AEP analysis in EEG from schizophrenic patients using PCA

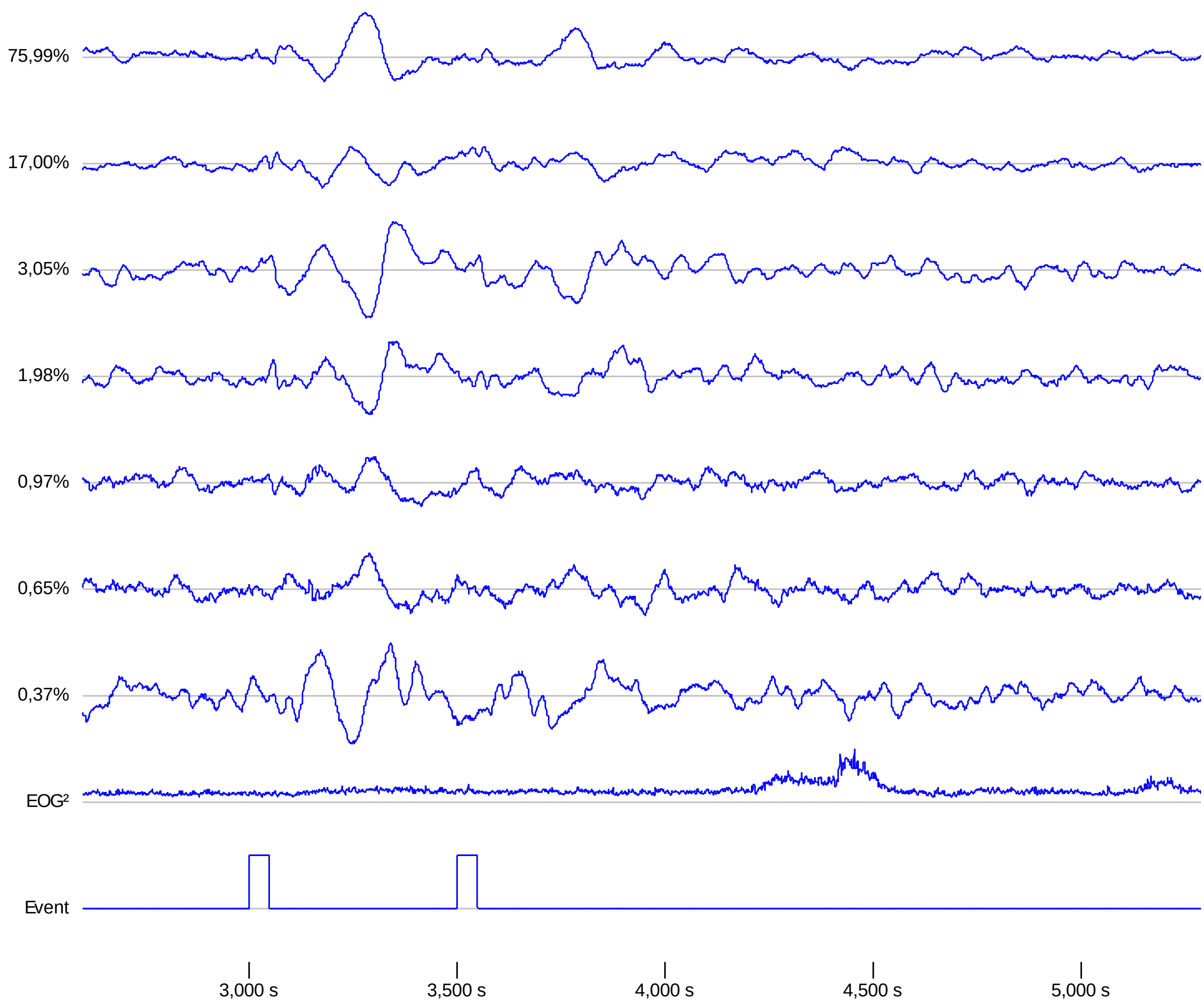
Michael Vinther, s973971 descartes@mailme.dk

Ørsted, DTU

June 7th, 2002

Supervisor: Kaj-Åge Henneberg

External supervisor: Sidse Arnfred



1 Table of contents

1 TABLE OF CONTENTS **2**

2 INTRODUCTION **3**

3 EEG **3**

3.1 Oscillation frequencies 4
3.2 Types of activity 4
3.3 Gating 6

4 PRINCIPAL COMPONENT ANALYSIS, PCA **8**

4.1 Introduction to PCA 8
4.2 The eigenvalue problem 9
4.3 Computation of principal components 13
4.4 Properties of PCA 14
4.5 Implementation 15

5 METHODS AND RESULTS **17**

5.1 Data set 17
5.2 Averaging 17
5.3 PCA 18
5.4 Gating 21
5.5 Discussion 24

6 CONCLUSION **26**

7 LITERATURE **27**

APPENDIX 1 JACOBI IMPLEMENTATION **28**

APPENDIX 2 EEG ANALYZER INTRODUCTION **29**

2 Introduction

In this report I will give a brief introduction to EEG recording and the basic types of activity that can bee seen in the EEG. Then principal component analysis (PCA) will be described as a method to represent the EEG recordings in a more efficient manner using fewer channels. The purpose of PCA is here to reduce the amount of data to be analyzed. The eigenvalue problem and a way of solving it will be discussed as the mathematical tool for performing a PCA.

Analysis of EEG recordings with responses to click sounds from a group of schizophrenic patients and a group of "normal" subjects will be used as an example application. This data set was made available to me by Sidse Arnfred, dept. of psychiatry, Bispebjerg Hospital.

3 EEG

The usual way of examining a "black-box" system where the parts cannot be separated and examined one by one is to apply some input signal and then analyze the output. One such system is the human brain, where inputs can be applied through our senses: Visual, auditory (hearing), somatosensory (feeling), olfactory (smelling) or gustatory (tasting). In this study output will be the electroencephalogram, EEG.

An EEG is recorded from a number of electrodes placed on the scalp (see Figure 1) and typically the ears are used as ground. At each electrode a superposition of the large number of brain cell (neuron) potentials is obtained. The results are weighed sums, where the weights depend on the signal path from the neurons to the electrodes. Because the same potential is recorded from more than one electrode, the signals from the electrodes are highly correlated. The potentials at the electrodes are in the μV range and therefore very sensitive to noise, so a special designed amplifier/sampler has to be used.

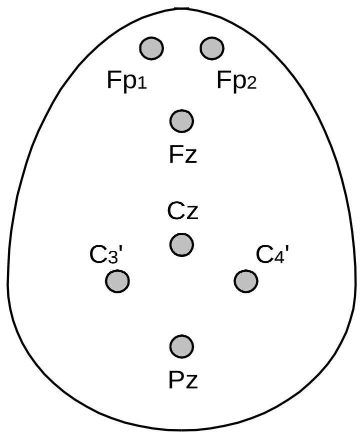


Figure 1: The data set available was recorded using 7 electrodes. The names and positions of the electrodes are shown here (the person is looking up ↑).

3.1 Oscillation frequencies

The first EEG recorded was alpha rhythms with frequencies about 10 Hz. This was done by Berger (1929), who also found and named beta activity. Beta is today used to describe the frequency range from about 12 Hz to 30 Hz. Experiments show that the amount of alpha activity varies when eyes are closed and opened, see Figure 2. Gamma frequency oscillations (30-80 Hz), which were found by Adrian (1942), are today believed to correlate with binding and attention[1]. Binding is the process of combining sensory input to form the perception of one or more objects.

In addition to alpha, beta and gamma, the ranges 0.25-4 Hz and 4-7 Hz are respectively named delta and theta. The frequency ranges are not strictly defined, and different papers use slightly different frequency intervals.

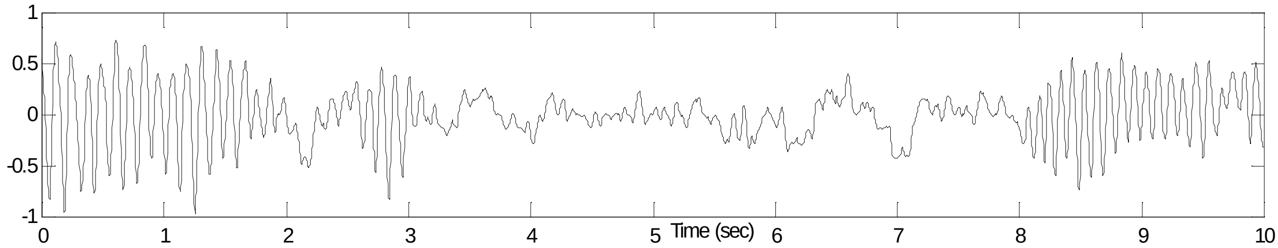


Figure 2: Ten seconds of EEG. The subject's eyes were closed during the first two seconds. The opening of the eyes first result in suppressed alpha activity, and then speeded alpha when the eyes are again closed after eight seconds[2].

3.2 Types of activity

3.2.1 Spontaneous

When recording the response to some stimuli, e.g. a sound, the EEG will always contain some activity that is uncorrelated with the stimuli. This is seen both during and in between stimulations, and is called spontaneous activity. It is caused by processes, which may be unrelated to the experiment being performed.

