Clinical Neurophysiology 118 (2007) 2519–2524

www.elsevier.com/locate/clinph

Test–retest reliability of resting EEG spectra validates a statistical signature of persons

Markus Näpflin [a,b], Marc Wildi [b], Johannes Sarnthein [a,c,*]

[a] Neurochirurgie, Universitätsspital Zürich, CH-8091 Zürich, Switzerland
[b] Zürcher Hochschule Winterthur, 8401 Winterthur, Switzerland
[c] Zürich Center for Integrative Human Physiology, Universität Zürich, Switzerland

Accepted 14 July 2007
Available online 24 September 2007

Abstract

Objective: When EEG is recorded in humans, the question arises whether the resting EEG remains stable. We compared the inter-individual variation in spectral observables to the intra-individual stability over more than a year.
Methods: We recorded resting EEG in 55 healthy adults with eyes closed. In 20 persons EEG was recorded in a second session with retest intervals 12–40 months. For electrodes AFz, Cz and Pz α peak frequency and α peak height were transformed into Z-scores. We compared the curve shape of power spectra by first aligning α peaks to 10 Hz and then regressing spectra pairwise onto each other to calculate a t-value. The t-value and differences of Z-scores for all pairs of sessions were entered into a generalized linear model (GLM) where binary output represents the recognition probability. The results were cross-validated by out-of-sample testing.
Results: Of the 40 sessions, 35 were correctly matched. The shape of power spectra contributed most to recognition. Out of all 2960 pairwise comparisons 99.5% were correct, with sensitivity 88% and specificity 99.5%.
Conclusions: Our statistical apparatus allows to identify those spectral EEG observables which qualify as statistical signature of a person.
Significance: The effect of external factors on EEG observables can be contrasted against their normal variability over time.

Keywords: EEG; α Peak; Oscillations; Regression; Generalized linear model; GLM; Cross-validation

1. Introduction

The intra-individual stability of the resting EEG has been addressed by a number of publications. In spectral analysis several parameters show high reproducibility over a period of months or years (Stassen, 1980; Gasser et al., 1985; Salinsky et al., 1991; Dustman et al., 1999; Kondacs and Szabo, 1999; Poulos et al., 2002; Maltez et al., 2004; Vuga et al., 2006). High stability of EEG spectra can also be found across multiple sleep recordings in individuals by clustering techniques (Buckelmüller et al., 2006). The clinical interest in intra-individual EEG stability stems from the study of therapeutic interventions and the follow-up of individual patients (Sarnthein et al., 2006). But when comparing an individual's EEG to a previous record, the intra-individual stability has to be set in relation with the inter-individual variation of EEG spectra.

As a new approach in EEG research, we have established a generalized linear model to recognize individual persons by their EEG spectra. Resting EEG power spectra recorded with eyes closed are usually dominated by a peak in the α frequency range 8–12 Hz (Niedermeyer and Lopes da Silva, 1999; Nunez et al., 2001). To characterize intra-individual stability of EEG spectra over recording sessions, we investigated differences in the absolute height of the peak, frequency of the peak and shape of the power spectra. We were interested whether this parsimonious set of

* Corresponding author. Address: Neurochirurgie, Universitätsspital Zürich, CH-8091 Zürich, Switzerland. Tel.: +41 44 255 1111.
E-mail address: johannes.sarnthein@usz.ch (J. Sarnthein).

1388-2457/$32.00 © 2007 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2007.07.022

observables together with a generalized linear model would result in satisfactory recognition probabilities and thereby establish a statistical signature of persons on the basis of their EEG.

2. Methods

2.1. Participants

The group of participants consisted of 55 healthy volunteers, (19–79 years, median 51 years, 18 women, 37 men). The study was approved by the Kanton Zürich Ethics Committee. All persons were informed about the aim and the scope of the study and gave written informed consent according to the declaration of Helsinki. All persons were screened for health problems using a detailed health questionnaire. The persons had no current or previous history of relevant physical illness and they were not currently taking drugs or medication known to affect their EEG. The EEG was measured for a second time (retest) in 20 persons (25–76 years, median 58.5 years, 7 women, 13 men). The time interval between two EEG sessions was longer than 1 year (retest interval 12–40 months, median 15 months).

