Auditory selective attention as a method for a brain computer interface

Michiel Kallenberg[1]
Supervisors: Peter Desain[1], Stan Gielen[2]

[1]*Nijmegen Institute for Cognition and Information, Nijmegen, The Netherlands*
[2]*Department of Medical Physics and Biophysics Radboud University, Nijmegen, The Netherlands*

The object of this study was to design a Brain Computer Interface (BCI) based on auditory selective attention (ASA). ASA is a promising paradigm for a BCI, as focusing attention does not requite a lot of training, whereas the possibility of offering a large number of possible targets facilitates a high bit rate. In this study subjects focused attention on one tone out of two. The two tones were separated in space and pitch, and each tone was frequency tagged by means of amplitude modulation (AM). AM tones are known to evoke an auditory steady state response (ASSR) at the am frequency, and previous research has demonstrated that the power of this ASSR is increased by selective attention. To detect the direction of the subject's attention, features were calculated that characterized the ASSR. Subsequently, a classifier was trained based on linear discriminant analysis. The best results were obtained with a feature that consisted of the real and imaginary parts of the Fourier transformed signal at the am-frequency. Electrodes above the auditory cortices yielded the best results. On perception data, single trial classification reached a classification rate of 80%. On attention data, the best classification rate was 68%. The current BCI achieved a bit rate of 3.78 bits/min, which is moderate compared to other BCI-systems. We will discuss several procedures for improvement.

Keywords: Brain-Computer Interface, auditory selective attention, frequency tagging

Correspondence to: Michiel Kallenberg, prof Bromstraat 92-3, 6525 BG, Nijmegen, the Netherlands, e-mail: michielkallenberg@gmail.com

1. Introduction

1.1 Brain computer interfaces

In 1973 Vidal proposed the idea to build a Brain Computer Interface (BCI) : a system that facilities direct communication between one's brain and a computer. [25] Thirty years later, more than twenty research groups all over the world are working on it. The general idea of a BCI-system is depicted in figure 1.

Signals of the brain are picked up, processed and used to control to a device. A Brain Computer Interface can serve several purposes. The most promising is to (partially) restore functionality in locked-in patients. In this scheme the BCI can serve as a communication prostheses.

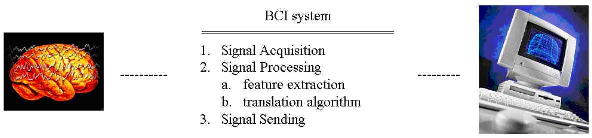

Figure 1. Basic design and operation of any BCI system. Signals from the brain are acquired by electrodes on the scalp or in the head and processed to extract specific signal features that reflect the user's intent. These features are translated into commands that operate a device.

1.2 Current systems

1.2.1 EEG

Many BCI's focus on the use of electroencephalography (EEG). EEG is the neurophysiological measurement of the electrical activity of the brain by recording from electrodes placed to the scalp. EEG is generally believed to reflect the weighted summation of post-synaptic potentials. It is based on synchronized activity from aligned neurons. The first recordings were reported by Hans Berger in the 1920s.

EEG has several limitations, of which the poor spatial resolution can be seen as the most important. This poor spatial resolution arises from broad variances in the conductance of the skull. Single trial based interferences are difficult because of a poor signal to noise ratio. Another problem lies in the artifacts that can arise from eye-, and muscle-movements.

However EEG has a lot of advantages as well. EEG is characterized by an outstanding time resolution, compared to neuro-imaging methods as fMRI and PET and the ease of use, portability, and low cost of set-up makes it attractive for BCI's.

The potential role for EEG as a basis for BCI has been further encouraged by recent basic and clinical research. Several studies have shed light on specific aspects of EEG rhythms and a variety of evoked potentials. The sites and mechanisms of origin and the relationship with specific aspects of brain function are made clearer. Numerous studies have demonstrated correlations between EEG signals and actual or imagined movements and between EEG signals and mental tasks. [26] These results have inspired BCI researchers to consider which EEG signals might be used for control and communication, and how they might best be used.

1.2.2 Examples: SCP's, sensorimotor-rhythms, P300

Most of current BCI-research focuses on slow cortical potentials, sensorimotor rhythms, and the P300 potential.

Slow cortical potentials (SCP's) are potential shifts in the scalp-recorded EEG that occur over 0.5–10 s. They reflect the level of excitability of the underlying cortical areas. Functions that involve cortical activation (deactivation) elicit positive (negative) SCP's. SCP's are detectable in every human brain, even if the motor periphery is completely disconnected from the central nervous system.

People can learn to control SCP's, as has been proved in studies of Birbaumer and colleagues. They has used this phenomenon to develop a thought translation device (TTD). They tested the system in patients with late-stage ALS and the device turned out to be helpful in restoring basic communication capabilities. [7]

Sensorimotor rhythms are rhythms in the EEG recorded over primary sensorimotor cortices. A typical rhythm of 8–12 Hz arises on the primary sensory or motor cortical areas when people are not engaged in processing sensory input or producing motor output. The amplitude of the sensory rhythms is related to the amount of sensory input and/or movement that is performed. Movement imagery can also bring about modifications in the rhythms. These findings led to the idea that BCI's can be constructed on the idea of control on the rhythm amplitudes. Indeed several sensorimotor rhythms based BCI's have proved to be successful. [8],[26]

The P300 is a positive wave that shows up in the time period between 250 and 800 milliseconds after the onset of a meaningful stimulus. The P300 is usually elicited in the so-called oddball paradigm. Subjects are presented with a series of stimuli, each belonging to one of two classes, with one class much more frequent than the other. The Oddball effect refers to the larger P300 elicited by stimuli in the infrequent class. [19].

Farwell and Donchin developed a BCI based on the P300. They presented the user a matrix of 6x6 filled with letters. The subject was instructed to choose one item of the matrix. Then the columns and rows were flashed one in a time. The EEG following each flash was recorded and scanned for a P300 signal. Combining the P300-identified column with the P300-identified row yields the chosen item.[1] Due to the high amount of items a subject can choose from (64) a lot of information can be communicated by one single trial.

1.2.3 Evaluating current systems

In evaluating the different BCI-methods one can select several relevant criteria. The first to be considered is the speed with which the system can operate, which can be expressed as the bit rate. In BCI research multiple definitions of the bit rate exist, however the definition of Wolpaw et al. is most accepted. [6] This definition integrates the accuracy of the signal-classification and the number of classes one trial can code for by the following equation:

$$B = \log_2(N) + P * \log_2(P) + (1-p) * \log_2\left(\frac{1-P}{N-1}\right) \quad (1)$$

where B=bit rate, P=proportion correct, N=number of classes.

Table 1 provides an insight in the bit rates that can be expected for the different BCI-systems. bit rates ranges from 7 to 50 bits/min; pretty low if compared to the 900 bits/min a typist can afford.

It is important to notice that the bit rate does not only depend on the intrinsic properties of the signal-class that is used, but also on the way it is processed. For instance, a couple of modifications in the P300 BCI (i.e. better classification techniques and more electrodes) increased the bit rate from 10 bits/min to almost 50 bits/min. [14].

Next to evaluating the communicating capabilities of a certain BCI, one has to look at the user friendliness. A parameter of special interest is the amount of training that is needed for a good result. Training schedules range from no training in P300 evoked potentials, to weeks or months of training for SCP's. Especially in patients with short life expectancies training should be reduced to the minimum.

1.3 Our approach: auditory selective attention

This project focuses on the use of auditory selective attention as a means for a BCI. Main advantage of this paradigm is that focusing attention does not require a lot of training. Moreover a high bit rate can be expected as, in one trial, the subject can select a class from a large set of alternatives. Previous research has shown auditory selective attention to be a promising paradigm for a BCI. In 2004 Hill et al used a dichotic listening task in which the stimuli were separated by their position in space, acoustic properties, and periodicities.

A linear support vector machine was trained to detect the direction of the subject's attention. Preliminary results were not as good as the existing BCI systems (4-7 bits/min); however, in the opinion of the authors much room was left for improvements by modifying the stimulus. [3]

In the present system we use a related design, however we make use of a different type of stimuli: In our experiment, subjects has to focus their attention on one tone out of two, which are separated in space and pitch.

1.4 Frequency tagging

In order to make sure that the tones are coming through in the EEG, each tone is watermarked with a tone-specific frequency-tag. The "tagging" occurs by means of Amplitude Modulation.