3.2.2 Evoked

Evoked potentials (EP) are phase-locked to the onset of stimuli meaning that every time the stimulus is applied, they appear at the same latency. Most kinds of sensory stimulation cause

[1] History of alpha, beta and gamma recordings from Herrmann, 2000

[2] Data from Herrmann, 2000

evoked potentials. The recordings, which will be analyzed in this report, contain auditory evoked potentials (AEP). These are found within the first 500 ms after stimuli.

The amplitude of evoked potentials is usually smaller than that of the spontaneous activity, and they are rarely visible in a single recording (see Figure 3). By averaging a number of recordings with the same stimulus, the other activities will cancel out and only evoked potentials remains because they are phase-locked (see Figure 4). In general the signal to noise ratio is improved with the square root of the number of epochs averaged, and with AEPs preferable more than 100 epochs should be used.

3.2.3 Induced

Like evoked potentials, induced potentials are directly caused by the experimental stimuli. But they appear with varying latency or phase, and hence they will be removed by averaging so other methods have to be applied to extract them.

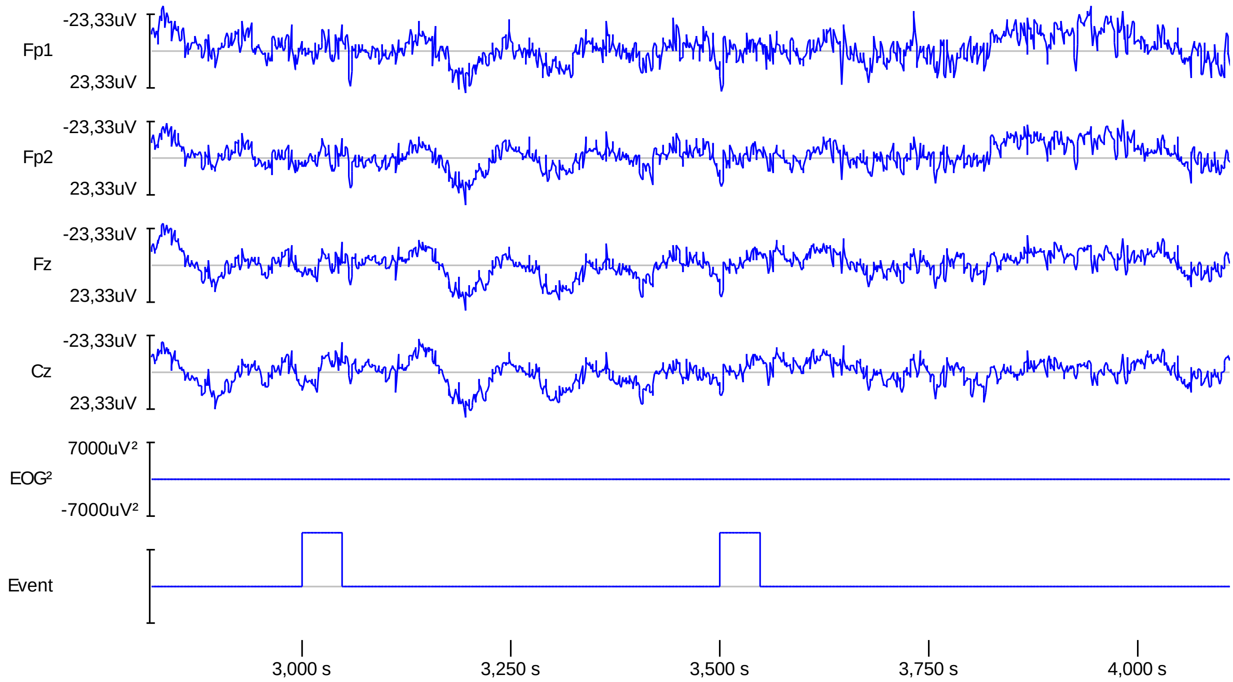


Figure 3: Four EEG channels sampled at 1 kHz. The EOG channel is used to exclude epochs with eye movement. The 50 ms pulses in the event channel signals the onset of a click with duration 1.6 ms (subject 114).

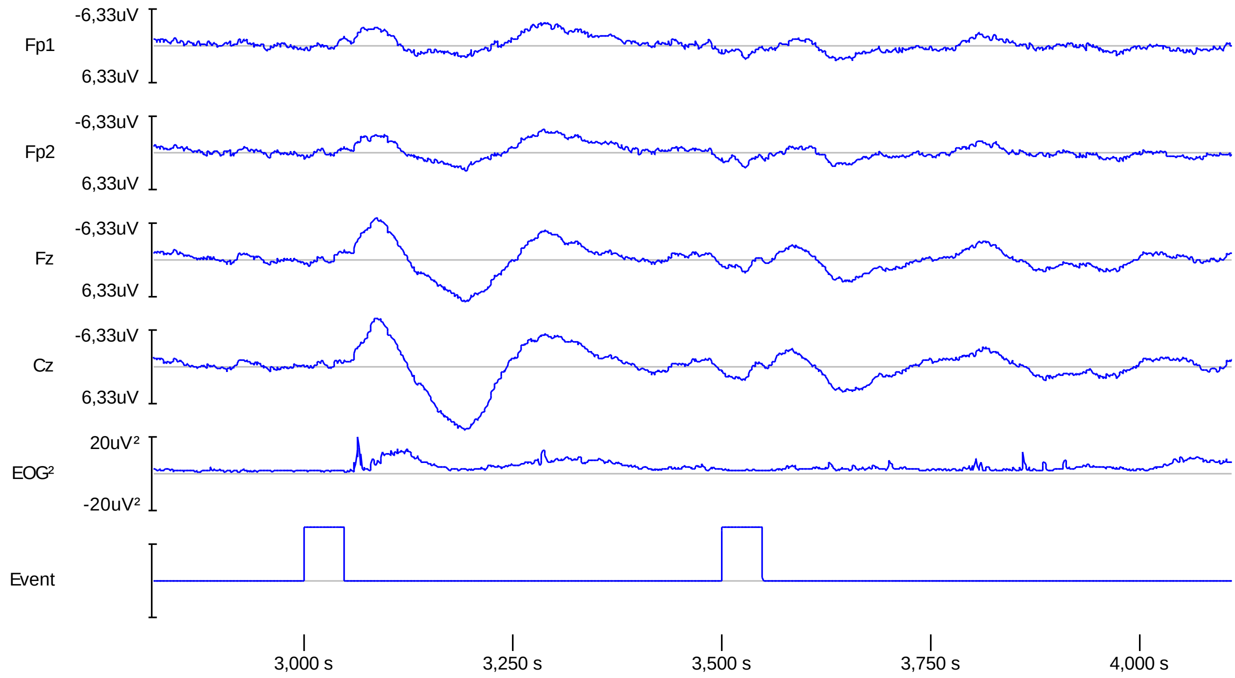


Figure 4: Average of 117 stimulus pair epochs. Here, the evoked potentials clearly stand out.

3.3 Gating

Our brain receives a large number of inputs from our senses all the time, and one of its primary tasks is to sort out what is important and what is not. One way of doing that is to focus only on changes. Figure 4 shows the evoked potentials resulting from two click sounds with an interval of ½ second. The response to the second click has much less amplitude than the first response because of this effect, which in this paradigm is called gating.

The ability to pay attention to the significant is very important to us, and it is an ability that is weakened with schizophrenic patients. Therefore the gating effect should be less visible with schizophrenic patients i.e. the amplitude difference should be smaller[3]. Of specific interest is the early response in the 50 ms latency range, as this is considered preattentive. As such the abnormality in the gating of the early responses could form the basis of the disturbance in attention and cognitive function observed in schizophrenic patients. The consistent findings in medicated patients have been reduced gating. However this has only been seen in one study of unmedicated patients.

[3] The difference in gating is demonstrated in e.g. Clementz & Blumenfeld, 2001

3.3.1 Comparing evoked responses

To compare the responses, some way of measuring the size has to be established. Traditionally the amplitude of a specific peak in the EEG has been used, see Figure 5. The peaks are named with a “P” or “N” for positive or negative and the approximate latency is ms at which they appear. As the latencies varies from subject to subject, the peak named N100 can be found after both 95 or 105 ms, so the “100” part should not be seen as exact time but just a convenient name.

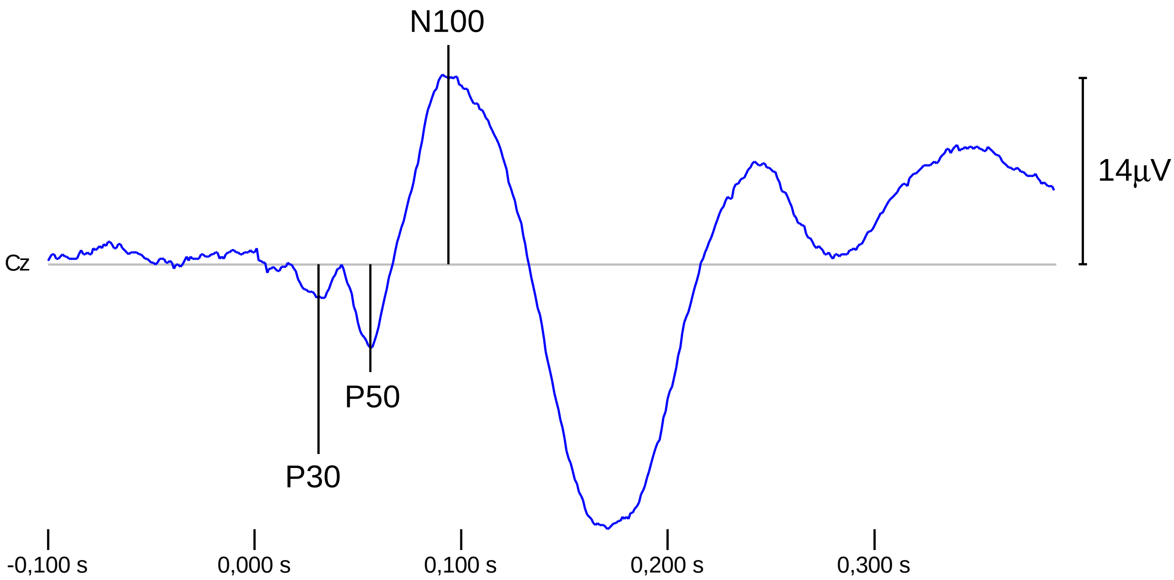


Figure 5: Typical AEP as result of a click at the time 0 s. Note that it is convention to invert the vertical axis when showing EEGs so that peaks with positive voltages point down. Here, 96 epochs from the Cz channel have been averaged to cancel out spontaneous and induced activity (subject 103).

