31          if s in e:
32              return l.index(e)
```

```
33
34    def extract_experiment_stimulation(file_to_import):
35        start_string_type_1 = "MEA 2 "
36        start_string_type_2 = "MEA2 "
37
38        if start_string_type_1 in file_to_import:
39            index_start = file_to_import.index(start_string_type_1) + 6
40        elif start_string_type_2 in file_to_import:
41            index_start = file_to_import.index(start_string_type_2) + 5
42        else:
43            raise ValueError('could not find start string patterns in %s' % (file_to_import))
44
45        stop_string = "Recording-0"
46
47        if not stop_string in file_to_import:
48            raise ValueError('could not find stop string pattern in %s' % (file_to_import))
49
50        index_stop = file_to_import.index(stop_string) - 1
51
52        return file_to_import[index_start:index_stop]
53
54    def lies_in_frequency_band(Hz):
55        # return conventional frequency band for Hz
56        # according to neuro science:
57
58        # delta: [0.5, 4>
59        # theta: [4, 8>
60        # alpha: [8, 13>
61        # beta: [13, 30>
62        # gamma: [30, 120>
63        # unspecified: [0, 0.5> or [120, inf>
64
65        # assuming appropriate preprocessing
66
67        if Hz >= 0.5 and Hz < 4:
68            return "delta"
69        elif Hz >= 4 and Hz < 8:
70            return "theta"
71        elif Hz >= 8 and Hz < 13:
72            return "alpha"
73        elif Hz >= 13 and Hz < 30:
74            return "beta"
75        elif Hz >= 30 and Hz < 120:
76            return "gamma"
77        elif Hz >= 0 and (Hz >= 120 or Hz < 0.5):
78            return "unspecified"
79        else:
80            raise ValueError("could not identify frequency band of %i" % Hz)
81
82    def extract_experiment_date(file_to_import):
83        try:
84            index_start = file_to_import.index("/") + 1
85        except ValueError:
86            index_start = 0
87        return file_to_import[index_start:index_start+10]
88
89    def days_between(d1, d2):
90        d1 = datetime.strptime(d1, "%Y-%m-%d")
91        d2 = datetime.strptime(d2, "%Y-%m-%d")
92        return abs((d2 - d1).days) + 1 # regard one day experiment as one day long
93
94    def make_unscrambler_friendly_list(numpy_array_or_list):
95        # converts number into string and replaces . comma with , comma
96        return np.array([ str(number).replace('.',',') for number in numpy_array_or_list ])
97
98    def unmake_unscrambler_friendly_list(numpy_array_or_list):
99        # converts number into string and replaces . comma with , comma
100       strings = [ str(number).replace(',','.') for number in numpy_array_or_list ]
101       return np.array([ np.float64(string) for string in strings ])
102
103   def calculate_signal_duration(L, dt):
104       # return signal duration in seconds using
105       # L: the number of samples in the singal S. len(S)
106       # dt: the periodic time of the sampling. 1/Fs
107       return L*dt
108
109   def calculate_signal_mean(S):
110       return np.mean(S)
```

```python
111
112    def calculate_signal_min(S):
113        return np.min(S)
114
115    def calculate_signal_max(S):
116        return np.max(S)
117
118    # --- outer loop start ---
119
120    for experiment_index in range(0, len(raw_filenames)):
121
122        # select experiment index
123        #experiment_index = 0
124
125        # make empty pandas DataFrame to later save to csv file
126        P_pandas_total_df = pandas.DataFrame()
127
128        # chose file to import
129        file_to_import = raw_filenames[experiment_index]
130        print("Experiment info: Select experiment " + file_to_import)
131
132        if raw_csv:
133            header_row = 5
134        elif not raw_csv: # modified 1s samples A-E (TTK19)
135            header_row = 0
136
137        # Load the 61 column file
138        print("Experiment info: Load the experiment")
139        raw_file = pandas.read_csv(file_to_import, sep=",", header=header_row, low_memory=True)
140
141        # Extract header. All colums except timestamp column
142        header = list(raw_file.head(0))[1:]
143
144
145        # --- Inner loop start