2.2. EEG recording sessions

The persons were seated in a dimly lit room shielded against sound and stray electric fields and were video-monitored. All EEGs were recorded in the morning between 9.00 and 12.00. The recording apparatus was continuously calibrated. The persons refrained from caffeinated beverages before the session to avoid the caffeine-induced θ decrease in the EEG (Landolt et al., 2004). During recording they assumed a comfortable position in a chair and placed their head on a chin-rest. The persons were instructed to close their eyes, to place their fingers on their eyelids, and to relax but to stay alert for 5–10 min. Finally, alertness was checked by asking whether they had stayed alert during recording.

EEG signals were measured with 60 surface electrodes, which were fixed in a cap at the standard positions according to the extended 10–10 system (Montage 11, Easycap, Herrsching, Germany). For the first 54 recording sessions, we used passive Ag/AgCl electrodes (Easycap, Herrsching, Germany). For the remaining 21 sessions we used active electrodes (ActiCap, Brain Products, Gilching, Germany). Electrode CPz served as reference during recording. Impedances were below 20 kΩ in all electrodes processed in further analysis. We used two additional bipolar electrode channels as eye monitors. EEG signals were registered using the SynAmps EEG system (Neuroscan Compumedics, Houston, TX, USA, common mode rejection 100 dB, gain 5000, range 1.1 mV, A/D conversion 17 nV/LSB, sampling rate 250 Hz, band pass filter 0.3–100 Hz, −12 dB/octave) and continuously viewed on PC monitor.

2.3. Data preprocessing and editing

The data were analysed offline in the Matlab (The Matworks, Natick, MA, USA) environment using EEGLAB (http://sccn.ucsd.edu/eeglab (Delorme and Makeig, 2004)) and custom programming on the basis of standard mathematical and signal analysis functions. The scalp EEG was re-referenced to the mean of the signals recorded at the ear lobes (Ag/AgCl electrodes) or at the mastoids (ActiCap). We confirmed alertness of the persons during the recording session by checking for the slowing of the α rhythm or slow rolling eye movements. The data were inspected in 5 s epochs and large muscle or eye movement artefacts were removed. For editing purposes, muscle artefacts were considered significant if the underlying EEG rhythms were not clearly seen. The EEG was decomposed into independent components using blind separation (independent component analysis, ICA). After removal of components containing eye movement or muscle artefacts, the signal was reconstructed. All records were edited by the same encephalographer to increase reliability.

2.4. Spectral parameter estimation

The spectral analysis was performed with the multitaper method, which offers optimal spectral smoothing, i.e. it allows to trade resolution in the frequency domain for reduced variance (Percival and Walden, 1993). Power spectral density was estimated with a window length of 5 s, which will be referred to as sweep in the following. Over all sessions, the minimum, median and maximum number of sweeps were 40, 58 and 120, respectively. We used fast Fourier transform length 10 s, bandwidth parameter $nw = 2$ and $k = 3$ tapers. To approximate a normal distribution, power estimates were log-transformed. The maximum in the 8–12 Hz α frequency range was taken as the dominant frequency (Fig. 1). Peak frequency and peak height were transformed into Z-scores on the basis of the first session of all 55 persons (Fig. 2).

2.5. Similarity in shape of power spectra

The spectra of power spectral density were first aligned to a common maximum (Fig. 3a). In this alignment the frequency axis was compressed or expanded linearly such that the origin was kept fixed and the dominant frequency came out to be 10 Hz for each spectrum. This multiplicative alignment ensures that the ratio of harmonics in the spectrum remains unchanged.