The peaks are very distinct in the example shown here, but in some recordings they can be hard to identify. Often they have to be found by visible inspection, and sometimes they cannot be identified at all. It is very time consuming to do it by hand, but it can be necessary because an automatic peak detection algorithm would make too many errors. Other ways of measuring the size of the responses could be the average power in some time interval or the numerical value of a frequency component in a Fourier transform of the data. The advantage with these approaches is that they can be performed automatically.

The primary focus in this study has been the preprocessing (averaging, filtering and PCA), and the results presented in the following will be generated by looking at average power.

3.3.2 Frequency bands

Sometimes a band pass filtering can accentuate the response to a stimulus, and often the standard EEG bands (alpha, beta, ...) are used as a starting point for selecting a filter. The evoked potentials shown in Figure 5 have only been high pass filtered with a cutoff frequency of 1 Hz to avoid baseline wander. A band pass filtering preserving e.g. only the gamma band will change the appearance of the potentials considerably, and generally filters with a high lower bound will produce more peaks. Especially when looking at higher frequencies like the gamma band, it becomes more relevant to look at average power instead of comparing peak amplitudes.

4 Principal component analysis, PCA

4.1 Introduction to PCA

PCA is a method to transform a multi-variable data set by a rotation. A rotation is found so that the first axis corresponding to the first component is rotated to the direction where the variance of the data set is greatest. The next component will then be the direction perpendicular to the first with the most variance and so on. Figure 6 show an example with a two-variable data set where the new axes are drawn.

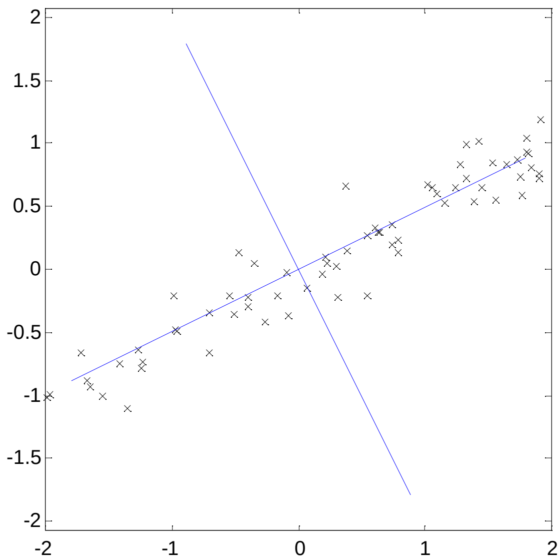


Figure 6, principal components of two-variable data set.

In this application the purpose of PCA is to reduce the number of channels to analyze. In the example it is obvious that most of the variance in the data set is along the first principal component. Assuming that variance equals information contents, one can extract almost all information in the measurement from just one channel.

The EEG recordings available here have seven channels. Because of the way signals propagate from a source in the brain to the electrodes, large signals will be measured at all electrodes and hence the channels will be highly correlated. The primary interest here is the large signals, as they relative easy can be extracted without too much noise. Therefore PCA is an appropriate tool to reduce the number of channels to analyze. The way it is used here is referred to in the literature as “spatial PCA”.

The concept of eigenvalues and eigenvectors has to be introduced before describing how to find the transformation, which gives the principal components.

4.2 The eigenvalue problem

l is an eigenvalue of the matrix $\mathbf{A}_{N\text{x}N}$ and $\mathbf{x}$ the corresponding eigenvector if

$$\mathbf{Ax} = \mathbf{x}\,l \tag{1}$$

This can only hold if

$$\det|\mathbf{A} - l\,\mathbf{I}| = 0 \tag{2}$$

Proof of that can be found in any linear algebra book. (2) can be expanded to a N’th-degree polynomial in l, whose roots are the eigenvalues. This means that there is always N eigenvalues, of which some can be equal. For a real, symmetrical matrix the eigenvalues are always real. (1) holds for any multiple of $\mathbf{x}$, but in the following only eigenvectors having length 1 will be considered.

(1) can be expressed in matrix form with a matrix $\mathbf{V}$ whose columns contain the eigenvectors and a diagonal matrix $\mathbf{D}$ with the eigenvalues in the diagonal:

$$\mathbf{A\,V} = \mathbf{V\,D} \tag{3}$$

$$\mathbf{V} = \begin{bmatrix} \begin{bmatrix} \\ \mathbf{x}_1 \\ \\ \end{bmatrix} & \begin{bmatrix} \\ \mathbf{x}_2 \\ \\ \end{bmatrix} & \cdots & \begin{bmatrix} \\ \mathbf{x}_N \\ \\ \end{bmatrix} \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} l_1 & & & \\ & l_2 & & \\ & & \ddots & \\ & & & l_n \end{bmatrix}$$

$\mathbf{V}$ is orthogonal. This means that all the eigenvectors are perpendicular to each other and hence forms an orthonormal basis. It also means that the transposed of $\mathbf{V}$ is equal to it’s inverse:

$$\mathbf{V}^{\mathrm{T}} = \mathbf{V}^{-1} \tag{4}$$

4.2.1 Diagonalization

A matrix **A** is said to be diagonalizable if a matrix **V** exists so that

$$\mathbf{D} = \mathbf{V}^{-1}\,\mathbf{A}\,\mathbf{V} \tag{5}$$

is a diagonal matrix. It is obvious that diagonalization is connected to the eigenvalue problem as (3) can be rewritten into (5). If **A** is already a diagonal matrix, the eigenvalues are simply the values along the diagonal and **V** is a unit matrix.

The transformation

$$\mathbf{A} \rightarrow \mathbf{Z}^{-1}\mathbf{A}\mathbf{Z} \tag{6}$$

is called a similarity transform. A diagonalization is hence a special case of a similarity transform. It can be shown that a this transform does not change the eigenvalues.

4.2.2 The Jacobi method

The Jacobi method[4] for calculating eigenvalues and -vectors is based on similarity transforms. The transformations are Jacobi rotations of the form

$$\mathbf{P}_{pq} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos f & & \sin f & & \\ & & & \ddots & & & \\ & & -\sin f & & \cos f & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \tag{7}$$

where the diagonal contains ones except for the elements *p,p* and *q,q*. All off-diagonal elements are zero except *p,q* and *q,p*. The idea is to do a series of rotations

$$\mathbf{A} \rightarrow \mathbf{P}_{pq}^{\mathrm{T}}\mathbf{A}\mathbf{P}_{pq} \tag{8}$$

which each forces an off-diagonal element *p,q* in **A** to zero. It is easily seen that $\mathbf{P}_{pq}^{\mathrm{T}}$ can replace $\mathbf{P}_{pq}^{-1}$ as the inverse rotation of f must be $-f$ and

$$\sin(-f) = -\sin(f)\,, \qquad \cos(-f) = \cos(f). \tag{9}$$

Every rotation creates a new zero, but at the same time ruins that created by the previous rotation. The value will, however, be less than the initial value, so by repeated rotations the off-diagonal values will decrease. It can be shown that by repeatedly selecting the numerically largest element and rotating that to, zero one can create a matrix with arbitrarily small values. As

[4] Press et al., 1987, chapter 11.

the computations in practice are done on a numerical computer with finite precision, the matrix will relatively fast be diagonalized.

sin(f) and cos(f) can be found by doing the transformation symbolically and setting the expression for a_{pq} (element p,q) equal to zero:

$$q = \frac{a_{qq} - a_{pp}}{2a_{pq}} \tag{10}$$

$$t = \frac{\operatorname{sgn}(q)}{|q| + \sqrt{q^2 + 1}} \tag{11}$$

$$\cos(f) = \frac{1}{\sqrt{t^2 + 1}} \tag{12}$$

$$\sin(f) = t\cos(f) \tag{13}$$

where sgn is the sign function: sgn(x) = 1 for $x \geq 0$ and –1 for $x<0$.

