Clinical Neurophysiology 119 (2008) 842–852

www.elsevier.com/locate/clinph

Non-provocative diagnostics of photosensitivity using visual evoked potentials

Joost Vermeulen [a,1], Stiliyan Kalitzin [b,*], Jaime Parra [c], Erwin Dekker [c], Albert Vossepoel [f], Fernando Lopes da Silva [d,e]

[a] *Quantitative Imaging Group, Department of Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*

[b] *Medical Physics Department, Epilepsy Institutes of The Netherlands Foundation (SEIN), Lorentz de Haas Laboratory, Achterweg 5, 2103 SW Heemstede, The Netherlands*

[c] *Department of Clinical Neurophysiology, Epilepsy Institutes of The Netherlands Foundation (SEIN), Lorentz de Haas Laboratory, Achterweg 5, 2103 SW Heemstede, The Netherlands*

[d] *Center of Neuroscience, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands*

[e] *Instituto Superior Técnico, Lisbon Technical University, Portugal*

[f] *Biomedical Imaging Group Rotterdam, Erasmus MC – University Medical Center Rotterdam, The Netherlands*

Accepted 8 November 2007
Available online 7 February 2008

Abstract

Objective: Photosensitive epilepsy (PSE) is the most common form of reflex epilepsy. Usually, to find out whether a patient is sensitive, he/she is stimulated visually with, e.g. a stroboscopic light stimulus at variable frequency and intensity until a photo paroxysmal response (PPR) occurs. The research described in this work aims to find whether photosensitivity can be detected without provoking a PPR.
Methods: Twenty-two subjects, 15 with known photosensitivity, were stimulated with visual stimuli that did not provoke a PPR. Using an ''evoked response representation'', 18 features were analytically derived from EEG signals. Single- and multi-feature classification paradigms were applied to extract those features that separate best subjects with PSE from controls.
Results: Two variables in the ''evoked response representation'', a frequency term and a goodness of fit term to a particular template, appeared to be best suited to make a prediction about the photosensitivity of a subject.
Conclusions: Evoked responses appear to carry information about potential PSE.
Significance: This result can be useful for screening patients for photosensitivity and it may also help to assess in a quantitative way the effectiveness of medical therapy.

Keywords: Epilepsy; Photosensitivity; Visual evoked potentials; EEG

1. Introduction

Photosensitive epilepsy (PSE) is the most common form of stimulus induced epilepsy (Parra et al., 2005; Kasteleijn-Nolst Trenité, 1989). It is found in 5% of the adult epileptic patients and in 10% of all children with epilepsy (Harding and Jeavons, 1994). It is estimated that it occurs in approximately 1 in 4000 of the population.

Diagnosing photosensitivity (Rubboli et al., 2004) usually involves provoking photo paroxysmal responses (PPRs) with intermittent photic stimulation (IPS). If stimulation is stopped as soon as PPRs appear, there is a small risk of inducing an actual seizure. Some patients, however, might refuse to have photic stimulation due to this small risk. In any case it would be desirable to develop an analytical method that could predict whether the responses to IPS in the sensitive patients would differ from those of normal subjects.

In this context one possibility would be to determine whether some properties of the visual evoked potentials

* Corresponding author. Tel.: +31 235588248.
[1] Tel.: +31 15 27 81416; fax: +31 15 27 86740.

1388-2457/$34.00
doi:10.1016/j.clinph.2007.11.177

would yield useful information in this respect. Research on visual evoked potentials (VEPs) in these cases, however, is rather sparse. Gokcay et al. showed that there were differences in the visual evoked potentials of patients with two forms of epilepsy: childhood epilepsy with occipital paroxysms (CEOP) and symptomatic occipital epilepsy (SOE) (Gokcay et al., 2003). The VEPs of both groups were also found to be different from control subjects. This research was done with IPS at 1 Hz, which in general does not provoke a photo paroxysmal response (PPR).

In recent years, an increase of incidence of photosensitive epilepsy, caused predominantly by increased exposure to video games and television, has prompted researchers to deepen the study of the pathophysiology of human photosensitivity (Wilkins et al., 2004).

In our group (Kalitzin et al., 2002b; Parra et al., 2003), we found that photosensitive and control subjects could be discriminated using the phase clustering index (PCI) of the EEG, or MEG, recorded during intermittent light stimulation at various frequencies, before a PPR was elicited. Using this stimulation paradigm, however, a PPR was ultimately triggered. The question we address in this work is whether it is possible to make a prediction about whether or not the patient is photosensitive using non-provocative stimulation.

Porciatti demonstrated a defect in the contrast gain control in a selected group of patients with idiopathic occipital epilepsy compared to normal controls (Porciatti et al., 2000). A sub-threshold paradigm, i.e. without eliciting PPRs, was used. The stimuli were black and white, and red and green sinusoidal gratings. The essence of their findings is that spatial contrast can play a dominant role in revealing the potential sensitivity of these patients to visual inputs. No attempt for prognostic classification, however, was reported in their paper.

Our approach focuses on quantifying features of the VEPs to IPS. A set of these features is then used to test whether photosensitive epileptic patients and non-sensitive control subjects can be discriminated. Our strategy is to use non-provocative stimulation that is the closest possible to the most common provocative protocol used clinically. Accordingly, we use here responses to simple flashes (and not checker boards) delivered at the generally non-provocative rate of 2 Hz.

Instead of making an ad hoc selection of features that may be expected to yield a good classification, we started with a set of eighteen features, and afterwards we selected, those that provide the best classification using statistical tests.