We then performed a pairwise comparison between the aligned power spectra k and j, where index j runs over all EEG recording sessions. The spectra are written as $x_i^{(k)}$ and $x_i^{(m)}$, where i denotes one of n frequency points. The spectra were regressed pairwise onto each other using the model

$$x_i^{(k)} = \beta_1 + \beta_2 \cdot x_i^{(m)} + e_i, \qquad i = 1, \ldots, n$$

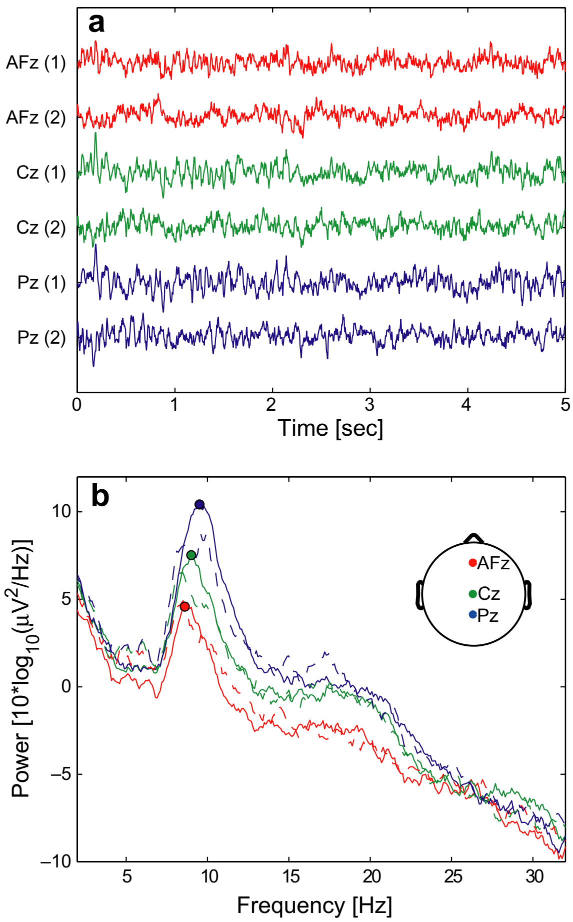

Fig. 1. EEG traces and power spectra. (a) Resting EEG traces with eyes closed are shown for three electrode sites (AFz red, Cz green, Pz blue). The first recording session (1) of person 8 (64 years old) was followed by the second recording session (2) after an interval of 18 months (first sweep in the EEG recording session, offset between traces 50 μV). (b) Power spectra averaged over all sweeps of the first recording session (test, solid lines) and the second recording session (retest, dashed lines).

where e_i denotes random noise. We included the whole frequency range [2…32 Hz] in the analysis (Fig. 3b). The ordinary least-squares estimation was used in the linear regression model. If the shape of two spectra is similar, the slope (β_2) is near to one and the standard error SE of coefficient β_2 is near to zero. The t-value

$$t_{\text{power}} = \frac{\beta_2}{\text{SE}}$$

describes the strength of the relationship between the two spectra. In this way we generate the observables t describing the similarity in shape of two power spectra.

2.6. Generalized linear model

To describe the similarity between two EEG recording sessions, values derived from the spectral parameters were entered in a multiple linear regression model (GLM).

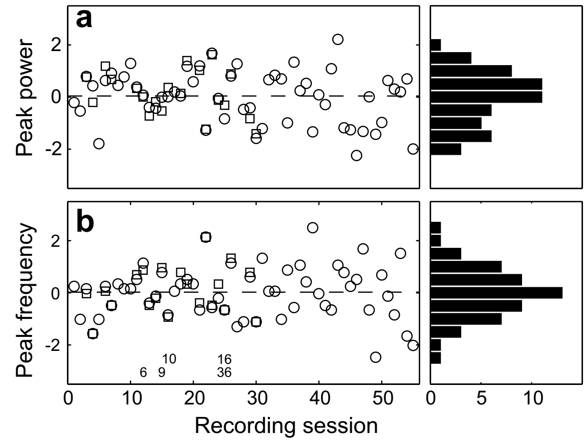

Fig. 2. Distribution of peak height and frequency. (a) The peak power at electrode site Pz was Z-transformed (mean 11.7 ± 7.0 10 * $\log_{10}$ $\mu V^2/Hz$) and plotted for first (○) and second (□) recording sessions. (b) Same as in (a) for peak frequency (mean 9.7 ± 1.1 Hz). The first recording sessions constitute the normal population. Numbers 6, 9, 10, 16 and 36 indicate those sessions, which have not been matched correctly. Peak power and peak frequency are normally distributed (Lilliefors test).