The result is a matrix with the eigenvalues in the diagonal. The corresponding eigenvectors are achieved by multiplying all of the transformation matrixes and in that way obtaining the combined transformation:

$$\mathbf{V} = \mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \ldots \mathbf{P}_n \tag{14}$$

4.2.3 Example

The eigenvectors and eigenvalues of the following matrix are sought:

$$A = \begin{bmatrix} 1 & 0.4 & 0 \\ 0.4 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{bmatrix} \tag{15}$$

The largest off-diagonal element is a_{12} = 0.4. The rotation, which produces a zero at that position, is found from expressions (10) through (13):

$$q = \frac{1-1}{2 \cdot 0,4} = 0 \tag{16}$$

$$t = \frac{1}{|0| + \sqrt{0^2 + 1}} = 1 \tag{17}$$

$$\cos(f) = \frac{1}{\sqrt{1^2 + 1}} = \sqrt{½} \tag{18}$$

$$\sin(f) = \sqrt{½} \tag{19}$$

That gives the rotation matrix

$$\mathbf{P}_1 = \begin{bmatrix} \sqrt{½} & \sqrt{½} & 0 \\ -\sqrt{½} & \sqrt{½} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

The result of the first similarity transform is then:

$$\mathbf{A}_1 = \mathbf{P}_1^{\mathrm{T}}\mathbf{A}\mathbf{P}_1 = \begin{bmatrix} 0.6 & 0 & -0.2121 \\ 0 & 1.4 & 0.2121 \\ -0.2121 & 0.2121 & 1 \end{bmatrix} \tag{21}$$

Now two elements are numerically equal. One can e.g. continue with a_{13}:

$$q = \frac{1-0.6}{2\cdot(-0.2121)} = -0.9428 \tag{22}$$

$$t = \frac{-1}{0.9428+\sqrt{0.9428^2+1}} = -0.4316 \tag{23}$$

$$\cos(f) = \frac{1}{\sqrt{0.4316^2+1}} = 0.9181 \tag{24}$$

$$\sin(f) = -0.4316\cdot 0.9181 = -0.3962 \tag{25}$$

Inserted into (7) this gives the rotation

$$\mathbf{P}_2 = \begin{bmatrix} 0.9181 & 0 & -0.3962 \\ 0 & 1 & 0 \\ 0.3962 & 0 & 0.9181 \end{bmatrix} \tag{26}$$

$$\mathbf{A}_2 = \mathbf{P}_2^{\mathrm{T}}\mathbf{A}_1\mathbf{P}_2 = \begin{bmatrix} 0.5085 & 0.0841 & 0 \\ 0.0841 & 1.4 & 0.1948 \\ 0 & 0.1948 & 1.0915 \end{bmatrix} \tag{27}$$

Now after two rotations it is clear that the off-diagonal elements have decreased. After five similar rotations the result is

$$\mathbf{A}_5 = \begin{bmatrix} 0.5000 & -0.0002 & 0.0000 \\ -0.0002 & 1.5000 & -0.0028 \\ 0.0000 & -0.0028 & 1.0000 \end{bmatrix} \approx \mathbf{D} \tag{28}$$

If two decimals are sufficient, the matrix is then diagonalized. The eigenvectors can now be found as the combined transformation:

$$\mathbf{V} = \mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \mathbf{P}_3 \cdot \mathbf{P}_4 \cdot \mathbf{P}_5 = \begin{bmatrix} 0.5656 & 0.5625 & -0.6031 \\ -0.7073 & 0.7070 & -0.0039 \\ 0.4242 & 0.4288 & 0.7976 \end{bmatrix} \tag{29}$$

4.2.4 Implementation

The method just described is implemented in the Matlab program in Appendix 1. A practical implementation is usually a little different. The search for the largest value is time consuming. Therefore all the elements are usually just selected systematically one at a time, and when all off-diagonal elements have been the target of the rotation, the selection is restarted. This requires a few more rotations, but these can be done faster. By exploiting that fact that the matrix is always symmetric one can also save a lot of computations[5].

4.3 Computation of principal components

The first component is found as the least mean square error fit to the entire data set. The next component is a least mean square error fit to the residue from the first fit and so on. The transformation matrix that gives this least mean square fit proves to be the matrix, which diagonalizes the covariance matrix of the data (apart from scaling)[6].

For N channels, the covariance matrix $\mathbf{R}$ is an $N \times N$ symmetrical matrix. The elements are defined as:

$$r_{ik} = r_{ki} = \frac{1}{T}\sum_{t=1}^{T}(s_{it} - \bar{\mathbf{s}}_i)(s_{kt} - \bar{\mathbf{s}}_k) \tag{30}$$

$\mathbf{S}$ is the data set with T samples and $\bar{\mathbf{s}}_i$ denotes the mean of channel i. If the channel power is normalized, $\mathbf{R}$ will contain correlation coefficients and the diagonal will be ones.

The matrix $\mathbf{V}$, which diagonalizes $\mathbf{R}$ can e.g. be found using the Jacobi method described above.

$$\mathbf{D} = \mathbf{V}^{-1}\,\mathbf{R}\,\mathbf{V} \tag{31}$$

After the rotation matrix $\mathbf{V}$ is found, the principal components, $\mathbf{F}$, can be computed as

$$\mathbf{F} = \sqrt{\frac{1}{\mathbf{D}}}\mathbf{V}^{\mathrm{T}}\mathbf{S} \tag{32}$$

5 Some more optimizations, which I will not describe here, are included in Press et al., 1987.

6 Glaser & Ruchkin, 1976 chapter 5.

The factor $\sqrt{1/\mathbf{D}}$ is sometimes omitted if scaling of the new components is arbitrary. It is defined as

$$\sqrt{\frac{1}{\mathbf{D}}} = \begin{bmatrix} \sqrt{\frac{1}{l_1}} & & & \\ & \sqrt{\frac{1}{l_2}} & & \\ & & \ddots & \\ & & & \sqrt{\frac{1}{l_N}} \end{bmatrix} \tag{33}$$

4.4 Properties of PCA

4.4.1 Signal power

The eigenvalues of **R** corresponds to the power contribution of the principal components to the data set:

$$l_i = P_i \tag{34}$$

This means that the first principal component is the one corresponding to the largest eigenvalue. For that reason the eigenvalues and eigenvectors are often sorted by power so that l_1 is the first principal component.

The size of the eigenvalues is an important property when selecting how many channels should be included in the analysis, and how many are to be thrown away. Usually the power contribution from each component in percent is calculated:

$$p_i = \frac{l_i}{\sum_{k=1}^{N} l_k} \cdot 100\% \tag{35}$$

One can then include enough channels to preserve e.g. 90% of the total signal power. In the present application, the seven EEG channels can be reduced to two containing about 90% of the original signal power. This is described in detail later.

4.4.2 Noise reduction

In addition to reducing the number of channels, the signal to noise ration can also be improved. As mentioned above, the signal from one powerful source is recorded at all electrodes. The new channels after the PCA transformation/rotation are actually weighed averages of the original channels, and averaging is a well-known method for noise reduction.

PCA has many applications in digital signal and image processing. One purpose is data reduction or compression as is the case here. Anther is noise removal and reconstruction. The data can be reconstructed from the principal components using the inverse transformation:

$$\mathbf{S} = \mathbf{V}\sqrt{\mathbf{D}}\ \mathbf{F} \tag{36}$$

$$\sqrt{\mathbf{D}} = \begin{bmatrix} \sqrt{l_1} & & & \\ & \sqrt{l_2} & & \\ & & \ddots & \\ & & & \sqrt{l_N} \end{bmatrix} \tag{37}$$

If some components with little power are assumed only to contain unwanted noise, they can be removed prior to the reconstruction. In that way the data is restored without the unwanted noise.

4.4.3 Component interpretation

PCA does not necessarily create channels corresponding to the original signal sources when used on a system where signals from different sources are mixed. It just collects as much variance as possible is as few channels as possible. Therefore one cannot expect that the PCA will separate signals from different parts of the brain or that it will separate the EEG from noise coming from sources outside the brain. To achieve such separation *independent component analysis*, which produces non-orthogonal transformations, could be applied. This is a more complicated method, which will not be described here.

4.5 Implementation

The effect of PCA on a highly correlated data set is demonstrated with the Matlab program in Figure 7 on the next page. The program uses the function JacobiEig shown in Appendix 1 for calculating eigenvalues and –vectors. It demonstrates both the scaled and unscaled rotation using equations (30) through (33). To the right of each program segment of is shown the plot resulting from the code.

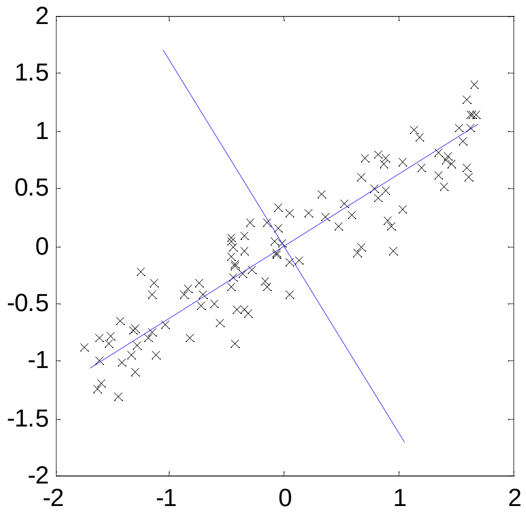
For analyzing the EEG data, an Object Pascal[7] implementation of the algorithm was made. In this implementation the scaled rotation was used, and the Jacobi method was optimized as described in 4.2.4[8]. An overview of the complete program is given in Appendix 2.