Amplitude Modulation means that the amplitude of a tone (carrier wave) is modulated by an envelope. In figure 2 this principle is clarified. Amplitude Modulation can be expressed in the following equations:

Signal-class	Training	Bit-rate	Problems
Slow Cortical Potentials	Extensive	7-12 bits/min	-
P300 evoked potentials	No training	12 bits/min	Habituation?
Mu and Beta rhytms	Moderate	20-25 bits/min	-

Table 1. valuation of current BCI systems

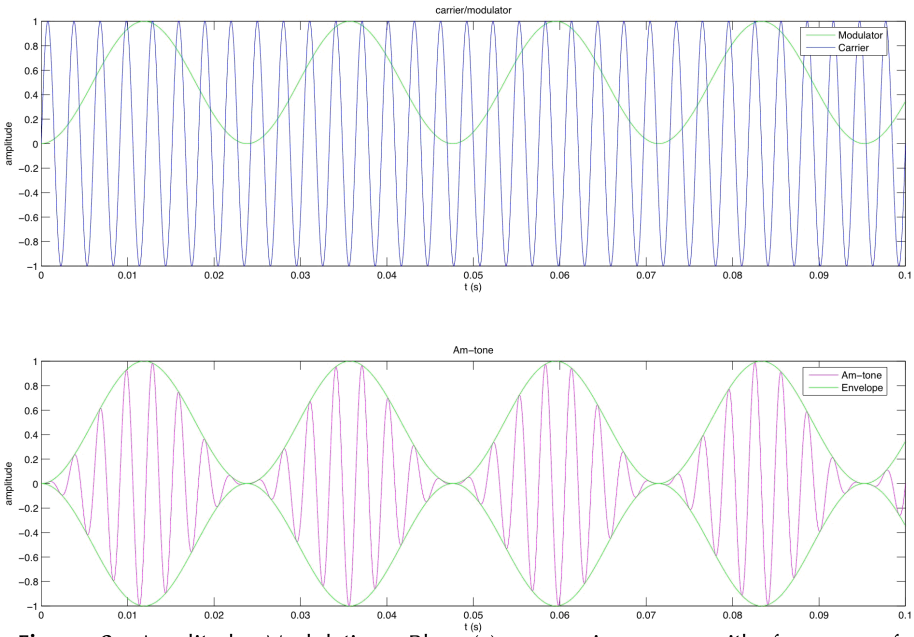

Figure 2. Amplitude Modulation. Blue (c) = carrier wave with frequency fc; green(m) = modulator with frequency fm. Pink (s) = am-tone. The AM-tone is constructed by the multiplication of the carrier wave with the modulator.

$$c(t) = \sin(2\pi t * fc) \quad (2)$$
$$m(t) = 1 - m_i * 0.5 * (\cos(2\pi t * f_m) + 1) \quad (3)$$
$$s(t) = c(t) * m(t) \quad (4)$$

where c(t)=carrier, fc=carrier frequency, m(t)=modulator, mi=modulation index, fm=modulator frequency, s(t)=signal; $0 \leq mi \geq 1$.

The concept of am-frequency tags stems from recent developed hearing tests. These hearing test all rely on the Auditory Steady State Response (ASSR), a brain-oscillation that is evoked by Amplitude Modulated tones. The interesting point of the ASSR is that its frequency corresponds to the frequency of the am-wave in the stimulus. The hearing tests benefit from this principle by presenting subjects several tones, which are than each coupled to a unique am frequency. The presence of absence of the ASSR at the frequency of the am-tag now reveals whether or not the tone is processed by the brain.

The ASSR was first described by Moller et al. in 1974. [15] Since then a lot of research has been done on its main characteristics and the processes that leads to its generation.

Currently two theories exist on the origins of the ASSR on the macroscopic level.

The first states that the ASSR reflects a phase resetting of ongoing brain-oscillations at am frequency. This phase reset leads to an increased power in the ERP, as oscillations of neurons add up over trials. On a single trial level, however no effect is seen in the power of the signal, as only the phase of the oscillation is changed. [13]

A second theory states that the ASSR reflects induced activity. According to this theory the ASSR reflects a separate neural oscillation, in addition to ongoing brain activity. The induced oscillation is facilitated by the rhythmic stimulation at best responding frequencies of the underlying neural network [21]

On the neuronal level models are developed that deal with the frequency of the ASSR. In fact it is not trivial that the ASSR is elicited at the ASSR-frequency. If we look at the pure stimulus, no energy is present at the am-frequency: the am-frequency manifests itself as a sideband to the carrier-frequency. The brain is apparently able to function as a demodulator, capable of extracting the am frequency from the signal.

Two models have been proposed that can account for this brain's competence to demodulate the signal: the first one is based on the cochlear transducer, the second on am-encoding neurons. The cochlear hair-cells and the auditory nerve form the first elements that comes into play, when a tone is presented to the ear. These elements, in assembly known as cochlear transducer, have properties that make them suitable to demodulate an am-tone. In figure 3 the working mechanism is explained: Vibrations of the air, induced by the tone, are captured by the hairs on the inner hair cells of the cochlea, thereby causing polarization and depolarization of the hair cells. Only depolarization causes the auditory nerve

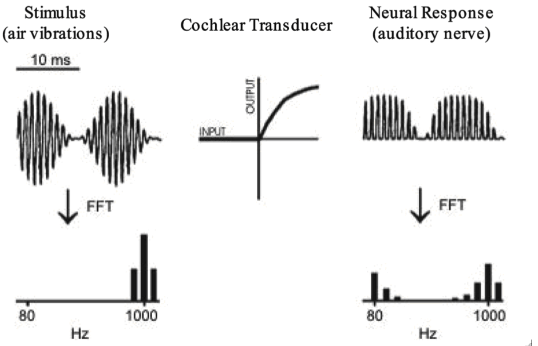

Figure 3. Demodulating by the cochlear transducer. By the vibration of the air, induced by the tone, the hairs on the inner hair cells bend, causing polarization and depolarization of the hair cells. Only depolarization causes the auditory nerve fibers to transmit action potentials. The output of the cochlea thus contains a rectified version of the acoustic am stimulus, which now has a spectral component at the AM frequency. (adapted from Lins et al. 1995)[10]

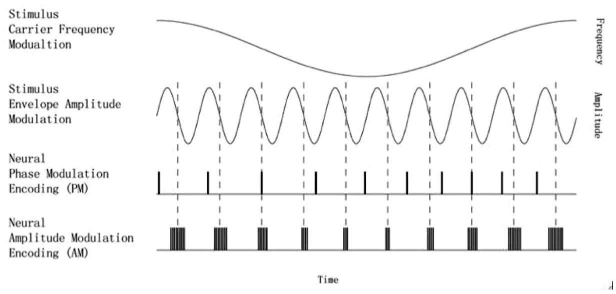

Figure 4. AM-encoding neurons are neurons that fire an action-potential burst in a phase-locked fashion to the envelope of the stimulus, while the firing rate within the bursts is matched to the carrier frequency. As a result the resulting spike train carries the am frequency. (adapted from Luo et al. 2006)[11]

fibers to transmit action potentials. The output of the cochlea thus contains a rectified version of the acoustic am stimulus. As a result the neural code now has a spectral component at the am frequency, thereby confirming the demodulating capabilities of the cochlea. [12],[10]

Neurons located in the auditory cortex, might also account for the demodulating capabilities of the brain. Figure 4 describes the so called am-encoding neurons. These neurons fire an action-potential burst in a phase-locked fashion to the envelope of the stimulus, while the firing rate within the bursts is matched to the carrier frequency. As a result the resulting spike train carries the am frequency.[10]

ASSR's are elicited along a wide range of am frequencies. However it has been suggested that for low (25-55Hz) and high (80-100Hz) modulation frequencies, distinct generators are responsible, residing in the cortex and the brainstem respectively. The ASSR has a maximum amplitude at am frequencies near 40Hz. [5],[12],[20]

The ASSR is characterized by an initial linearly rising slope during the first 200 ms after the am onset, then an interval of enhanced ASSR amplitude around 300 ms, followed by a steady state interval with constant amplitude, and finally a fast decay after stimulus offset. [22]

1.5 Effect attention on the ASSR

Until now five studies are done on the effect of attention on the ASSR.