146
147        for electrode_index in range(0, len(header)):
148
149            # select electrode index
150            #electrode_index = 0
151
152            # Load the selected column signal
153            print("Electrode info: Select electrode " + header[electrode_index])
154            S = np.array(raw_file[header[electrode_index]])
155
156            # compute length of signal in samples
157            L = len(S) # [#samples]
158            # compute length of signal in seconds
159            signal_duration = calculate_signal_duration(L, dt)
160
161            # compute power spectral density (PSD) with scipy (DFT using FFT)
162            print("Electrode info: Compute PSD")
163            f_fft_scipy, Sxx_fft_scipy = signal.periodogram(S, fs=Fs, scaling="density")
164
165            # downsample PSD so that len(P) ~ 5000
166            df = f_fft_scipy[1]-f_fft_scipy[0]
167            f_downsampled = f_fft_scipy[::int(1/df)]
168            P_downsampled = Sxx_fft_scipy[::int(1/df)]
169
170            # select subset of colums using
171            # simple low pass filtering using frequency_treshold
172            P_downsampled_selected = P_downsampled[0:frequency_treshold+1]
173            f_downsampled_selected = f_downsampled[0:frequency_treshold+1]
174
175            # convert elements into string 1.2345e-12 format into 1,2345e-12 because of unscrambler
176            P_unscrambler_friendly_list = make_unscrambler_friendly_list(P_downsampled_selected)
177
178            # make pandas series of PDF signal
179            P_pandas_s = pandas.Series(P_unscrambler_friendly_list)
180
181            # make pandas dataframe of P_pandas_s with associated electrode number
182            # removing [pV] part of header string
183            # and adding [pV^2/Hz] (Power density in PSD)
184            P_pandas_df = pandas.DataFrame( {header[electrode_index][0:-5] + " [pV^2/Hz]" : P_pandas_s} )
185
186            # concatenate the coloumn in P_pandas_df into P_pandas_total_df
187            P_pandas_total_df = pandas.concat([P_pandas_total_df, P_pandas_df], axis=1)
188
```

```
189              # delete things
190              del P_pandas_s
191              del P_pandas_df
192
193              # --- inner loop end, go to next electrode_index
194
195
196          # add other colums with other data and metadata extracted from file_to_import
197
198          # add average PSD per frequency column
199          PSD_averages = np.array([])
200          for row in P_pandas_total_df.itertuples():
201              PSD_average = np.mean(unmake_unscrambler_friendly_list(np.array(row)[1:]))
202              PSD_averages = np.append(PSD_averages, np.array([PSD_average]))
203
204          # convert elements into string 1.2345e-12 format into 1,2345e-12 because of unscrambler
205          PSD_averages_unscrambler_friendly_list = make_unscrambler_friendly_list(PSD_averages)
206
207          PSD_averages_s = pandas.Series(PSD_averages_unscrambler_friendly_list)
208          PSD_averages_df = pandas.DataFrame( {"Average Power Density per Hz [np.mean(P[Hz])]" : PSD_averages_s} )
209
210          # nb adding!
211          P_pandas_total_df = pandas.concat([P_pandas_total_df, PSD_averages_df], axis=1)
212
213
214          # add signal duration column
215          length_time_list = make_unscrambler_friendly_list([signal_duration]*len(P_downsampled_selected))
216
217          l_s = pandas.Series(length_time_list)
218          l_df = pandas.DataFrame( {"Duration [s]" : l_s} )
219
220          # nb adding!
221          P_pandas_total_df = pandas.concat([P_pandas_total_df, l_df], axis=1)
222
223
224          # add total energy of average PSD signal column
225          PSD_averages = unmake_unscrambler_friendly_list(np.array(P_pandas_total_df["Average Power Density per Hz [np.mean(P[Hz])]"]))
226          E_PSD = np.sum(PSD_averages*(Fs/df)) # Fs/df because the PSD was scaled by unit Hz using df
227          E_PSD_list = make_unscrambler_friendly_list([E_PSD]*len(P_downsampled_selected))
228          E_PSD_s = pandas.Series(E_PSD_list)
229          E_PSD_df = pandas.DataFrame( {"Total Energy [sum(Average Power Density per Hz[Hz]*(Fs/df))]" : E_PSD_s} )
230
231          # nb adding!
232          P_pandas_total_df = pandas.concat([P_pandas_total_df, E_PSD_df], axis=1)
233
234
235          # add mean of mean of signals column
236          means = np.array([])
237          for electrode_index in range(0, len(header)):
238              S = np.array(raw_file[header[electrode_index]])
239              means = np.append(means, np.array([calculate_signal_mean(S)]))
240          mean_list = make_unscrambler_friendly_list([calculate_signal_mean(means)]*len(P_downsampled_selected))
241
242          mean_s = pandas.Series(mean_list)
243          mean_df = pandas.DataFrame( {"Signal mean [pV]" : mean_s} )
244
245          # nb adding!
246          P_pandas_total_df = pandas.concat([P_pandas_total_df, mean_df], axis=1)
247
248
249          # add mean of min of signals column
250          mins = np.array([])
251          for electrode_index in range(0, len(header)):
252              S = np.array(raw_file[header[electrode_index]])
253              mins = np.append(mins, np.array([calculate_signal_min(S)]))
254          min_list = make_unscrambler_friendly_list([calculate_signal_mean(mins)]*len(P_downsampled_selected))
255
256          min_s = pandas.Series(min_list)
257          min_df = pandas.DataFrame( {"Signal min [pV]" : min_s} )
258
259          # nb adding!
260          P_pandas_total_df = pandas.concat([P_pandas_total_df, min_df], axis=1)
261
262
263          # add max of signal column
264          maxes = np.array([])
265          for electrode_index in range(0, len(header)):
266              S = np.array(raw_file[header[electrode_index]])
```