---

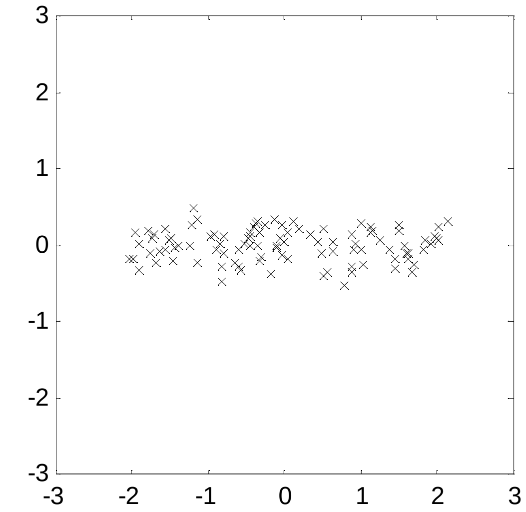
[7] Object Pascal is the language used in the Delphi compiler from Borland.

[8] My implementation of the Jacobi method is based on JACOBI.PAS and EIGSRT.PAS from Press et al., 1986.

```
% Create data with correlated channels
x = (rand(100,1)-0.5)*2.2;
S = [1.6*x x+randn(100,1)*0.25]';
plot(S(1,:),S(2,:),'kx')
% Compute covariance matrix
R = cov(S');
% Find eigenvectors and -values
[V,D] = JacobiEig(R);
% Show new axes
hold on
A = 2*V';
plot([-A(1,1) A(1,1)],[-A(1,2) A(1,2)])
plot([-A(2,1) A(2,1)],[-A(2,2) A(2,2)])
```




```
% Rotate data without scaling
F = V'*S;
figure
plot(F(1,:),F(2,:),'kx')
```




```
% Extract diagonal with eigenvalues
d = spdiags(D,0);
% Rotate and scale data
F = diag(sqrt(1./d))*V'*S;
figure
plot(F(1,:),F(2,:),'kx')
```

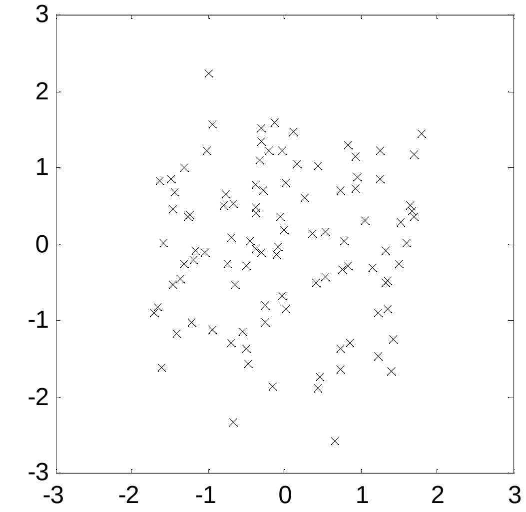


Figure 7: Demonstration of PCA with a highly correlated two-dimensional data set.

5 Methods and results

5.1 Data set

The data set used here consists of seven-channel recordings from 24 healthy subjects and 17 unmedicated schizophrenic patients, all males under the age of 55. It is sampled at 1000 Hz and high pass filtered with a cut-off frequency of 1 Hz. All subjects were presented with 120 click stimulus pairs (S1 and S2) delivered through earphones. Each click had duration 1.6 ms, there were 500 ms between onset of the clicks in a pair and 8 to 9 seconds from onset of one stimulus pair to the next, see Figure 8. Experiments have shown that when the pause between paired stimuli is at least 8 seconds, the amplitude of the P50 peak from S1 is not attenuated as result of the preceding stimuli[9]. This means that with an interval of between 8 and 9 seconds one stimulus pair should not affect the result of the next.

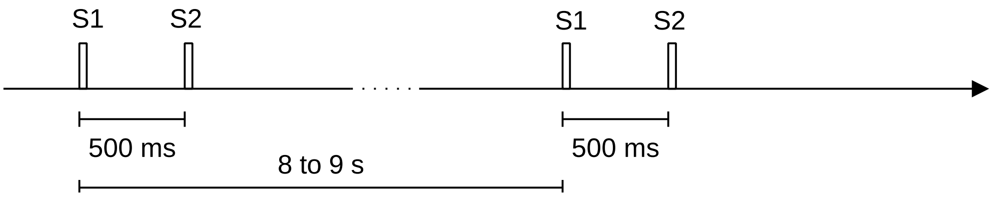


Figure 8: Timing of clicks in AEP recordings.

Besides the seven EEG channels there was an EOG channel with the voltage difference between two electrodes placed above the eyes and an event channels with markers for S1 and S2.

All results were produced using the EEG Analyzer program introduced in Appendix 2. The program is based on several other projects I have been working on, and except for PCA the implementation will not be described in detail.

5.2 Averaging

The first step of the analysis is averaging of the paired stimuli epochs to extract the evoked potentials from the other activity. Usually not all epochs can be used because of EMG (movement) and other artifacts, so only the good epochs should be included in the average. A threshold of ±70μV was used in the EOG channel, and all epochs with a sample outside this range from 3 seconds before and until 4 seconds after S1 onset was excluded. Correspondingly a threshold of ±99μV was used in the EEG channels. The A/D converter saturates at 100μV, so by

[9] Stimulus timing is discussed by Arnfred et al., 2001.

using the range ±99µV the artifacts produced by the saturation of large potentials is avoided. When only examining the 500 ms responses to the two stimuli, a narrower interval than 7 seconds could have been used, but this would prevent looking for potentials e.g. produced by the expectation of the stimuli. The effect of averaging was demonstrated in Figure 3 and Figure 4 in section 3.2. The number of epochs included in the average for each subject is shown in Table 1.

Subject	Epochs	Subject	Epochs	Subject	Epochs	Subject	Epochs
101	61	117	80	128	105	312	*46*
102	91	118	110	129	78	314	*43*
103	96	119	115	130	114	316	76
104	114	120	88	300	111	318	*1*
105	91	122	116	302	54	319	101
106	110	123	110	303	110	321	92
107	116	124	109	304	*18*	501	99
108	95	125	109	305	*3*	506	91
110	100	126	101	310	*49*	507	105
112	103	127	97	311	100	508	113
116	106						

Table 1: Number of good paired-stimuli epochs included in the averaged evoked potential. Numbers 100-130 are healthy subjects, 300-508 are schizophrenic patients.

A typical artifact seen in the first few epochs with many subjects was eye movement as result of the surprise from the loud 104 dB click. Some subjects have very few good epochs, which reduce the ratio of background activity to evoked potentials and hence the averages should not be used in further studies. I decided to exclude averages with fewer than 50 good epochs from the following experiments after visual inspection of the results. Other criteria could have been used, e.g. a minimum required SNR, and as discussed later comparing results with averages of different number epochs can also be inexpedient.

5.3 PCA

Principal component analysis of the averaged recordings was done for the remaining 35 subjects. A cross correlation coefficient matrix was computed for averages from each subject, which means that the transformation performed varied a little from subject to subject. Figure 9 show the seven channels from subject 118 before and after transformation, where the cross correlation used in the PCA was computed over 7 seconds.