Linden et al. performed the first EEG-experiments in 1987. In their study two paradigms were used: a dichotic listening task, in which subjects had to ignore one stream of tones in one ear, while paying attention to the stream in the concurrent ear; and a general auditory attention task, in which subjects either had to attend to the stimulus or ignore the stimulus by performing a distractor task, such as reading a book. In both cases, subjects had to count deviants that were based on carrier frequency or intensity, in the attending condition.

The stimuli consisted of 500 Hz-tone bursts which were presented at the rate of around 40 Hz. The data was analysed by averaging over trials of 500 ms duration. In the evoked potential, effects of attention were found; however no effect of attention was found on the amplitude and phase of the ASSR. [9]

Ross et al replicated parts of the Linden-study with MEG. They performed a general auditory attention task, in which the distractor task consisted of counting slideshow-pictures in three categories, whereas the attention task consisted of detecting a deviant in which a sudden change in am frequency was hidden. In contrast to Linden et al., they found an increase of the ASSR-amplitude between 200 to 500 ms after stimulus onset, to attended stimuli as compared to non-attended stimuli. They attributed the difference in findings between them and Linden to the type of deviant detection. They argued that the discrimination task used in Linden et al distracts the subjects attention from the am-frequency and therefore canceled the effect. However, it might also be the case that their improved brain imaging methods (MEG, dipole source analysis), unmasked the effect.[22]

The finding of an increased ASSR amplitude was found as well by Tiitinen et al.[24] and Gander et al. [2] in similar paradigms. Until now no clear explanation is available for the raised ASSR-amplitude. There used to be a claim that addressed the increase in amplitude to induced gamma band oscillations, however Muller et al did three odd-ball experiments that suggested that this was not the case. According to them the effect of attention is located in the primary auditory cortex, where it influences the ASSR in a direct manner. [17]

1.6 Hypothesis

This study aims to exploit the found effects of attention on the ASSR. We hypothesize that a classification based on the amplitude or phase of the ASSR is able to reveal the direction of the subjects attention.

2. Material and methods

2.1 Subject

This study describes the results of one subject, who was a 50 years old, righthanded man. The subject was untrained on auditory selective attention and amplitude modulated tones.

2.2 Design

Subjects performed an auditory selective attention task, in which they had to focus attention on one out of two streams. The streams, which consist of a sequence of amplitude modulated tones, were both separated in pitch (ie. carrier frequency) and space.

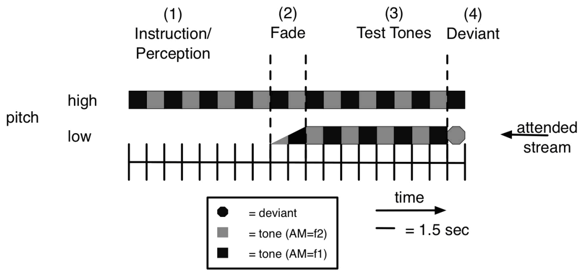

Figure 5. Stimulus. Each trial consists of five stages. In this trial subjects should focus attention on the low tone and detect the deviant at the end of the trial.

Subjects accomplished three sessions of 43 trials. Each trial consisted of four stages. (see fig 5). In the first stage (perception/instruction) the subjects were presented with one stream (either of high or low pitch), which tells the subjects to focus attention on that particular pitch during the whole trial. In the second stage (fade-in) the second, concurrent stream faded in, which the subjects were instructed to ignore. In the third stage (selective-attention) both streams were present with equal intensity and in this stage the subjects had to maintain their focus on the instructed pitch/stream. In the fourth, and last, stage a deviant was presented in one of the two streams. The subjects had to indicate by a button press whether or not they heard a deviant in the attended stream.

Subjects were told to keep their eyes fixated on a fixation cross. Furthermore they were instructed to abstain from movements till the end of the trial. In between the trials there was a self-paced pause.

A trial had an average duration of 33 seconds: Each trial consisted of 8 instruction tones, 2 fade-in tones and 8 to 10 test tones. Each tone had a duration of 1.5 seconds; there was no pause between the tones of a stream. In one in four trials no deviant occured, while in the rest of the trials the deviant was placed on the 8th, 9th or 10th tone.

2.3 Stimuli

The stimuli consisted of amplitude modulated tones: Carrier-frequencies were 1000 and 463 Hz (ie. a separation of 13.3 semitones), modulating frequencies were 42 and 32 Hz. The modulation index was set to one (see equation 2.3). The deviant differed from the standard tones by a raised amplitude (+1.3 dB).

Stimuli were generated in Matlab (The Mathworks) with a sample rate of 44.1 kHz. Audio was presented by passive loudspeakers at a distance of 0.75 m from the subject at +45 and -45 degrees.

2.4 Data acquisition

EEG was recorded using a 256 electrode system with active Ag/AgCl electrodes (BioSemi). Active electrodes are known to be efficient in keeping impedance level low. Before the start of the measurement it was made sure that the offset was kept below a threshold. Gel was inserted underneath each electrode using a semi-automatic procedure driven by air pressure. The data was digitized with a samplerate of 256 Hz.

2.5 Data-analysis

Data-analysis is done in three steps. First the EEG-signal is preprocessed, eliminating artifacts resulting from EOG and EMG. Next the features are extracted, and finally these features are used to classify the data.

2.5.1 Preprocessing

Preprocessing of the data is applied using the open source Matlab-toolbox Fieldtrip (F.C. Donders Centre, Radboud University Nijmegen). Detection of artifacts is based on thresholding of z-transformed data. Z-values are calculated by concatenating the signals over trials for each channel individually. The algorithms first bandpass-filter the data; the EOG channel is filtered with a bandpass filter between 1 and 15 Hz , the EMG channel is filtered between 110 and 128 Hz. Trials in which outliers occur that have a value larger than five times the standard deviation are removed. (22.8 % of the data) Next to artifact detection, preprocessing comprises detrending of the data, which removes the mean and a slope in each trial.

2.5.2 Feature extraction

Detection of the ASSR: the importance of phase-information As stated in the introduction there is an ongoing debate on the nature of the ASSR: Some authors claim that the ASSR stems from a phase resetting of the ongoing oscillations, where others claim that the ASSR results from an additive neural response.

This debate is very important in the selection and extraction of the features that detect an ASSR. If one assumes that the phase-resetting hypothesis holds, one should select a feature that tracks the phase of the oscillation, whereas if one supports the additive-response hypothesis one should design a feature that measures the power of the oscillation.

Figure 6 clarifies this process. In the upper panel the event related potential (ERP) of the E17-electrode is depicted for two conditions. In condition one the subject was stimulated with a 42 Hz am-tone, in condition two the subject listened to a 32 Hz am-tone. In the ERP we can see that, after the period of P1-and N1 waves, a sustained field arises. This period, which last from 500 ms to the end of the tone, is now further processed by means of spectral analysis.

First a Fourier transform is applied to the selected data, than the real and imaginary parts are selected in the bins that represents the ASSR-frequencies. In this figure we zoom in on the 42 Hz-ASSR, so we look at the 42 Hz bin and compare the real and imaginary parts from the 42 Hz-ASSR condition with the real and imaginary parts from the 32 Hz-ASSR condition. This comparison reveals the characteristics of the 42Hz-ASSR: The 32 Hz-ASSR condition does not evoke an ASSR at 42 Hz; so it can serve as a good baseline to the 42 Hz-ASSR evoked in the 42 Hz-ASSR condition.