```
267            maxes = np.append(maxes, np.array([calculate_signal_max(S)]))
268       max_list = make_unscrambler_friendly_list([calculate_signal_mean(maxes)]*len(P_downsampled_selected))
269
270       max_s = pandas.Series(max_list)
271       max_df = pandas.DataFrame( {"Signal max [pV]" : max_s} )
272
273       # nb adding!
274       P_pandas_total_df = pandas.concat([P_pandas_total_df, max_df], axis=1)
275
276
277       # metadata extracted from the string file_to_import
278
279       # add day number of experiment
280       day_list = make_unscrambler_friendly_list([days_between(date_seeded, extract_experiment_date(file_to_import))]*len(P_downsampled_selected))
281
282       day_s = pandas.Series(day_list)
283       day_df = pandas.DataFrame( {"Day nr. from seeding date " + str(date_seeded) : day_s} )
284
285       # nb adding!
286       P_pandas_total_df = pandas.concat([P_pandas_total_df, day_df], axis=1)
287
288
289       # add date of experiment yyyy-mm-dd
290       date_list = make_unscrambler_friendly_list([extract_experiment_date(file_to_import)]*len(P_downsampled_selected))
291
292       date_s = pandas.Series(date_list)
293       date_df = pandas.DataFrame( {"Experiment date [yyyy-mm-dd]" : date_s} )
294
295       # nb adding!
296       P_pandas_total_df = pandas.concat([P_pandas_total_df, date_df], axis=1)
297
298
299       # add stimulation type
300       stimulation_list = make_unscrambler_friendly_list([extract_experiment_stimulation(file_to_import)]*len(P_downsampled_selected))
301
302       stimulation_s = pandas.Series(stimulation_list)
303       stimulation_df = pandas.DataFrame( {"Stimulation info" : stimulation_s} )
304
305       # nb adding!
306       P_pandas_total_df = pandas.concat([P_pandas_total_df, stimulation_df], axis=1)
307
308
309       # add conventional frequency bands colums, according to neuroscience
310       f_band_list = make_unscrambler_friendly_list([lies_in_frequency_band(int(np.ceil(freq))) for freq in f_downsampled_selected])
311
312       f_band_s = pandas.Series(f_band_list)
313       f_band_df = pandas.DataFrame( {"Frequency band" : f_band_s} )
314
315       # nb adding!
316       P_pandas_total_df = pandas.concat([P_pandas_total_df, f_band_df], axis=1)
317
318
319       # lastly, add "Hz" as header of index column
320       P_pandas_total_df.index.name = "Hz"
321
322
323       # save P_pandas_total_df to csv
324       print("Experiment info: Save to csv")
325       P_pandas_total_df.to_csv("PSD/" + file_to_import[0:-4] + "_PSD_pV_and_filename_metadata.csv", sep=";", encoding="ascii")
326
327       # delete things
328       del PSD_averages_s
329       del PSD_averages_df
330       del l_s
331       del l_df
332       del E_PSD_s
333       del E_PSD_df
334       del mean_s
335       del mean_df
336       del min_s
337       del min_df
338       del max_s
339       del max_df
340       del day_s
341       del day_df
342       del date_s
343       del date_df
344       del stimulation_s
```

```
345        del stimulation_df
346        del f_band_s
347        del f_band_df
348
349        del P_pandas_total_df
350        del raw_file
351
352        # --- outer loop end, go to next experiment_index
353
354    # combine all files in PSD folder
355
356    # get filenames
357    filenames = sorted(glob.glob('PSD/*.csv'))
358
359    # make empty pandas DataFrame to later save to csv file
360    P_pandas_combined_df = pandas.DataFrame()
361
362    for filename in filenames:
363        # Load the csv file
364        print("load " + str(filename))
365        raw_file = pandas.read_csv(filename, sep=";", header=0, low_memory=True)
366
367        # concatenate
368        print("concatenate")
369        P_pandas_combined_df = pandas.concat([P_pandas_combined_df, raw_file], axis=0)
370
371        del raw_file
372
373    # save P_pandas_total_df to csv
374    print("save combined csv")
375    P_pandas_combined_df.to_csv("PSD/combined.csv", sep=";", encoding="ascii", index=False)
376
377    del P_pandas_combined_df
```