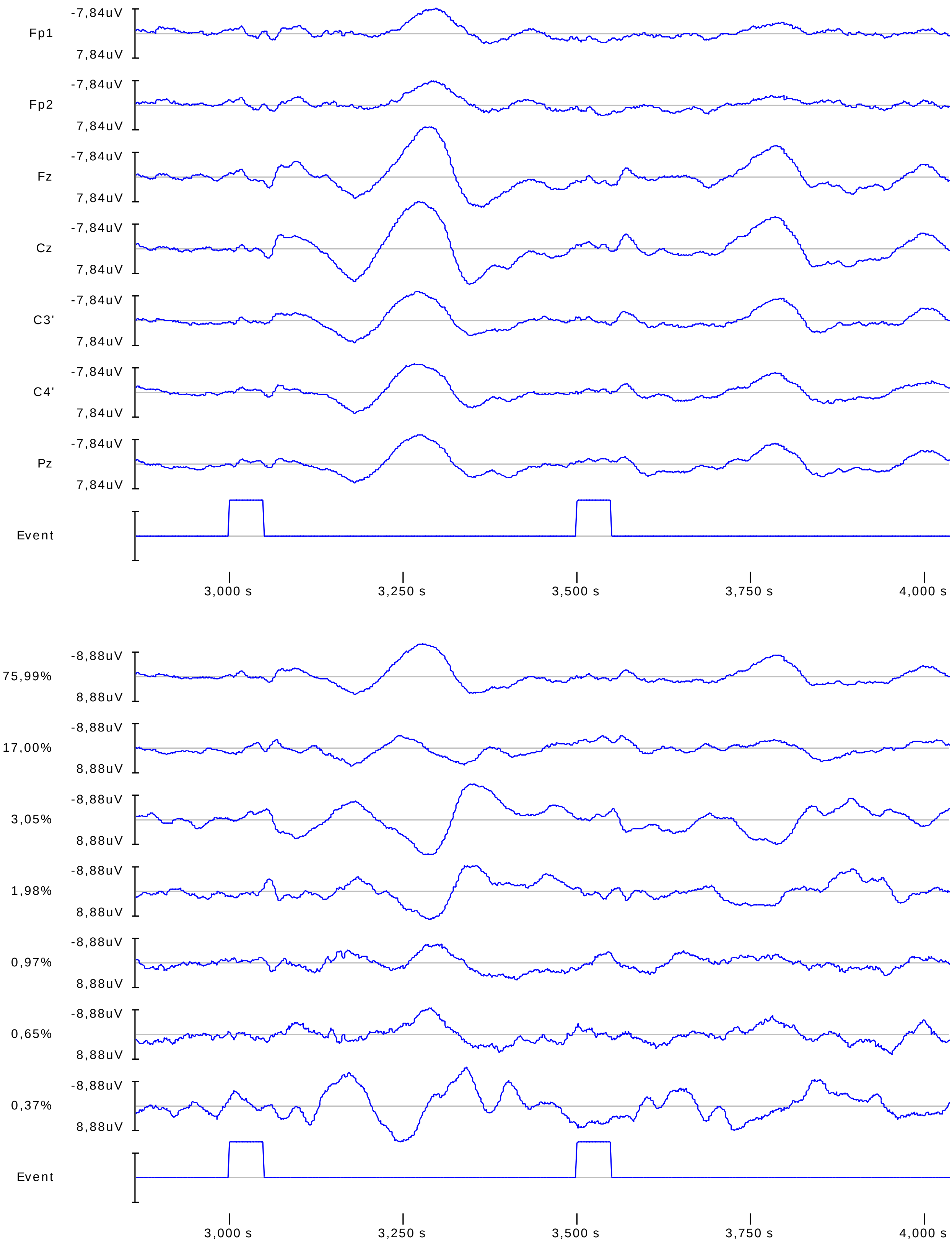


Figure 9: The seven EEG channels after the averaging and the principal components with their contribution to the signal power in %. PCA was performed on 7 second epochs (subject 108).

In the example above, the eigenvalues corresponding to the first two principal components were 5,39 and 1,19. When correlation coefficients are used, the sum of the eigenvalues are 7 as the cross correlation coefficient matrix has ones in the diagonal, so by using (35) the power contribution from the components can by found by multiplying the eigenvalues by $1/7$. The result is that the two components accounts for 76% + 17% = 93% of the signal power. The same PCA performed on all 35 recordings show that the two components account for on average 71% (std.dev. 7%) and 20% (std.dev. 6%) or in total 91% of the signal power.

If the transformation of the evoked potentials varies too much, comparing a certain component (e.g. the second) from one subject to another does not make sense. The average channel weights in the transformation matrices from all subjects are shown in Table 2 together with their standard deviations. When 7 seconds segments are used in computation of the correlation matrices, the standard deviations are relatively small and the difference in transformations is acceptable. Table 3 show the corresponding values for transformations made from 1 second epochs starting at S1, and here the standard deviations are considerable larger. Looking at the individual transformation matrices I saw that some coefficients had different sign for different subjects, even for the first component, and this cannot be explained by difference in electrode placement or signal path to the electrodes.

	Electrode	Fp1	Fp2	Fz	Cz	C3’	C4’	Pz
Component 1	**Avg. weight**	0.136	0.135	0.188	0.193	0.182	0.180	0.160
71%	**Std. dev.**	0.028	0.030	0.012	0.017	0.017	0.017	0.021
Component 2	**Avg. weight**	-0.497	-0.491	-0.129	0.135	0.212	0.235	0.357
20%	**Std. dev.**	0.104	0.083	0.063	0.063	0.067	0.084	0.135

Table 2: Average channel weights for the two first principal component. PCA was performed at the 7 second averages for all remaining subjects.

	Electrode	Fp1	Fp2	Fz	Cz	C3’	C4’	Pz
Component 1	**Avg. weight**	0.122	0.120	0.181	0.187	0.182	0.180	0.156
74%	**Std. dev.**	0.056	0.059	0.016	0.022	0.021	0.021	0.031
Component 2	**Avg. weight**	-0.433	-0.432	-0.064	0.123	0.219	0.226	0.328
20%	**Std. dev.**	0.320	0.306	0.132	0.117	0.124	0.119	0.211

Table 3: Average channel weights for the two first principal component. PCA was performed at the averages cropped to 1 second starting from S1.

With shorter epochs the first component typically contains more of the signal power i.e. is a better match for the original channels, but the transformation also varies more between subjects. In the following, PCA will be performed on 7 second epochs to make the results more comparable, and only the first two components will be considered.

About 100 Jacobi rotations were required to find the eigenvalues and -vectors for the correlation matrices in this implementation with 64 bit floating point values. Other, more efficient, methods for solving eigenvalue problems exists[10], but for matrices less than 10 by 10 the Jacobi method performs well, and the 100 rotation takes no noticeable time to perform on a standard PC.

5.4 Gating

The previous results were obtained without filtering. In this section I will describe an approach for searching for the expected difference between the two groups by looking at different frequency bands.

The high frequency contents has much less power than the low frequency contents (demonstrated in Figure 10), hence a PCA will create components, which primarily matches the low frequency activity when performed on the unfiltered data. Therefore the PCA will be performed after the filtering, and thereby produce components matching the actual frequency band only.

Figure 10 show the five frequency ranges that will be examined. In the first one (1-100 Hz) most EEG activity is contained, so primarily noise from other sources should be excluded. The next four are the alpha and beta ranges and a low and a high gamma band. The filters used are FIR filters with 501 filter coefficients created using the filter designer in the EEG Analyzer program. Because of the high filter order, the filters are close to ideal with about 10 dB/Hz slope in the transitions bands.

Recall that the 50 ms latency range is where a gating difference is expected. Gating effect is therefore computed using average power in a 50 ms interval after each stimulus as

$$r = \frac{P_{S1} - P_{S2}}{P_{S1}} \tag{38}$$

where P_{S1} and P_{S2} are averages of the squared samples. Normal gating results in $r>0$.

---

[10] Press et al., 1987

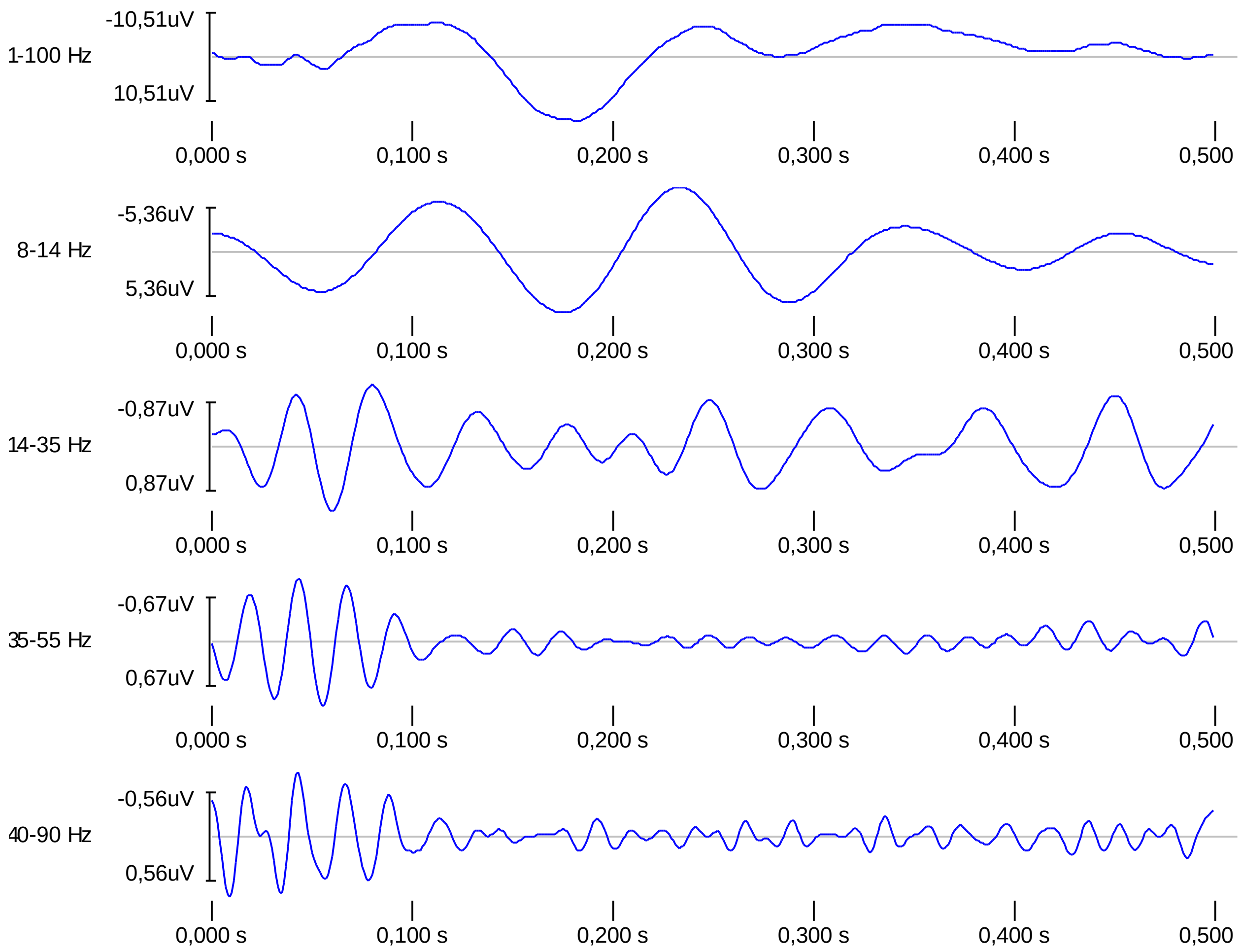


Figure 10: Frequency bands used. The plots show the first component after filtering followed by PCA. The time 0 s is at S1 onset (subject 103).