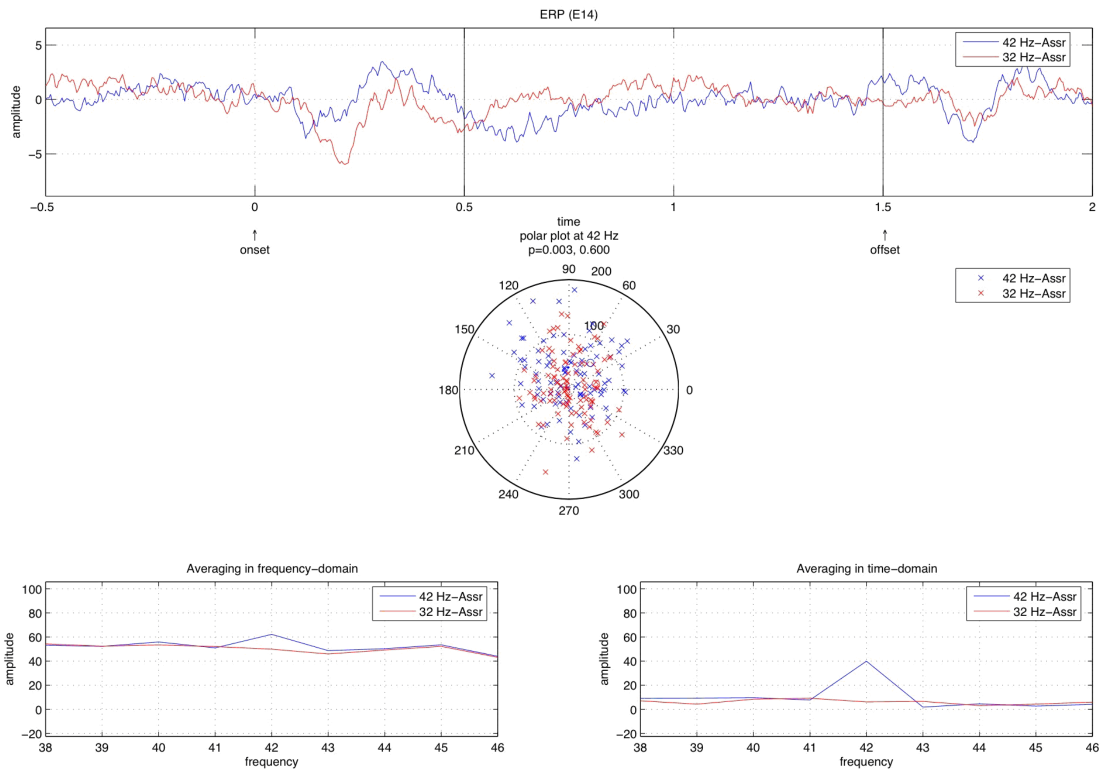

Figure 6. The importance of phase information in ASSR detection. Upper panel: ERP of 42Hz ASSR and 32 Hz ASSR; the steady state is reached after 500 ms. Middle panel: polar plot at 42Hz; ASSR phases are more clustered than non-ASSR phases. Lower panel: mean 42Hz-power plotted after two different methods of averaging.

If we now have a closer look at the polar plot in the second panel by looking at the distributions of the data points, two main findings can be extracted: First it can be seen that in the the 42 Hz condition the phases are less uniformly distributed as compared to the 32 Hz condition (Rayleigh test: p 0.003 vs p 0.600).

Second it can be seen that the powers are equal among conditions. As the difference between the two conditions reflects the pure ASSR, one may state that at least partly, the 42 Hz ASSR represents a phase-reset of the 42 Hz oscillations.

This finding has two important consequences. The first deals with the intERPretation of mean-power spectra reported in the literature; the second deals with the design of the features we want to use.

Figure 7 shows two methods that can be used to calculate the power spectra of a signal. The first method corresponds with first averaging in the time domain, the second correspond with averaging in the frequency domain. If we calculate the mean power spectra by means of the first method an effect of ASSR will be seen, as in the no-ASSR condition the oscillations will cancel out each-other, while in the ASSR-condition the oscillations will add up. However, if we use the second method, no effect will be seen as the amplitude of the spontaneous oscillations are equal in both conditions.

2.5.3 Selecting features

In the previous section we saw that an ASSR can be partly interpreted as a phase reset of ongoing oscillations. In the classification of the ASSR phase information may thus serve as an important feature.

Table 2 sums up the features that were examined. Two sets of features were created; in the first set the output of the FFT is kept in the complex plane, whereas in the second set the output is transferred to phase and power.

Feature	Description
R & I	real and imag. parts of FT at bins of interest
Rn & In	real and imag. parts of FT at bins of interest normalized to unit-circle (ie. phase)
Po	Power of FT at bins of interest
Phi & Po	Phase and power at frequency of interest
Phi	Phase at frequency of interest

Table 2. Overview of the features that were examined.

2.5.4 Classification

Classes

This project aims to classify the direction of the subjects selective attention. However, we also like to classify perception data as a simpler case and reference. At the perception side, two types of classification are interesting; The first concerns detecting of the ASSR. The second deals with detecting and classifying the stimulus that is presented to the subject.

Table 3 shows the contrasts that were examined: in the ASSR detection set four contrasts are explored and in the tone detection set two. For the classification of the attention data two contrasts are available.

Discriminant Analysis

Classification is done by applying discriminant analysis. To distinguish between both classes the algorithm fits a quadratic curve to the data. The

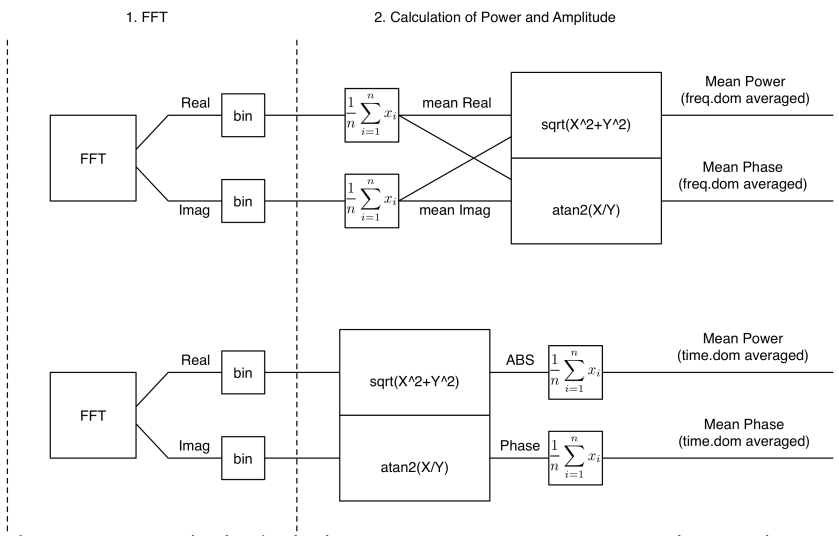

Figure 7. Two methods of calculating power spectra: averaging in the time domain vs averaging in the frequency domain.

Perception		Attention
ASSR detection	Tone detection	Attention detection
c1m1 vs c1m2 (f1 Hz) c1m1 vs c1m2 (f2 Hz) c2m1 vs c2m2 (f1 Hz) c2m1 vs c2m2 (f2 Hz)	c1m1 vs c2m2 c1m2 vs c2m1	a1c1m1c2m2 vs a2c1m1c2m2 a1c1m2c2m1 vs a2c1m2c2m1

Table 3. Contrasts used in the classification. In total, two carriers and two modulators were examined. c = carrier frequency (c1 = 1000 Hz, c2 = 463 Hz), m = modulator frequency (m1 = 42 Hz, m2 32 Hz), f = analysis-frequency (f1 = 42 Hz, f2 = 32 Hz), a = attention (ax = attention on carrier x) (e.g. c1m1 = a 1000 Hz tone, modulated with 42 Hz).

calculation of the parameters of the curve occurs by fitting multivariate normal densities with covariance estimates stratified by class. After calculation of the parameters likelihood ratios are calculated to assign observations to a class.

To prevent overfitting a leave-N-out cross-validation is used. This method works as follows: First the data is split randomly in 85% training-data and 15% testdata. Then the classifier is trained with the training-data, which is followed by classifying the test-data by the trained classifier. Then the error rate is calculated, by calculating the percentage of misclassified samples. To obtain a robust classification-rate the whole procedure is repeated twenty times, delivering twenty classification-rates. The mean of these classification-rates is than taken as the crossvalidated-error-rate.

Combining trials

To improve the accuracy of the classification multiple trials can be combined. In table 4, three methods for combining trial data are proposed.

Method	Characteristics
Mini-ERP	Not phase locked oscillations cancel out
Concatenate	Improved frequency resolution
Probabilities	Applicable for any classification method

Table 4. Methods of combining trials.

The first method constructs a mini ERP by averaging the data over multiple trials. Main advantage of this method is that it has increased sensitivity for evoked oscillations. By the process of averaging, oscillations which are not phase-locked cancel out each other, while phase locked oscillations persist. Detection of the ASSR might benefit from this method as the ASSR implies phase locking.

The second method concatenates the data. In this case only the interval in which the ASSR is at steady state is used. As the length of the signal increases the frequency resolution can improve.