2. Methods

2.1. Patients

Twenty-two subjects were studied. Information of the patients can be found in Table 1.There were three normal

Table 1
This table gives information about the patients

Subject	Gender	Age	Spont. seizures	Clinical	Syndrome	AED	Photosensitive
1	F	51	Y	MS, GTSC, EM	JME	LTG	Y
2(2×)	F	36	Y	MS, GTSC	JME	VPA	Y
3	F	57	Y	MS, GTCS	IGE, JME	GBP,LEV	Y
4	M	20	Y	CPS	Posttraumatic focal epilepsy	CBZ,LTG	–
5	F	22	–	GTSC, MS	JME	VPA	Y
6	M	15	Y	CPS, GTSC	SGE	OCB	Y
7	M	16	Y	GTSC, MS, FC	GEFS+	VPA	Y
8	M	24	Y	AS, MS	JME	VPA	Y
9	F	48	Y	GTSC	SGE	LTG	Y
10	F	12	–	GTSC	Pure PSE	–	Y
11	F	35	Y	GTSC, MS	JME	–	Y
12	F	14	Y	AS, GTCS	JAE	VPA ESM	–*
13	F	12	Y	MS, GTSC, EM	JME	VPA ESM	Y
14	F	35	Y	CPS	TLE, lesional	LEV CLB	–
15	F	19	Y	GTCS, MS	JME	VPA,LZP	Y
16	M	19	Y	GTCS, MS	JME	–	–*
17	F	16	Y	GTCS, MS	JME	LTG	Y
18	F	7	Y	GTCS	JME	–	Y
19	F	11	Y	AS, MS	Doose syndrome	VPA	Y
20	F	30	–	Control		–	–
21	F	11	–	Control		–	–
22(2×)	F	25	–	Control		–	–

Abbreviations for clinical responses: MS, myoclonic seizure; GTSC, generalized tonic–clonic seizures; EM, eyelid myoclonia; CPS, complex partial seizures; FC, febrile convulsions; AS, absence seizures; S, subclinica; U, uncomfortable sensation. *Abbreviations for AED (Anti Epileptic Drug):* LTG, lamotrigine; VPA, valproic acid; GBP, gabapentine; LEV, levetiracetam; OCB, oxcarbazepine; CBZ, carbamazepine; ESM, ethosuximide; CLB, clobazam; LZP, lorazepam. *Abbreviations for syndromes:* JME, juvenile myoclonic epilepsy; JAE, juvenile absence epilepsy; IGE, idiopathic generalized epilepsy; SGE, symptomatic generalized epilepsy; GEFS+, generalized epilepsy with febrile seizures plus; PSE, photosensitive epilepsy; TLE, temporal lobe epilepsy. The '*' indicates that patients were photosensitive in the past, but not anymore at the time of the recording.

subjects, with no familiar antecedents of epilepsy. Two patients had focal epilepsy (one posttraumatic and one with lesional temporal lobe epilepsy), without evidence of photosensitivity, neither during several EEG recordings, nor on clinical grounds. In addition, two patients with idiopathic generalized epilepsy (one with juvenile absence epilepsy and the other with juvenile myoclonic epilepsy) did show some degree of photosensitivity in previous recordings, but this condition was controlled with medication, such that no evidence of photosensitivity was found during the day of the recording. These patients were considered for the purpose of this study as non-sensitive and are indicated with an '*' in Table 1.

The other 15 patients showed photo paroxysmal discharges during appropriate intermittent photic stimulation. Two patients had symptomatic generalized epilepsy, the others belonged to the group of idiopathic generalized epilepsies, mostly juvenile myoclonic epilepsy.

2.2. Data acquisition

Most data were acquired at SEIN using a LaMont 32 channel amplifier and Harmonie 6.0 software (Stellate Systems, Montreal). The international 10/20 system was used for the electrode positions and the EEG data were sampled at 800 Hz. Some data were obtained from the VUMC (Free University Medical Center in Amsterdam) magneto encephalography (MEG) center. The VUMC data were acquired at different sampling rates, varying from 250 Hz to 625 Hz. To overcome this inequality, all data were downsampled to 250 Hz.

VEP investigation was carried on by means of a commercially available stroboscopic photic stimulator (Grass PS33). The patients were seated on a chair in front of the stroboscopic stimulator that was placed at 30 cm from their eyes. The recordings obtained in the MEG center were performed with a stimulator specifically designed to project images in the shielded MEG environment (Parra et al., 2003). In the MEG center, EEG data were measured in addition to MEG data. Only the EEG data were used for this analysis.

In order to record visual evoked potentials (VEPs) the patients were stimulated with a flashing light at 2 Hz and an analog trigger signal was recorded synchronously with the EEG traces; only the first 333 ms of the evoked responses was used.

2.3. Signal representation

The interesting parts of the EEG recordings for this study are the sequences immediately after the stimulus. Using the trigger timing information the data are cut and arranged into a 3D response "cube": channels (c) × stimulus number (s) × time after stimulus (t), $R(t)_{cs}$. For each channel one can select the response to any individual stimulus; this is called a trial. After this procedure, the best linear fit of each trial is subtracted ("detrending"). Our basic stimulus response model is:

$$R(t)_{cs} = N(t)_{cs} + 1_s a_c S(t+\tau_s) \tag{1}$$

here R, N and S represent, respectively, the measurement, the noise and the "true" response signal. In the above representation, $S(t)$ is the time-evolution factor while the coefficient a_c quantifies the distribution of the response over the EEG derivations and the coefficient τ_s represents possible trial-to-trial jitter in the responses. The latter can be due to imprecise trigger localization or due to jitter caused by fluctuations of the EEG on-going activity. We assume further that the noise N has all the 3 dimensions: channels, stimuli and time. We have also assumed here that the jitter can depend on the stimulus but not on the channels which limits our interpretation to non-propagating patterns.

Notice that for all stimulation trials the same $a_c S(t)$ represents the stationary physiological response. To account for this, the symbol $1_s \equiv [1,1,\ldots,1]$ is introduced.

In short, Eq. (1) assumes that the evoked response of a patient is stationary and that the system resets itself after each response.