So, one subject at a time the following sequence is performed for each frequency band:

1. Filter
2. PCA
3. Find average power in the two intervals in component 1 and 2
4. Compute gating effect, *r*, in component 1 and 2

The results are shown in Figure 11 and Figure 12, where the gating of each subject is marked with a circle. The horizontal line is group mean and the vertical line is ± the standard deviation. Group 1 is the healthy subjects and group 2 the schizophrenic patients.

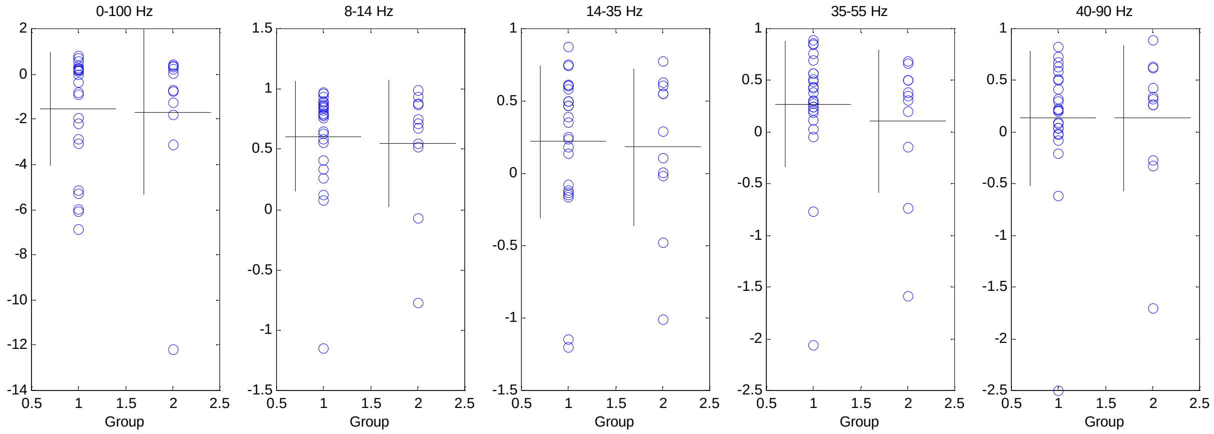


Figure 11: Average power in the first 50 ms after stimuli, component 1.

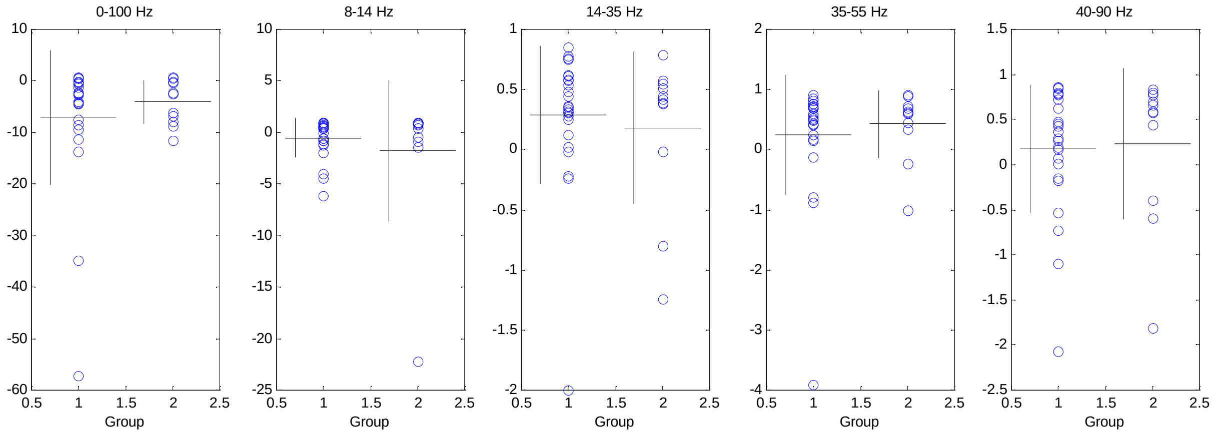


Figure 12: Average power in the first 50 ms after stimuli, component 2.

None of the plots show a significant group difference, but the gating effect is visible in all the frequency bands except for 1-100 Hz, especially in the first component. Some subjects falls far beyond the expected range and even seem to have higher S2 response than S1 response, and examination of the individual results show that it is not the same subjects which are far from the mean every time.

I tried repeating the process with average power over 500 ms from each stimulus in the first component, hoping that the longer periods would give more consistent results (see Figure 13).

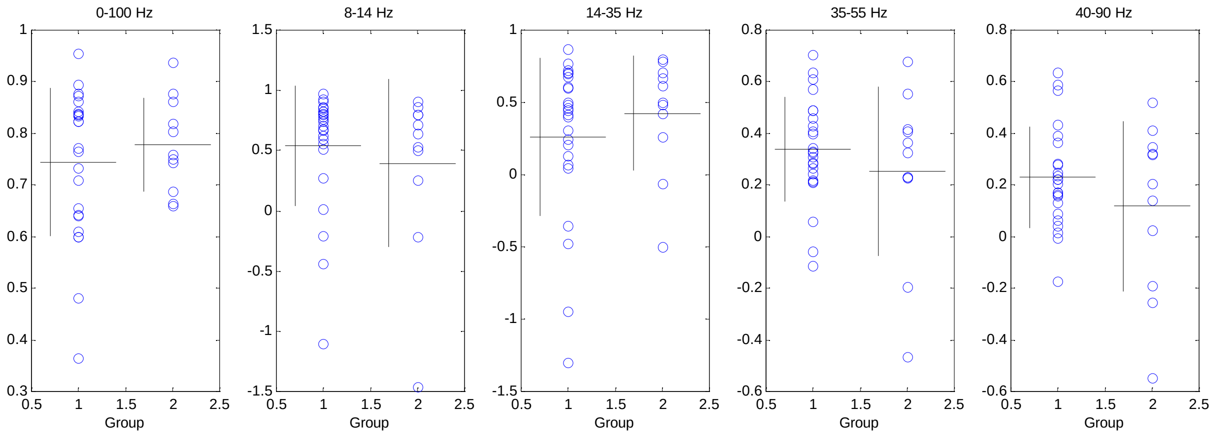


Figure 13: Average power in the 500 ms after stimuli, component 1.

As expected, the variance in these results is far less and the gating is clearly visible in all five frequency ranges, but still there is no significant difference between the groups.

5.5 Discussion

Usually one would want data being compared in some study to have been preprocessed in exactly the same way, but in this case the varying transformations might actually produce more comparable results. The EEG has already been “transformed” on its way through the brain and the skull before being recorded at the electrodes. Not all heads have the same shape and size, and the electrodes are placed a little differently at each recording, and some of these differences might be equalize by the PCA. If too short epochs are used, however, the differences can be amplified because the PCA will match specific features.

One has to be careful when interpreting the results, as group differences might be caused by other factors than the one being examined. An example could be noise: Assuming that the background noise level in a recording is constant, the gating being computed is actually

$$r = \frac{(P_{S1EP} + P_{Background}) - (P_{S2EP} + P_{Background})}{P_{S1EP} + P_{Background}} = \frac{P_{S1EP} - P_{S2EP}}{P_{S1EP} + P_{Background}} \qquad (39)$$

This is seen to decrease towards zero if the noise level rises.

Averaging tend to remove more high frequency contents than low frequency contents, because the phase looking have to be more accurate for the high frequencies to survive the averaging: If a high frequency component with a period of 10 samples has 5 samples phase-jitter from one epoch to the next, it will cancel out, but a low frequency component with a period of

100 samples with the same 5 samples jitter will only decrease a little in amplitude as result of the averaging.

As the recordings from the schizophrenic patients generally have fewer good epochs (see Table 1), their averages will contain more high frequency background activity, which is unrelated to stimuli. Hence the share of gamma activity originating from the evoked potentials is less. This suggests that an apparent difference in gating could actually just be a difference in number of epochs averaged. A possible solution to this problem might be to use the same number of epochs for all subjects, which would mean discarding good epochs from some subjects.

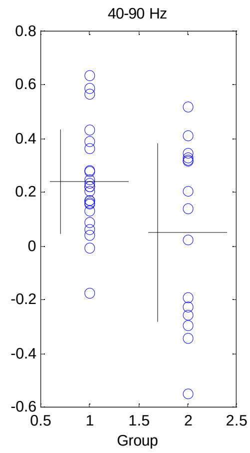


Figure 14: Gating computed on the first component using average power over 500 ms after the stimuli. When I first saw this, it looked like a significant difference between the groups, but after removing some poor recordings from the group of schizophrenic patients the difference almost disappeared (see Figure 13). This observation led to the above conclusions.

Computing the average power in the two 50 ms periods after the stimuli did not turn out be a workable approach. 50 ms is too little compared to the period of the low frequencies in the EEG (alpha waves have a period of about 100 ms), even though waves are phase locked to the beginning of the periods. A wider non-rectangular window might have produced more stable results, but as mentioned in the introduction to evoked potentials, the method of comparing average power is better suited for higher frequency bands. The wide 500 ms window more clearly showed the gating, but of course it was unable to show a gating difference in the 50 ms latency range.

6 Conclusion

In this report EEG recording, evoked potentials and the gating concept has been introduced along with some standard ways of analyzing the EEG activity.