The third method combines the single-trial output of the classifier. The classification algorithm calculates for each single trial the probability that the sample stems from one of the classes. These probabilities are multiplied over trials and the class with the highest probability is selected.

2.6 Statistics

To test whether the classifier performed above chance level a statistical test was introduced. Statistical testing is complicated because we have to deal with a multiple comparison problem, as there are 256 electrodes.

The null-hypothesis (H0) we tested stated that none of these electrodes had a classification rate that exceeded chance level (e.g. 0.5). The alternative hypothesis stated that there was at least one electrode that performed better than chance level, thereby implying that the best electrode we found performed above chance level. These hypotheses lead us to take the maximum classification rate as test-statistic.

In the next example this principle will be explained: Suppose the maximum classification rate over all electrodes is c. How likely is this finding under the null-hypothesis?

The probability that we find, in one electrode, a maximum classification rate (c) that does not exceed x can be found by calculating the value of the cumulative distribution function at x:

$$p(c < x) = P(x) \qquad (5)$$

where $P(x) = \int_{-\infty}^{x} p(y)dy$, which is known as the cumulative distribution function (CDF).

The probability that none of the 256 electrodes have a maximum classification rate that exceed c can be calculated as:

$$p\left(\max_{i \in (1,256)} c_i < x\right) = P(x)^{256} \qquad (6)$$

If we take the complement of this value, we find the p-value that is associated with a maximum classification rate of x or higher. Unfortunately the calculation of P(x) is complicated. The cumulative

ASSR	Contrast	Features				
		R & I	Rn & In	Phi & Po	Phi	Po
42 Hz	c1m1 vs c1m2	0.696*	0.684*	0.700*	0.705*	0.636*
	c2m1 vs c2m2	0.816*	0.782*	0.795*	0.780*	0.729*
32 Hz	c1m1 vs c1m2	0.625	0.623	0.618	0.604	0.593*
	c2m1 vs c2m2	0.807*	0.764*	0.795*	0.775*	0.730*

Table 5. Classification rates of the perception data. c = carrier frequency (c1 = 1000 Hz, c2 = 463 Hz), m = modulator frequency (m1 = 42 Hz, m2 32 Hz), (e.g. c1m1 = a 1000 Hz tone, modulated with 42 Hz)(* $p < 0.05$).

distribution function (CDF) under the null-hypothesis is not very trivial. One would assume that classification of random data (as is done under the null-hypothesis) would generate a binomial CDF; however this turned out to be not the case. Apparently the training of the classifier, as was done in the crossvalidation procedure, contradicts the standard assumptions. Therefore we estimated the H0-CDF by repeating the classification 5000 times with random data.

3. Results

3.1 Classification

The main goal of this study was to classify the direction of the subjects auditory selective attention. However, in order to examine the proper analysis-protocol we first investigated the classification of the perception data.

3.1.1 Perception

ASSR detection

First we examined the detection of the presence of the ASSR. Table 5 shows the results of this ASSR-detection, for the two ASSR-frequencies used in this study.

The table represents single-trial classification. For each feature the percentage of correct classified trials is given for the best channel.

From the table multiple findings can be extracted: First one can say that single trial-ASSR detection is good: For the best feature, classification rates range from 0.625 to 0.816 percent correctness. Overall, detection of the ASSR is best at the low carrier-frequency, and interestingly there is an interaction effect of carrier and ASSR. At the low carrier, the 42 Hz ASSR is equally detectable as the 32 Hz ASSR, whereas at the high carrier the 42 Hz ASSR carrier is better detectable than the 32 Hz ASSR.

Second one can see that R&I is the best feature. Classification with phase information leads to better results than classification on power. Detection with a feature that combines phase and power information is superior. Figure 8 shows the details of the classification of the 42Hz ASSR in one channel (D18, on top of the auditory cortex). The upper plot shows the classification based on feature R&I, the lower panel shows Phi&Po. The algorithm assigns the red surface to c2m1 and the blue to c2m2. As expected the curvature that splits the surfaces is a parabola.

The dotted lines in figure 8 represent the border of the four quadrants in a polar plot. It can be clearly seen that the red dots are more clustered than the blue dots. That is, the phases of the ASSR oscillation are more tightly locked to the stimulus than the phases of the non-ASSR oscillation. The classifier uses this information by assigning 1.5 quadrant to the ASSR while assigning the rest of the quadrants to the non-ASSR-oscillations. Also power information is used by the classifier. In general oscillations with low power are classified as non-ASSR, while oscillations with a large power are classified as ASSR.

Figure 9 shows a topoplot of the classification rates of the two different ASSR's on the two different carriers. For carrier 1 (463 Hz, left subplots) no clear location pops out, however for carrier 2 (1000 Hz, right subplots) classification is best in rightfrontal regions.

As explained in the Material and Method section combining information over trials may raise the accuracy of classification. In this study three methods of combining trials were investigated. In figure 10 the performance of these methods are plotted. It can be seen that all methods are effective in raising the performance of classification. Classification-rates of the ASSR on carrier 2 even almost get to 100%. That means that only six trials are needed to detect whether or not there is an ASSR elicited. The methods are almost equal in performance.

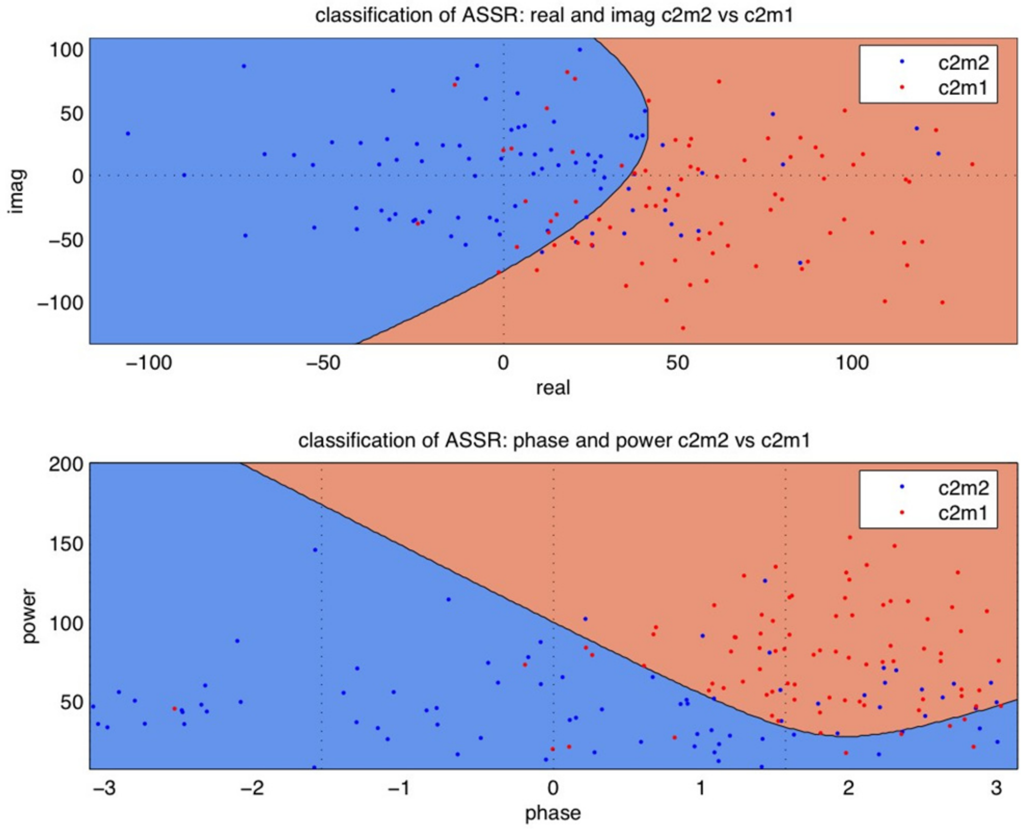

Figure 8. Details of the classification of the 42 Hz ASSR in channel D18 (above the auditory cortex).