The objective of the following pre-processing phase was to determine or estimate the individual factors in (1). In the pre-processing phase several techniques were used as described in Sections 2.3.1, 2.3.2, 2.3.3. Different strategies were followed in order to find an optimal representation of the response signal for our classification objective.

2.3.1. Alignment procedure

Our first optional strategy was to find an adequate estimation for the "misalignments" τ_s in (1). To account for possible misalignment a computational scheme was developed (Kalitzin et al., 2002a) based on estimating the correlation functions between the evoked responses. Here an alternative alignment procedure was applied that is also correlation-based but less computationally expensive. The procedure can be found in Supplemental Appendix A. The τ_s found here can be inserted in formula (1) to compensate if required for the initial misalignment.

2.3.2. Taking the mean over trials

To further reduce the complexity of our data and according to the assumption that the terms $a_c S(t)$ are the actual evoked responses per channel, we average (1) over all trials to obtain

$$\begin{aligned} P_c(t) &\equiv \langle R(t)_{cs}\rangle_s = \langle N(t)_{cs}\rangle_s + \langle 1_s S(t+\tau_s)\rangle_s a_c \\ &= a_c \overline{S}(t) + \eta_c(t) \end{aligned} \tag{2}$$

Here, the notation $\langle\,\rangle_a$ is used to indicate averaging. In the last formula the left-hand side defines the channel-dependent triggered response, the average $\eta_c(t) \equiv \langle N(t)_{cs}\rangle_s$ is the residual noise component, and the definition $\langle 1_s S(t+\tau_s)\rangle_s \equiv \overline{S}(t)$ has been applied. In the case when no jitter is assumed ($\tau_s \equiv 0$) we have simply $\overline{S}(t) = S(t)$. In the sequel we use the notation $S(t)$ both for the aligned

and the non-aligned estimations of the response's shape function.

In addition to the averaged response (2), the response statistical properties can be quantified by the phase clustering index (PCI) (Kalitzin et al., 2002b; Parra et al., 2003). Explicitly,

$$PCI_c(\omega) \equiv \frac{|\langle F(\omega)_{cs}\rangle_s|}{\langle |F(\omega)_{cs}|\rangle_s} \tag{3}$$

where:

$$F(\omega)_{cs} = FT\{R(t)_{cs}\}$$

is the Fourier transform of the signal. We consider here the overall PCI as in formula 3 with the following channel-averaging added:

$$PCI(\omega) \equiv \frac{\langle |\langle F(\omega)_{cs}\rangle_s|\rangle_c}{\langle |F(\omega)_{cs}|\rangle_{sc}} \tag{4}$$

2.3.3. Principal component analysis between the channel responses

Since the EEG is measured at several positions on the head, one expects the amplitude to differ between channels resulting in factors a_c different for each channel. To separate the common response shape $S(t)$ from the weight factors a_c in formula (2) we can use the singular value decomposition technique that minimizes the total quadratic variation of the channel-specific noise $\eta_c(t)$. This technique uses the covariance matrix

$$C = \mathrm{cov}_t(P_c(t))$$

Selecting the eigenvector V_c corresponding to its largest eigenvalue we define the principal component (PC)

$$S(t) = \sum_c P_c(t) \cdot V_c \tag{5}$$

The remaining components, orthogonal to (5), may then be interpreted as "noise". From formula (2) it can be seen that the factor a_c can be calculated as follows:

$$a_c = \frac{\sum_t P_c(t) \cdot S(t)}{\sum_t S(t) \cdot S(t)} \tag{6}$$

Finally, to give an estimate of the three-dimensional noise, we write (1) in the form

$$\begin{aligned} N(t)_{cs} &= R(t)_{cs} - 1_s a_c S(t+\tau_s) \\ &= R(t)_{cs} - 1_s \frac{\sum_{t'} P_c(t') \cdot S(t')}{\sum_{t'} S(t') \cdot S(t')} S(t+\tau_s) \end{aligned} \tag{7}$$

In the above formula all quantities are known from the steps described in Sections 2.3.1,2.3.2,2.3.3. We assume in the sequel the arbitrary normalization $\sum_t S(t) \cdot S(t) \equiv 1$. Given the Pythagorean theorem, this means that all spatial amplitude variation can be found in the channel weights a.

In Fig. 1 the effect of what has been described in Section 2.3.3 is demonstrated. The image shows two artificial "signals" (blue lines), their mean (red line) and the representation obtained with PCA (green line). From the red line it appears as if nothing was measured. The green line shows that with the PCA weights, the shape of the signal is maintained.

2.3.4. Representation approaches for response classification

In this work four different approaches for the quantification of the evoked potential (EP) data are pursued. They consist of combinations of what has been described in Sections 2.3.1, 2.3.2, 2.3.3. The text in bold is used to designate a specific representation for future reference.

1. Basic representation. Only the procedure of Section 2.3.2 is performed. For each occipital channel (O1 and O2) the mean is taken over all responses. This yields 2 evoked responses per patient, one for each occipital channel. The subsequent features (Section 2.4) are calculated separately and averaged at the end.

2. PCA representation. Like the first method, but now the resulting signals of the different channels are processed according to Section 2.3.3. This representation yields one evoked response (the principal component) per patient.

3. Aligned basic representation. The same as the basic representation but after applying the alignment procedure described in Section 2.3.1.

4. Aligned PCA representation. The fourth and most computational expensive method involves all three pre-processing steps 2.3.1, 2.3.2, 2.3.3.