In particular PCA was described both as a tool for analyzing EEG and as a general signal processing tool. The Jacobi method used for solving eigenvalue problems performed well in this application, and PCA effectively reduced the number of EEG channels from seven to two preserving about 90% of the variance.

The methods were implemented in a program with a graphical user interface so that it was easy to perform and see the results of various experiments on the data set. This program might also be used as a basis for further studies of EEG activity.

Unfortunately the tests showed no difference between the groups, but such a difference has also only been seen in one previous study of unmedicated patients. If a difference had been found, it could have been used to test the relevance of using PCA by comparing the size of the difference with and without using principal components.

Michael Vinther

7 Literature

Arnfred, Sidse M.; Chen, Andrew C.N.; Eder, Derek N.; Glenthøj, Birthe Y. and Hemmingsen, Ralf P.

A mixed modality paradigm for recording somatosensory and auditory P50 gating

Elsevier Science Ireland Ltd. 2001

Clementz, Brett A. and Blumenfeld, Laura D.

Multichannel electroencephalographic assessment of auditory evoked response suppression in schizophrenia

Springer-Verlag 2001

Glaser, Edmund M. and Ruchkin, Daniel S.

Principles of Neurobiological Signal Analysis

Academic Press, 1976

Hellesen, Bjarne and Larsen, Mogens Oddershede

Matematik for Ingeniører, bind 3

DTU, Institut for Anvendt Kemi

Den private ingeniørfond, 1997

Herrmann, Christoph S.

Gamma activity in the human EEG

Max-Planck-Institute of Cognitive Neuroscience, 2000

Press, William H.; Teukolsky, Saul A.; Flannery, Brian P. and Vetterling, William T.

Numerical Recipes in Pascal: The art of scientific computing

Cambridge University Press, 1986

Press, William H. et al.

Numerical Recipes: The art of scientific computing

Cambridge University Press, 1987

Proakis, John G. and Manolakis, Dimitris G.

Digital Signal Processing

Third edition

Prentice Hall 1996

Appendix 1 Jacobi implementation

The following program is a simple Matlab implementation of the Jacobi method for finding eigenvalues and eigenvectors. The primary purpose of the program is to demonstrate the method and hence it is not optimized for speed and it is not suited for badly scaled problems.

```
%[V,D] = JacobiEig(A) returns a diagonal matrix D with eigenvalues and
%a matrix V whose columns are the corresponding eigenvectors so that
%A*V = V*D.
function [V,D] = JacobiEig(A)

Size = size(A,1);
E = eye(Size);
V = E; % Start with unit matrix
for Rotations=[1:Size^2*20] % Limit number of rotations
  % Find maximum off-diagonal element
  Max = 0;
  for r=1:Size-1
    for c=r+1:Size
      if abs(A(r,c))>Max % New Max found
        p = r; q = c;
        Max = abs(A(r,c));
      end
    end
  end

  % Compare Max with working precision
  if Max<eps
    break % A is diagonalized, stop now
  end

  % Find sin and cos of rotation angle
  theta = (A(q,q)-A(p,p))/(2*A(p,q));
  t = 1/(abs(theta)+sqrt(theta^2+1));
  if theta<0
    t = -t;
  end
  c = 1/sqrt(t^2+1);
  s = t*c;

  % Build rotation matrix
  P = E;
  P(p,p) = c;
  P(q,q) = c;
  P(p,q) = s;
  P(q,p) = -s;

  % Do rotation
  A = P'*A*P;
  V = V*P;
end
D = diag(spdiags(A,0)); % Return diagonal
```

Appendix 2 EEG Analyzer introduction

EEG analyzer is a 32-bit MS-Windows program created using Borland Delphi.

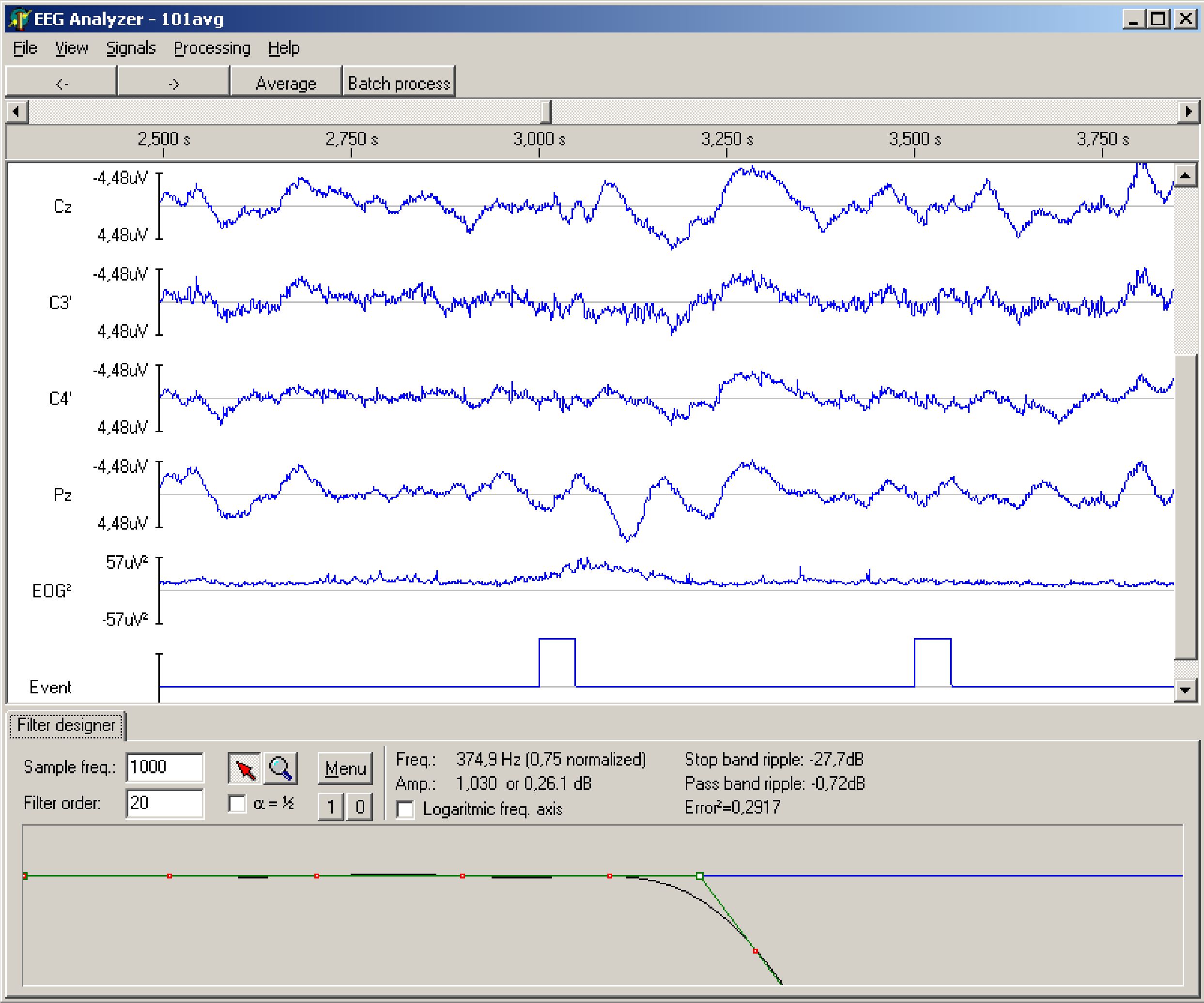


Figure 15: Screenshot from EEG analyzer. The two arrow buttons in the upper left corner changes the time scale.

Files

The program reads and writes both ordinary space delimited text files and a binary file format. The text files can be used to exchange data with other programs, but they are slow to read and write, they take up more disk space and they cannot contain extra information like the channel names. Only text files with 9 channels sampled at 1000 Hz is currently supported, as this was the format I received the test data set in.

View menu

In the view menu the scaling of the displayed signals can be changed, and the vertical axis showing the channel voltage range can be enabled or disabled. The menu item *Copy view to clipboard* copies the contents of the signal scroll box to Windows' clipboard in a vector format suited for printing, and most EEG plots in the report has been generated using this feature.

Signals menu

The signal menu only has two items: *Set noise threshold* is used for setting the allowed range for EEG and EOG activity in the averaging process. The thresholds are showed as green lines for each channel in the signal scroll box. When *Remove channels* is selected, the user is prompted for how many of the least significant PCA components should be removed

Processing menu

The *Average* item performs the averaging of the epochs based on the noise threshold set. The epoch length is auto-detected from the event channel. *PCA* performs the transformation based on either a covariance or a cross correlation coefficient matrix and show the transformation matrix used. The remaining menu items apply different filters to the EEG channels, of which I will only mention *User designed filter* here. This filter can be created using the filter designer in the bottom of the program window. Here, the amplitude response of a desired filter can be drawn using the mouse, and the program will compute filter coefficients for a linear phase FIR filter using the frequency sampling method[11]. The green line is the desired filter characteristics, the black line is the actual filter characteristics and the red dots show the location of frequency samples. All filters used here were created using this filter designer.

Batch process

The batch process is a feature which allows the user to define some sequence of filtering, PCA, cropping and computing average signal power. This sequence can then be automatically performed on a number of files and the results will be output in a text file readable by e.g. Microsoft Excel or Matlab, where statistics can be computed and visualized.

[11] Proakis & Manolakis, 1996, chapter 8.