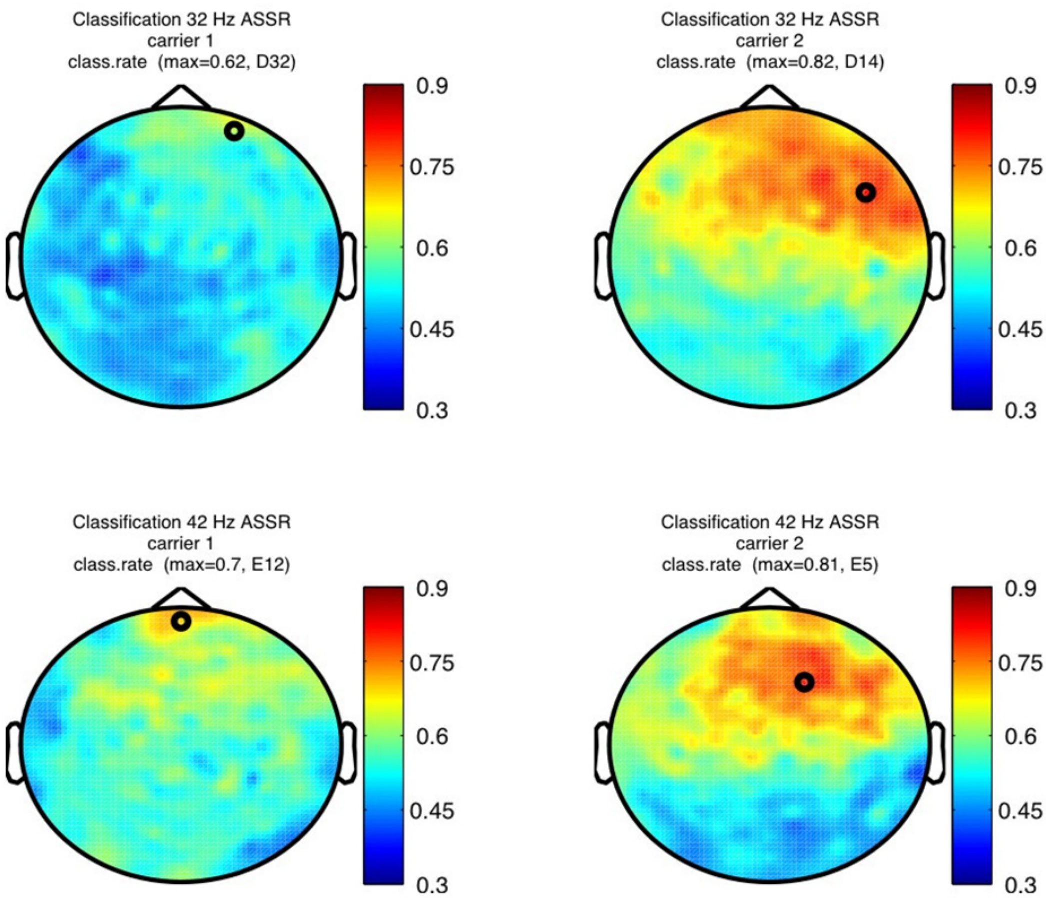

Figure 9. Topoplots of the accuracy of the classification of the ASSR. On the left side carrier 1 (463Hz) is showed, on the right side carrier 2 (1000Hz). For carrier 1 no clear location pops out, however for carrier 2 classification is best in rightfrontal regions.

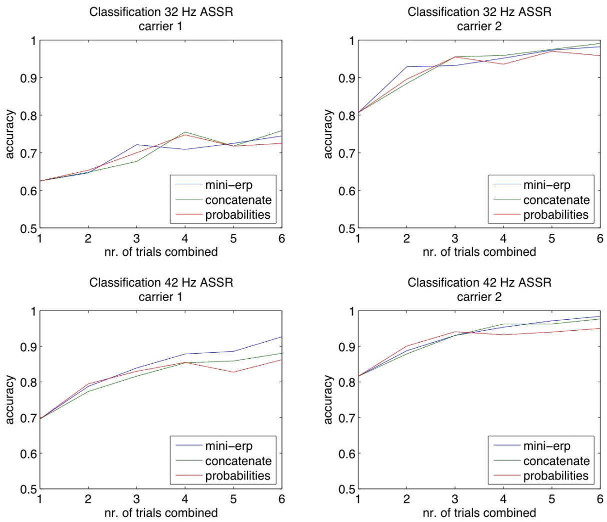

Figure 10. Combining trials. Three methods of combining trials were investigated. All methods are effective in raising the performance of classification.

Tone detection

Next to ASSR detection, detection of the tone is interesting. In tone detection two contrast are available. In each contrast two ASSR's are present, one on the first tone and one on the second tone. Therefore in classification one can select features on three levels: either from one of the two ASSR's or from both.

Table 6 shows the single-trial performance of the tone detection for the best channel. In both contrasts the tones are correctly classified in approximately 80 percent of the cases. In contrast c1m1 & c2m2 tones are best classified if one uses the parameters of both ASSR's. However in the contrast c1m1 & c2m2 classification is best if one only looks at the 42 Hz ASSR, which is on carrier 2. R&I is the best feature.

3.1.2 Attention

Table 7 shows the results of the classification of the attention data. The table shows single trial classification for the best channel. Overall performance is moderate. For the best feature classification rates ranges from 0.571 to 0.691. Interestingly the combination c1m2&c2m1 yielded better results than c1m1& c2m2. Classification rates are highest when the features are selected for both ASSR's. For the combination c1m1&c2m2 n. R&I is the best feature. This feature represents the phase of the oscillation, plotted in the complex plane. For the combination c1m2&c2m1 R&I is the best feature.

Figure 11 shows the topoplot of the classification rates of the best feature for the two different contrasts. For the combination c1m1&c2m2 no clear location pops out. For the combination c1m2&c2m1 however, clusters are detectable. High classification-rates are found in the electrodes located above the auditory cortices.

Tone	Freq	Features				
		R & I	Rn & In	Phi & Po	Phi	Po
c1m1 vs c2m2	42 Hz (c1)	0.709*	0.689*	0.689*	0.678*	0.631*
	32 Hz (c2)	0.780*	0.756*	0.750*	0.752*	0.693*
	42, 32 Hz	0.791*	0.776*	0.817*	0.769*	0.726*
c1m2 vs c2m1	42 Hz (c2)	0.797*	0.771*	0.788*	0.755*	0.733*
	32 Hz (c1)	0.666*	0.655*	0.621	0.669	0.609
	42, 32 Hz	0.784*	0.788*	0.779*	0.771*	0.719*

Table 6. Single-trial-performance of the tone detection (perception data). c = carrier frequency (c1 = 1000Hz, c2 = 463 Hz), m = modulator frequency (m1 = 42 Hz, m2 = 32 Hz) (* $p < 0.05$).

attention a1 vs a2	freq	Features				
		R & I	Rn & In	Phi & Po	Phi	Po
c1m1&c2m2	42 Hz	0.571	0.605	0.595	0.592	0.586
	32 Hz	0.588	0.600	0.585	0.582	0.585
	42,32 Hz	0.582	0.611	0.580	0.597	0.556
c1m2&c2m1	42 Hz	0.647*	0.642*	0.635*	0.644*	0.610
	32 Hz	0.592	0.613	0.610	0.584	0.603
	42,32 Hz	0.679*	0.653*	0.621	0.629	0.626*

Table 7. Accuracy of single trial-classification of attention data. a = attention, c = carrier frequency, m = modulator frequency (* $p < 0.05$).

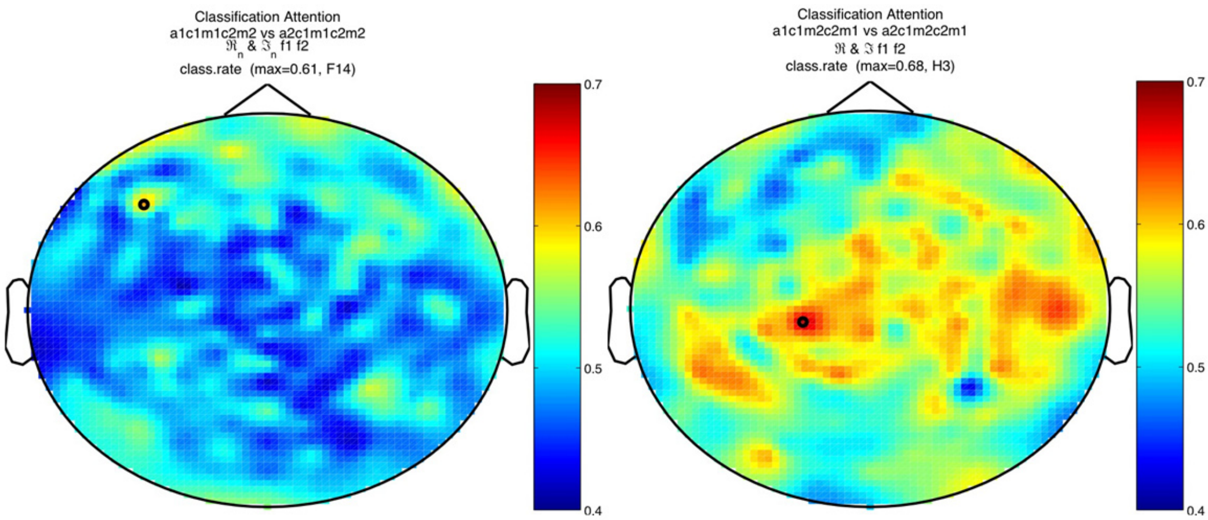

Figure 11. Topoplots of the classification-rate of the attention-data. For the combination c1m1&c2m2 no clear location pops out. For the combination c1m2&c2m1 a cluster is detectable above the auditory cortices.