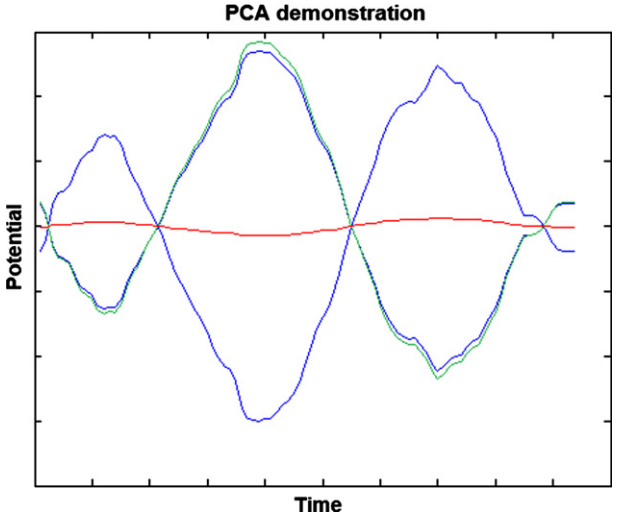

Fig. 1. Demonstration of the effect of the PCA procedure. There are two artificial "signals" (blue lines), their mean (red line) and the representation obtained with PCA (green line). In this particular case, it can be seen that two signals with the same shape, but with only a different scaling, are considered to be almost the same. The red line shows the mean of the two signals. According to the red line, nothing or very little was measured. Maintaining the shape is beneficial when for example one measures the same signal at opposite sides of a dipole. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Quantification of the EP

After obtaining a representation of the EP, we proceed with quantifying the features of the signal. In the previous chapter we have reduced the response signal complexity, according to the chosen representation from Section 2.3.4, to several lower dimensional objects. For representations 2 and 4 for example, we are left with the response function (5), weights (6), jitter delays τ_s and the quantities PCI(ω) defined in (4). In representations 1 and 3 the responses are kept separate for the individual channels but the processing is the same. For these two representations, the features per channel are averaged.

To compare different responses, additional data reduction is useful. In Supplemental Appendix C an explanation of features that are used is given. To simplify the notations, the proposed features are denoted by: $(f.n) = \{\text{feature } n$, "$n$" stands for the number of the feature$\}$.

Features were derived from the representation described in Section 2.3.4, an envelope function and from a response model.

To obtain the envelope function, we first compute the analytic signal.

$$H = S(t) + iHilbert(S(t))$$

The absolute of the analytic signal is the envelope function.

$$A(t) \equiv |H(t)| \tag{8}$$

This function has no zero crossings and is therefore well suited for computing features such as moments.

The response model is computed by trying to minimize the distance between the actual response from Section 2.3.4 and a model function by means of which it is attempted to accurately describe the response.

The response model, k, is:

$$k(t) = e^{\lambda \cdot t} \cdot \sin(2 \cdot \pi \cdot v \cdot t) \tag{9}$$

where v represents the frequency and λ is a measure for damping. Note that the frequency v has units Hertz because of the 2π factor in $k(t)$.

We assume that shape is more important than the overall scale factor, hence we quantify the goodness of fit by the distance function D:

$$D(\lambda, v) = \sum (S(t) - SF \cdot k(t))^2 \tag{10}$$

Here,

$$SF = \frac{S(t) \bullet k(t)}{k(t) \bullet k(t)}$$

and the dot denotes a scalar product. A direct parameter search method is used to minimize D, i.e. a range of values for λ and v are tested. v values were tested between 0.25 and 100 Hz in steps of 0.25 Hz; λ values were tested between -0.1 and 0.1 in steps of 0.001. Fig. 2 illustrates an example of an actual response and its model.

Note that this can be iterated several levels deep. One can subtract $k(t)$ from the response and repeat the procedure to obtain the parameters of the second order fit. Table 2 shows a complete list of the features that are used in the analysis.

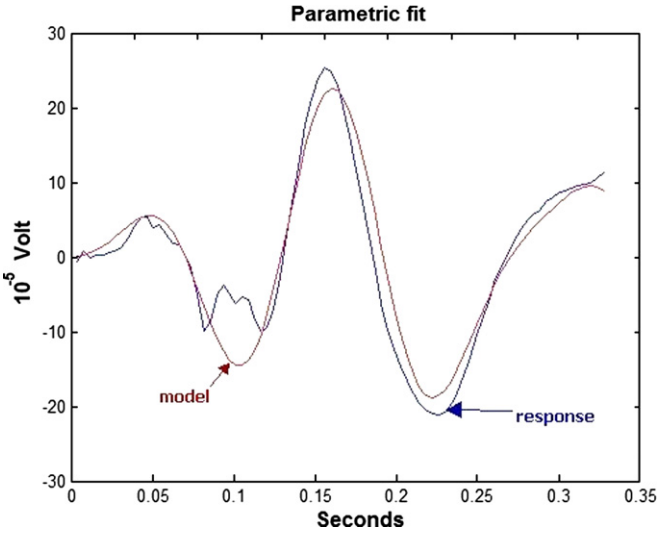

Fig. 2. Illustration of a response and the corresponding response function. The blue line shows the measured response. The red line shows the result of the parametrical fit. In this case a 2nd order fit was applied. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Because of the uncertainties intrinsic to the source localization inverse problem, as well as questions concerning parameters such as unequal contact impedances, we have not attempted to extract any information related to the spatial distribution of the response. Therefore we have ignored the information contained in the coefficients a_c from formula (6).

2.5. Classification

2.5.1. Introduction

In this section we explain how the above quantities are used to classify signals of different conditions. In this work the statistical power of the VEP features to discriminate photosensitive from non-sensitive subjects is analyzed.

2.5.2. Kolmogorov–Smirnov test

The Kolmogorov–Smirnov test (Massey, 1951) is used to compare the distributions of feature values of sensitive and non-sensitive subjects. This particular test has the advantage that a priori no assumptions need to be made about the distributions.