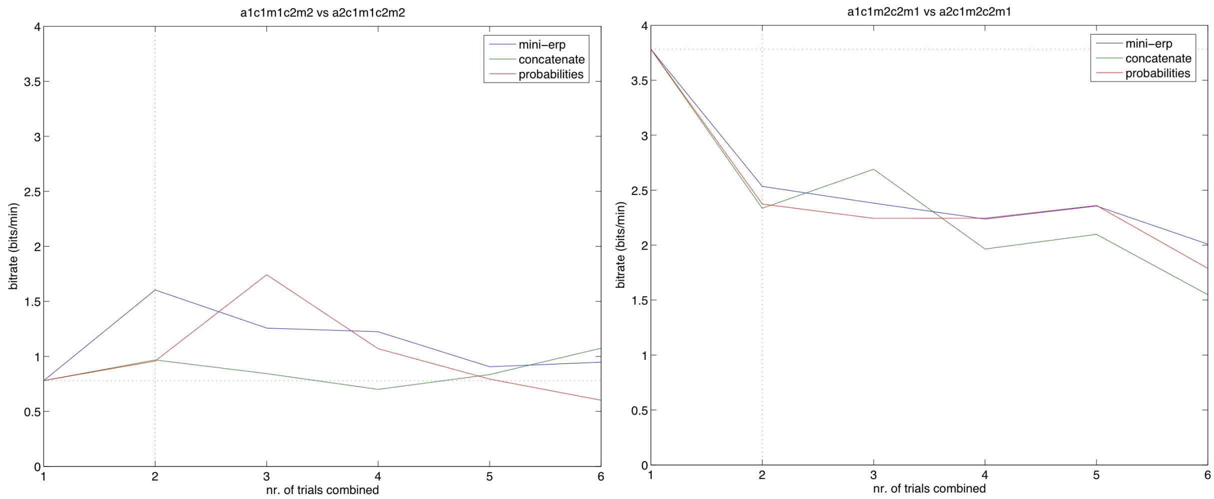

Figure 12. Bit rates. Combining over trials has a positive effect on c1m2&c2m1; for c1m1&c2m2 the bit rate drops.

3.2 Bit rate

Figure 12 shows the bit rates that we achieved in a BCI. For single trial classification the bit rate is 3.78 bits/min, for c1m2&c2m1 and 0.78 bits/min, for c1m1&c2m2.

As described in the Materials and Method section combining trials may raise the bit rate. Three methods of combining trials were investigated: the mini-ERP, concatenation, and probabilities.

For each contrast combining trials raised the accuracy of the classification; however the effects on the bit rate are mixed, due to the longer time it takes to classify. For c1m2&c2m1 combining trials decreased the bit rate, whereas for c1m1&c2m2 combining trials had a positive effect on the bit rate. For this combination, the best result was reached when the method of combining probabilities was chosen. The combination of three trials yielded a classification rate of 71%, which results in a bit rate of 1.62 bits/min.

4. Conclusion and discussion

This study focused on the use of auditory selective attention as a method for a Brain Computer Interface. Our hypothesis was that in the EEG, amplitude and/or phase of a frequency tagged tone alters if a subject focuses on this tone.

As a result we can classify on which of two tones a subjects focuses. Frequency tagging is done by amplitude modulation.

Before we examined the effect of attention, we examined the perception data as a reference. The results will be discussed in the same order. Finally, we will discuss the resulting BCI.

4.1 Perception

On the perception data we performed two analyses: the first considered the detection of the ASSR, the second considered the classification of the perceived tone. Detection of the ASSR was best at the low carrier frequency (463 Hz), while the 42 Hz am-frequency performed better than the 32 Hz am-frequency. The latter finding was not surprising: similar results were reported in the literature.

However, the role of the carrier frequency is more controversial. Ross et al [20] showed that the amplitude of the ASSR is maximal at higher carrier frequencies, while John et al [5] showed the opposite.

The best feature was the real and imaginary parts of the FFT in the am-bin. Features based on phase information were (slightly) better than features based on power, thereby suggesting a role of the phase-resetting hypothesis in the generation of an ASSR. Electrodes above right-frontal regions performed best in the classification of the ASSR. IntERPretation of this finding is difficult as classification rate does not simply reflect potentials, or power of an oscillation, thereby making source estimation less trivial. However in figure 4.1 it was argued that a good electrode-performance reflects the detection of an increase in power or the phase reset of an ongoing oscillation. If we model the ASSR as a dipole, electrodes that picks up currents that results from this dipole will therefore show a good classification rate. Therefore one could argue that the source of the ASSR lies underneath the electrodes that performed well, although the orientation of the dipole is of importance.

Often the ASSR is modeled as a dipole originating in the primary auditory cortex. Regardless of whether a tone is presented to one ear or both, the ASSR is elicited in both hemispheres. Our data differed slightly from this finding, because topological symmetry was not complete. However, deviations were small.

Tone detection was best if one extracts the features from carrier 2.

4.2 Attention

Classification of the attention data was much harder than classification of the perception data. Maximum classification-rate was 0.691 for single trial classification in the best channel. The combination c1m2 & c2m1 yielded better results than the combination c1m1 & c2m2.

Interestingly, a parallel between perception and attention can be seen here: In perception data we saw that the best results were obtained by the ASSR on the low carrier (c2). Besides, the 42 Hz ASSR (m1) was better detectable than the 32 Hz ASSR (m2). Now, in the (superior) combination c1m2 & c2m1 the best carrier (c2) carried the best am (m1), whereas in the combination c1m1 & c2m2 the best carrier (c2) carried the weakest am (m2). Apparently an attention effect is best detected when an ASSR effect is strongest.

Attention effects were largest in electrodes on top of the auditory cortex. This finding supported results found by Muller et al. [17] and Ross et al. [21].

4.3 Evaluating the BCI

As stated in the introduction the most important criterion to compare BCI-systems is the bit rate. If our experiment is implemented in a BCI, we expect a BCI a bit rate of 3.78 bits/min.

This bit rate is moderate compared to state-of-the art-BCI-systems –which operates with a bit rate of 20 bits/min– 3.78 bits/min. However as we will show later, in this system there is room left for improvement.

To date there is only one more BCI system that uses auditory selective attention. [3] In this system a bit rate of the same range was found. A second criterion to evaluate the BCI-system is the amount of training that is needed. Pilot studies showed that directing attention on one tone does not require a lot of training. In this study, subjects were not trained.

4.4 Recommendations

As stated before, there may be enough room left to improve current BCI-system, although doubt is raised whether the effect of attention will be large enough to be classified at a signal trial level. Improvements can be made by adjusting the stimuli, the classification methods and the system.

4.4.1 Stimuli

In the current experiment stimuli were presented by passive loudspeakers at some distance and angle from the subject. As a result, even tough tones were separated in space, each tone could reach both ears. Consequently, the brain was confronted with both an am tone and a phase delayed copy of that am tone. Possibly these signals could interfere thereby obscuring the ASSR. Presenting the stimuli by a headphone would overcome this issue.

The deviant that subject had to detect differed from their standards by means of its amplitude. As suggested by Ross et al. [22] this type of deviant might distract the subjects attention from the modulation. As a result ASSR effects might be diminished. Better results might be obtained if one designs a deviant that differs in amplitude modulation.