2.5.3. One-dimensional classification

To further investigate the features, the discriminatory information of features is tested directly. This was done using simple one-dimensional threshold classifiers where the values above a certain threshold are considered to belong to one class and those below the threshold to the other. The classifier selects an optimal threshold for each

Table 2
A list of the 18 features that are used in the analysis

f.1	$\omega_{average}$	Mean frequency
f.2	$\bar{t}$	Average time of the response
f.3	σ^2	Variance of the signal
f.4	$\delta\tau$	Misalignment of the response
f.5	Σ	Skewness
f.6	K	Kurtosis
f.7	PCI_0	PCI of the driving frequency
f.8	PCI_m	Maximal phase clustering index
f.9	ω_m	Frequency corresponding to f.8
f.10	$mPCI$	Mean PCI
f.11	$m\omega$	PCI weighted mean frequency
f.12	$rPCI$	Relative phase clustering index
f.13	λ	First order damping factor of the response model
f.14	λ_2	Second order damping factor of the response model
f.15	ν	First order frequency of the response model
f.16	ν_2	Second order frequency of the response model
f.17	D	Distance between the first order response model and the evoked response
f.18	D_2	Distance between the second order response model and the evoked response

feature separately so that the number of misclassified subjects is minimal.

2.5.4. Feature combinations: voting scheme

Using combinations of features may yield better classification results. One simple method to combine the discrimination power of more than one quantity is to perform independently a classification according to each ''member'' and then to postulate certain voting rules for a final decision.

If V_{ia} is the vote of feature i on patient a and W_i is a weight of feature i, then for C_a, the probable classification of patient a we have:

$$\sum_i (W_i \cdot V_{ia}) + A = C_a$$

Here V_{ia} contains the votes (0 or 1) of each feature on each patient based on the optimal threshold classification described in Section 2.5.3 and A is a constant representing possible bias in the classification.

The weight coefficients W_i and the bias term A are determined to minimize the overall classification error, as explained in Supplemental Appendix B.

With W_i and A determined, C_a can be calculated. Finally, we find the common threshold for the class identifiers C_a that minimizes the classification error.

2.5.5. Feature combinations: linear classifier

The results with a linear classifier are also used in this work. If, as in the previous paragraph, W_i is some weight of feature i and Q_{ia} is the actual value of feature i on patient a, then we postulate

$$\sum_i (W_i \cdot Q_{ia}) + A = C_a$$

To determine the optimal weights W_i and bias A, we refer again to Supplemental Appendix B.

With W_i and A determined, C_a can be calculated.

In this case also a threshold is set for C_a to minimize the classification error.

Although the voting scheme and linear classifier formulas look the same, they are different with respect to how the features are interpreted. The voting scheme translates the values of the features to a yes–no (01) vote per feature. In case of the linear classifier the values of the features are used to build a feature space in which it is attempted to separate the photosensitive from the control subjects.

2.6. Statistical significance of the classification

2.6.1. Determining a significance threshold

For any classification approach, the significance level is determined by the probability that the same results can be achieved by a ''random guess'' procedure. We require in this work statistical significance that limits the latter probability to at most 5%.

To make sure that the chance is <5%, that the results are as good as random, 10,000 classification attempts on random data were done with a classifier. Since 10,000 classification attempts are done, one can assume that any result the error of which ranks among the best 500 results, gives <5% chance to be random. Therefore the system is allowed to make as many errors as the 500th best attempt. This gives a distinct maximum allowed error. This maximum allowed error is called *error*(max).

2.6.2. Procedure to find good feature combinations

To find the good feature sets, all possible 2D, 3D, 4D and 5D combinations of the 18 (Table 2) features are tested. This is done with both the voting scheme and the linear classifiers. From these feature sets a selection is made based on the significance from Section 2.6.1.

First, the size of the feature space is determined by looking at the *error*(max) from Section 2.6.1 and the error of the best feature sets for each feature space size. The error

of this best set needs to be as low as possible. In addition, the results need to be significant.

The training and testing here and in Section 2.6.1 are done on the same sets. This is no problem as this test serves only to determine the size of the feature space and to remove feature sets that do not perform in accordance with the minimum level that is set by *error*(max).

Once the proper dimensionality of the feature set is determined, all combinations that have an error $<$*error* (max) are used for further investigation.

The results of this procedure are feature sets of proper dimensionality that can separate the two classes with a certain degree of significance.

2.7. Cross-validation tests

To find which feature sets yield the best classification results, two tests are applied. First a leave-one-out cross-validation test is done. Leave-one-out cross-validation involves using a single observation from the data set as the test data, and the remaining observations as the training data. This is repeated until each observation in the sample is used once as test data. The errors of all repetitions are averaged.

Additionally, a hold-out validation test is done. In the hold-out test a set of measurements is randomly split in a training (70%) and a testing (30%) part following prior probabilities. The hold-out cross-validation procedure is repeated 50 times and the results are averaged.

Only those feature combinations that have significant classification performance and of proper dimensionality (see Sections 2.6.1 and 2.6.2) are used in this test.

2.8. Inspection of relevant features

The best sets from the cross-validation tests are inspected taking into account the occurrence of specific features and feature combinations. The feature sets that had a cross-validation error $<$25% are further considered. The percentage (25%) is chosen arbitrarily.

Additionally the best features are studied using ROC (Receiver Operating Characteristic) curves. ROC curves are used to quantify the performance of a classifier in terms of specificity and sensitivity. The last are determined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{and} \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

where:

FP: number of False Positive classifications
TP: number of True Positive classifications
FN: number of False Negative classifications
TN: number of True Negative classifications

Notice that TP + FN represents all subjects that are actually positive (photosensitive) and the TN + FP represents all subjects that are actually negative (not photosensitive).

2.9. Robustness check

To get an impression about the consistency of our results, multiple measurements of subjects are used. Of two subjects, one of which is photosensitive, there were two data sets. These subjects will be referred to as subjects A and B. From each of these two subjects, the measured evoked responses are inspected visually. Additionally, it is verified that both measurements appear near each other in the feature space.

3. Results

3.1. Kolmogorov–Smirnov test

In Supplemental Figures S1a–d the probability is given, for each feature (*f.n*), that the distributions of sensitive and non-sensitive subjects are the same. This is done for the four different representations.

Comparing Supplemental Figures S1a and S1b with Supplemental Figures S1c and S1d shows that the alignment procedure is not very beneficial. This suggests that the alignment destroys discriminatory information and/or that the misalignment may contain discriminatory information.

The best results can be seen in Supplemental Figure S1a. In this case the basic representation is used.

The figures show clearly that the distributions of feature *f*.15 (*v*) for the two groups of subjects are very different. This is also the case for features *f*.8, *f*.11, *f*.17 and *f*.18.

From this point onwards we consider only the EP representations where alignment is not involved.

3.2. One-dimensional classification

In 1D classification the simple threshold from Section 2.5.3 is used. The error for *f*.15 is 13.6% (=3/22, cf. Supplemental Figure S2a). This is when the basic representation is used. The result of the method that involves PCA appears a little less favorable. In this case, *f*.8, *f*.15 and *f*.17 were the best features. Each of them had an error of ~18% (=4/22 cf. Supplemental Figure S2b). The Kolmogorov–Smirnov test also gave some indication of this result. Notice that all errors are fractions of 22, i.e. 1/22, 2/22, etc.

3.3. Procedure to find good feature combinations

This procedure is described above in Sections 2.6.1 and 2.6.2.

The results in Supplemental Tables S1–S4 were obtained with the voting scheme and the linear classifier. Supplemental Table S1 shows the results of the basic representation with the voting scheme. Supplemental Table S2 shows the results of the basic representation with the linear classifier.

Supplemental Tables S3 and S4 show equivalent data for the PCA representation. These tables only show the best performing feature sets in case that no training/testing split is done. As mentioned in Section 2.6.2, the tables serve only to determine appropriate dimensionality.

The '*' in the tables indicate that more than one feature combination with similar performance was found. In Supplemental Tables S1 and S4 the best results are obtained with 3 or more features; the results do not improve when including more features. This means that 3 features can contain the most relevant information. In Supplemental Table S3 the best results are achieved with S2 or more features. Supplemental Table S2 shows that the results with the linear classifier on the basic representation do not improve with more than two features. As expected when training and testing are done on the same set, in all Tables (Supplemental Tables S1–S4) more features yield better results.

It can be seen that the use of the basic representation (Supplemental Tables S1 and S2) yields better results than the PCA representation (Supplemental Tables S3 and S4). Previous tests also demonstrated better performance of the basic representation. Therefore the PCA representation is no longer considered.

The results also show that the linear classifier clearly outperforms the voting scheme classifier.

Like earlier using the Kolmogorov–Smirnov test, *f*.15, but also *f*.18, appears to be important. This is indicated by the fact that they show up in the best performing feature combinations. This can also be concluded from the weights they have in combinations in case the voting scheme classifier is used. The weights from the linear classifier are not