Another option to improve the bit-rate of the BCI system is to present more concurrent stimuli in one trial, or to use shorter tones. By increasing the number of tones the subjects can select from, the amount of information that can be transferred in a single trial will increase. It is important however that classification rate does not drop. From previous experiments it is known that am-tones with carrier frequencies separated more than an octave do not interfere. If the am frequency is in the higher frequency bands (80 Hz) the carrier frequencies can even be separated less. am frequencies are less susceptible to interference, however it is advisable to keep them well separated. If the am frequencies are too close, one has to design sharp bandpass-filters of inferior quality. Moreover the frequency resolution has to be high, which can only be established by analyzing over long time periods, which drops the bit-rate. In stead, the aim is to use shorter tones thereby increasing bit-rate.

4.4.2 Feature extraction and classification

In the current experiment features were extracted by taking the FFT of the ASSR's steady-state-period after which the real and imaginary parts were selected in the am-frequency-bin. However in this approach temporal information is lost, as we used a fixed, non sliding time window. Time-frequency analysis (TFA) is able to solve this problem. In the ASSR-literature mainly two TFA-methods are used: complex demodulation and the Hilbert transform. Applying those methods might lead to better feature-design and therefore bit-rate may improve. Better results can also be obtained by combining data of multiple channels. Currently, features and classification are solely based on data of one channel. Channel-data can be combined on three different levels: the signal, the feature and the classification level. An improvement on the signal level would be source analysis. This could be done by dipole-modeling, as has been successful in other ASSR-studies. Applying beamforming methods, as has been done in tactile frequency tagging experiments, might not function for auditory experiments, as we are dealing with correlated sources. Combining channels on the feature level would imply a feature like inter-channel correlation. Channel combination on the classification level refers to methods in which for instance a certain number of channels should have classified the trial as class A to assign class A to the trial.

4.4.3 System

The current BCI-system may be improved by using more modalities. For instance one can combine the system with a tactile paradigm. Recently a tactile BCI was developed by Hupse et al.[4]. The BCI they developed also operated with frequency tagged stimuli.

A combination with visual stimuli may also function. Also in this domain it has been shown that (spatial) selective attention has an effect on frequency tagged stimuli [16],[18].

4.5 Conclusion

To conclude, in perception data classification of the ASSR was possible at a single trial level. Best feature was the real and imaginary parts of the FFT at the bin in which the frequency of the frequency tag was represented. Best classification rates were found in midfrontal regions. For attention data classification rates were moderate, resulting in a bit rate of 3.78 bits/min.

Auditory selective attention on frequency tagged tones therefore seems to be a moderate method for a BCI, however there are possibilities left to improve the system. To draw final conclusions, though, these possibilities should be tested in a follow up study including more subjects.

Acknowledgements

We thank the BCI-team for their support: Peter Desain for his enthusiasm and insightful comments, Stan Gielen for his constructive feedback, Bas de Kruif for his help with the analysis, Rebecca Schaefer, Alex Brandmeyer and Philip van den Broek for their help in the setup and Rianne Hupse and Denise van Barneveld for their friendly support. Thanks goes also to Christoph van der Reijden and Lucas Mens for their collaborative support, and Wilbert Kallenberg for his statistical advice.

References

[1] E Donchin, KM Spencer, and R Wijeshinghe. The mental prosthesis: Assessing the speed of a p300-based brain-computer interface. IEEE Transactions on Rehabilitation Engineering, 8(2):174–179, 2000.

[2] P Gander. Attentional modulation of the 40-hz auditory steady state response and other auditory evoked potentials during acoustic training. Master's thesis, McMaster University, 2005.

[3] NJ Hill, TN Lal, K Bierig, NP Birbaumer, and B Schoelkopf. An auditory paradigm for brain computer interfaces. www.kyb.mpg.de/publications/pdfs/pdf2794.pdf, 2004.

[4] R Hupse and P Desain. Master's thesis, Radboud University Nijmegen, 2006.

[5] MS John, D Purcell, A Dimitrijevic, and TW Picton. Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli. J Am Acad Audiol, 13:246–259, 2002.

[6] J Kronegg, S Voloshynovskiy, and T Pun. Analysis of bit-rate definitions for brain-computer interfaces. In 2005 Int. Conf. on Human-computer Interaction (HCI'05), Las Vegas, Nevada, USA, 2005.

[7] A K ̈ubler, N Neumann, J Kaiser, B Kotchoubey, T Hinterberger, and Birbaumer NP. Brain-computer communication: self-regulation of slow cortical potentials for verbal communication. Arch Phys Med Rehabil, 82:1533–1539, 2001.

[8] A K ̈ubler, F Nijboer, J Mellinger, J Vaughan, T Pawelzik, H Schalk, DJ McFarland, N Birbaumer, and JR Wolpaw. Patients with als can use sensorimotor rhythms to operate a brain-computer interface. Neurology, 64(10):1775–1777, 2005.

[9] RD Linden, TW Picton, G Hamel, and KB Campbell. Human auditory steady-state evoked potentials during selective attention. Electroencephalogr Clin Neurophysiol, 66(2):145–159, Feb 1987.

[10] OG Lins and TW Picton. Auditory steady-state responses to multiple simultaneous stimuli. ElectroencephalographyandClinicalNeurophysiology, 96:420–432, 1995.

[11] H Luo, Y Wang, D Poeppel, and JZ Simon. Concurrent encoding of frequency and amplitude modulation in human auditory cortex: Meg evidence. J Neurophysiol, 2006.

[12] H Luts. Diagnosis of hearing loss in newborns. PhD thesis, Catholic University of Leuven, 2005.

[13] S Makeig, M Westerfield, TP Jung, S Enghoff, J Townsend, and TJ Courchesne, E.and Sejnowski. Dynamic brain sources of visual evoked responses. Science, 295(5555):690–694, 2002.

[14] P Meinicke, M Kaper, F Hoppe, M Heumann, and H Ritter. Advances in Neural Inf. Proc. Systems (NIPS02), volume 15, chapter Improving Transfer Rates in Brain Computer Interfacing: A Case Study, pages 1107–1114. MIT Press, 2003.

[15] AR Moller. No abstract responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones. Exp Neurol., 45(1):105–117, Oct 1974.

[16] ST Morgan, JC Hansen, and SA Hillyard. Selective attention to stimulus location modulates the steady-state visual evoked potential. Proc Natl Acad Sci U S A., 93(10):4770–4774, May 1996.

[17] MM Muller. Oscil latory cortical activities in the human brain. PhD thesis Universita?t Konstanz, 1997.

[18] MM Muller, TW Picton, P Valdes-Sosa, J Riera, WA Teder-Salejarvi, and SA. Hillyard. Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 hz range. Brain Res Cogn Brain Res., 6(4):249–261, April 1998.

[19] A Osman. An Invitation to Cognitive Science, volume 4, chapter Brainwaves and Mental Processes: Electrical Evidence of Attention, Perception, and Intention, pages 866–915. Scarborough, D and Sternberg, S, 1998.

[20] TW Picton, MS John, A Dimitrijevic, and D Purcell. Human auditory steady-state responses. Int J Audiol, 42:177–219, 2003.

21] B Ross, AT Herdman, and C. Pantev. Stimulus induced desynchronization of human auditory 40-hz steady-state responses. J Neurophysiol, 94(6):4082–4093, 2005.

[22] B Ross, TW Picton, AT Herdman, SA Hillyard, and Pantev C. The effect of attention on the auditory steady-state response. Neurology and Clinical Neurophysiology, 22:1–4, 2004.

[23] G Schalk, F McFarland, T Hinterberger, N Birbaumer, and JR Wolpaw. BCI2000: a general-purpose brain-computer interface (bci) system. IEEE Trans Biomed Eng., 51(6):1034–43, June 2004.

[24] HT Tiitinen, J Sinkkonen, K Reinikainen, K Alho, J Lavikainen, and R. Naatanen. Selective attention enhances the auditory 40-hz transient response in humans. Nature, 364(6432):59–60, July 1993.

[25] JJ Vidal. Toward direct brain-computer communication. Annu. Rev. of Biophys. Bioeng., pages pp. 157–180, 1973.

[26] JR Wolpaw, N Birbaumer, DJ McFarland, G Pfurtscheller, and TM Vaughan. Brain-computer interfaces for communication and control. Clinical Neurophysiology, 113(6):767–791, June 2002.