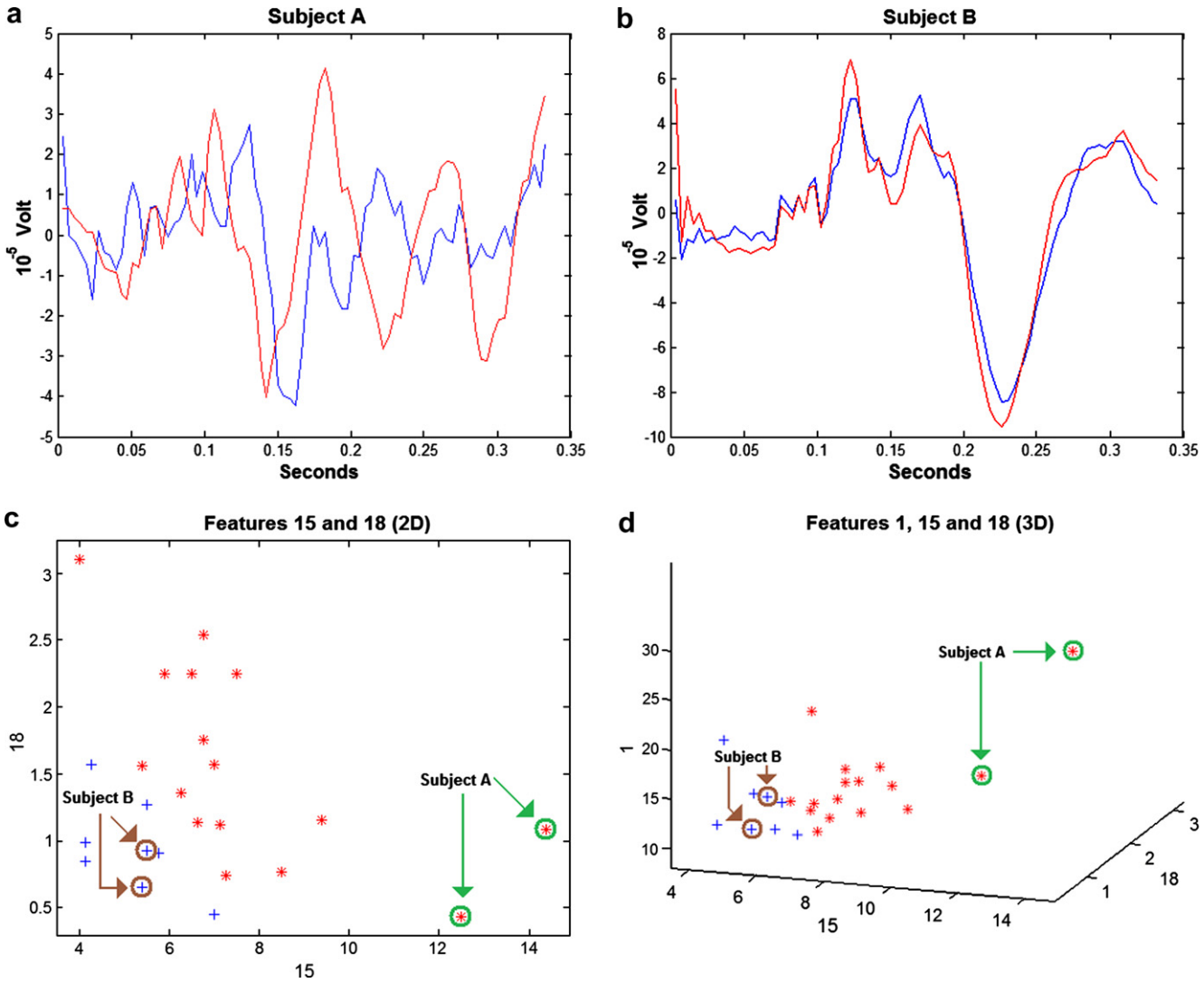

Fig. 3. The EPs of subjects A and B are shown in (a) and (b). The blue and red line indicate the result of two distinct measurements. The two measurements of each subject look very similar to the human eye. It can be seen in the 2D and 3D feature spaces of (c) and (d) that the data sets of the same subject are positioned near each other. Blue crosses represent control subjects, red asterisks represent photosensitive patients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

considered, as they hold scale information in addition to weight information.

Given the information from above it seems reasonable and safe to focus on finding good 3D feature combinations using the linear classifier on the basic representation.

3.4. Cross-validation

Cross-validation tests are done according to Section 2.7. Although $\frac{16\cdot17\cdot18}{6} = 816$ different unique 3D combinations can be made with 18 features, only those feature combinations that have significant classification performance are used (see Section 2.6). With the linear classifier used on the basic representation there were 206 3D sets with significant classification performance. As mentioned in Section 3.3, the leave-one-out and hold-out tests are done using the linear classifier on the basic representation.

3.4.1. Leave-one-out cross-validation

Supplemental Figure S3a shows the average classification error with the leave-one-out cross-validation test.

The three best sets had no errors. The corresponding feature sets are [18 15 1], [18 15 7] and [18 15 14]. Supplemental Table S5 shows the best 10 sets and their errors.

3.4.2. Hold-out validation

Supplemental Figure S3b shows the average classification errors and their standard deviations of the hold-out validation test.

The best set here is [18 16 15], it had an average error ~5%. Supplemental Table S6 shows the best 10 sets.

Investigation of the best feature sets has shown that they all rely on features 15 (v) and 18 (second order D). Note that other tests already indicated the usefulness of $f.15$ and $f.18$.

3.5. Most relevant features

As described in Section 2.8, we focus here on the occurrence of different features, in those feature sets that produce an error smaller than 25% in the cross-validation tests.

3.5.1. Best feature sets of the leave-one-out test

89 feature combinations produced an error <25%. In Supplemental Figure S4a the error fractions of the best feature combinations are shown. In these 89 best combinations feature $f.15$ is used 63 times (see Supplemental Figure S4b). Other frequently occurring features are $f.8$, $f.17$ and $f.18$. They are used 25, 27 and 25 times, respectively.

3.5.2. Best feature sets of the hold-out out test

Seventy-six feature combinations produced an error <25%. In Supplemental Figure S5a the error fractions of the best feature combinations are shown. The results are similar to the results of Section 3.5.1. In these 76 best combinations feature $f.15$ is used 59 times (see Supplemental Figure S5b). Other frequently occurring features are again $f.8$, $f.17$ and $f.18$. They are used 23, 24 and 23 times, respectively.

Supplemental Figures S6a and S6b show graphs of the feature combinations that performed well according to our analysis (see, e.g. Supplemental Table S5).

3.5.3. Sensitivity and specificity

ROC curves of $f.15$, $f.18$ and the combination of $f.15$–$f.18$ are shown in Supplemental Figure S7. The features $f.15$ and $f.18$ are combined using the weights of the linear classifier.

$$f.(15./18) = w(f.15)\cdot f.15 + w(f.18)\cdot f.18$$

where $w(f.15) \approx 0.29$ and $w(f.18) \approx 0.71$. This information cannot be used to directly draw conclusion about the relative importance of features. To do that scale information must be incorporated. It can be seen in, e.g. Supplemental Figure S6 that $f.15$ and $f.18$ have different scales.

As expected, the best sensitivity/specificity combination is obtained when $f.15$ and $f.18$ are combined.

3.6. Robustness check

This check, described in 2.9, gives an impression of the consistency of the measurements and more importantly the position of a subject in the feature space. Only the data of the basic representation are taken into account here.

Figs. 3a and b show the evoked responses of subjects A and B. The blue and red line indicate the result of two distinct measurements. Especially the two measurements of subject B look very similar to the human eye, although this is less obvious for subject A. More relevant though are the positions of the two measurements of these two subjects in the relevant feature space. It can be seen in the 2D and 3D feature spaces of Figs. 3c and d that the data sets of the same subject are positioned not far from each other, especially in the case of B.

4. Discussion

The results of this study indicate that it is possible, using non-provocative visual stimulation, to predict from the visual evoked potentials whether or not a subject is photosensitive.

Different approaches to represent the evoked response were tested. The good performance of the basic and PCA representations suggests that the alignment destroys discriminatory information. The basic representation gives the best results.

Combinations of features were found that could be used to separate the sensitive and non-sensitive subjects without provoking a seizure. The most important feature of the response appears to be $f.15$ (v, the first order frequency from the response model).

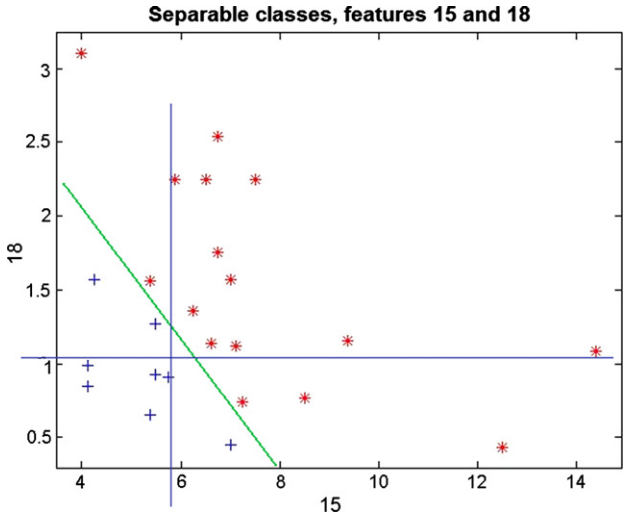

Fig. 4. This illustration shows the usefulness of features 15 and 18. A green diagonal line is drawn to demonstrate that the two classes can be separated perfectly. Additionally two blue lines are drawn to demonstrate the thresholds of the voting scheme. The data from the basic representation are used here. The data are not scaled. Blue crosses represent control subjects, red asterisks represent photosensitive patients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

These results, although based on a relatively small sample, show that it is possible to discriminate between photosensitive and non-photosensitive subjects. Nevertheless additional data would be necessary to make meaningful learning curves.

The main result was obtained using a weighted voting system what is very basic, since no relation between classifiers is assumed. According to this approach different parameters are taken individually and a classification is obtained based only on that information.

Additionally we tested a linear classifier. The scatter plot in Supplemental Figure S6a clearly shows that a combination of *f*.15 and *f*.18 (Table 2) can be used to separate the two classes with a linear classifier.

Although additional features may improve the performance, it is clear that the two features *f*.15 and *f*.18 are mainly responsible for good classification. Of these two, *f*.15 appears to be more important. The interpretation is that the discrimination of photosensitive subjects has a strong frequency dependency. Indeed the importance of *f*.15 in this respect is in accordance with the hypothesis that high frequencies are associated with the transition between normal and seizure activity (Parra et al., 2003; Jirsch et al., 2006; Le Van Quyen et al., 2006).

The interpretation of feature *f*.18 is less straightforward since it represents the second order distance between the response model and the measured evoked response. The value of *f*.18 tends to be lower for control subjects. From this we can conclude that response model function describes more closely the responses of control subjects than responses of photosensitive subjects.

As mentioned before, appropriate linear combination of features *f*.15 and *f*.18 can separate EPs from sensitive from those from non-sensitive subjects and therefore might be a good candidate for practical, non-provocative diagnostics. This is illustrated in Fig. 4 and is also shown in Supplemental Table S2.

In conclusion the method described here can be applied in clinical practice in order to test photosensitivity in patients without using relatively high frequencies of intermittent photic stimulation, and thus avoiding discomfort for the patient and increasing the safety of the stimulation protocol.

Acknowledgements

We thank the Epilepsy Institutes of The Netherlands Foundation (SEIN) for providing the EEG facilities and clinical data. The assistance and helpful comments of Demetrios Velis and Wouter Blanes were much appreciated.

We are thankful for the help and assistance of the colleagues at the Quantitative Imaging group of the TU Delft. The helpful comments, suggestions and feedback during progress meetings and research discussions were of great benefit to this project.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.clinph.2007.11.177.

References

Gokcay A, Celebisoy N, Gokcay F, Ekmekci O, Ulku A. Visual evoked potentials in children with occipital epilepsies. Brain Dev 2003;25:268–71.

Jirsch JD, Urrestarazu E, LeVan P, Olivier A, Dubeau F, Gotman J. High-frequency oscillations during human focal seizures. Brain 2006;129(6):1593–608.

Harding GFA, Jeavons PM. Photosensitive epilepsy. London: Mac Keith Press; 1994.

Kalitzin S, Parra J, Velis D, Lopes da Silva F. Enhancement of phase clustering in the EEG/MEG gamma frequency band anticipates transitions to paroxismal epileptiform activity in epileptic patients with known visual sensitivity. IEEE Trans Biomed Eng 2002a;49(11):1279–86.

Kalitzin S, Zbijewski W, Parra J, Velis D, Manshanden L, Lopes da Silva F. Correlation-based alignment of multichannel signals and application to paroxysmal events. IEEE Trans Biomed Eng 2002b;49(9):1068–70.

Kasteleijn-Nolst Trenité DGA. Photosensitivity in epilepsy. Electrophysiological and clinical correlates. Acta Neurol Scand 1989;80(125):1–149.

Massey Jr FJ. The Kolmogorov–Smirnov test of goodness of fit. J Am Stat Ass 1951;46(253):68–78 (The Matlab implementation was used for the Kolmogorov–Smirnov test).

Parra J, Kalitzin S, Iriarte J, Blanes W, Velis D, Lopes da Silva F. Gamma-band phase clustering and photosensitivity: is there an underlying mechanism common to photosensitive epilepsy and visual perception? Brain 2003;126:1164–72.

Parra J, Kalitzin S, Lopes da Silva F. Photosensitivity and visually induced seizures. Curr Opin Neurol 2005;18(2):155–9.

Porciatti V, Bonanni P, Fiorentini A, Guerrini R. Lack of cortical contrast gain control in human photosensitive epilepsy. Nat Neurosci 2000;3(3):259–63.

Le Van Quyen M, Khalilov I, Ben-Ari Y. The dark side of high-frequency oscillations in the developing brain. Trends Neurosci 2006;29(7):419–27.

Rubboli G, Parra J, Seri S, Takahashi T, Thomas P. EEG diagnostic procedures and special investigations in the assessment of photosensitivity. Epilepsia 2004;45(s1):35–9.

Wilkins A, Bonanni P, Porciatti V, Guerrini R. Physiology of human photosensitivity. Epilepsia 2004;45(1):7–13.