The nine values were derived from the EEG recorded from the three individual electrodes during the two sessions: the differences in Z-scores of peak height and peak frequency describe the similarity of the two peaks and the t-value of the linear regressions assesses the similarity in shape of the power spectra.

Since every observation describes a comparison of two EEG recording sessions, the response variable denotes whether it is an intra- or inter-individual comparison and is therefore binomially distributed. This leads to a GLM with "Logit" as canonical link function for binary response variables:

$$y_i = \frac{e^{\beta \cdot X}}{1 + e^{\beta \cdot X}} + E_i$$

$$\beta \cdot X = \beta_0 + \sum_{j=1}^{9} \beta_j \cdot x_i^{(j)}$$

where $\beta \cdot X$ is a linear combination of coefficient vector β with predictors X.

$x_i^{(j)}$ are elements of predictors X, detailed in Table 1.

y_i is the response variable for all pairwise comparisons i. It describes whether the comparison is intra- or inter-individual ($y_i = 1$, if the two recording sessions are from the same person, $y_i = 0$ otherwise).

E_i are random errors.

Of the persons with two recording sessions we compared the first and second session pairwise with all sessions of all persons (Fig. 4). The comparison with the highest recognition probability $\hat{y}_i$ indicates recognition of a person. In order to test its stability and to avoid overfitting, the GLM was fitted for each person separately under exclusion of the data of this person. This means that comparisons including person k were not used to estimate coefficients which predict the recognition probability for person k.

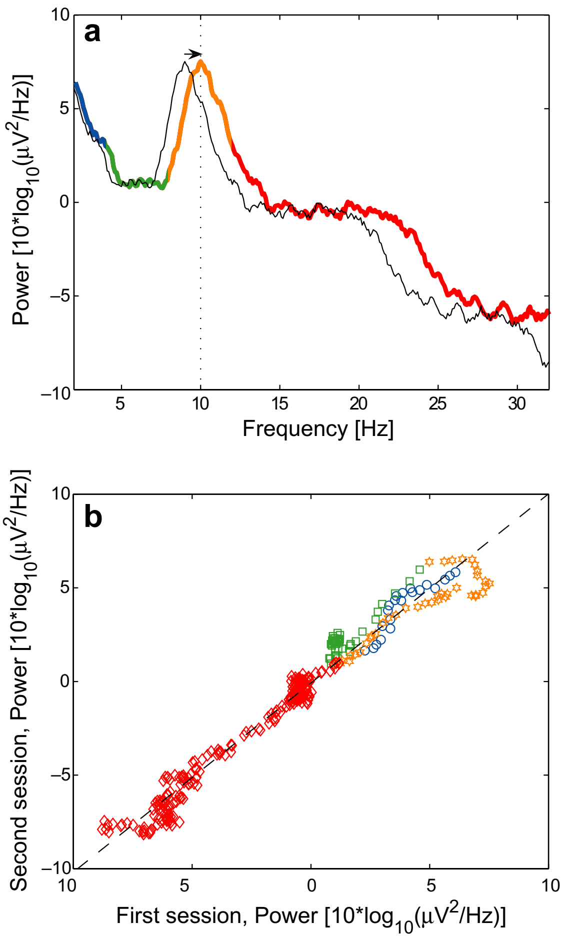

Fig. 3. Similarity of power spectral shape. The spectra were calculated using the data from electrode Cz of person 8 (sessions 8 and 28). (a) When aligning power spectral peaks, the frequency axis is compressed or expanded linearly such that the origin is fixed and all spectra have a common maximum at 10 Hz. (b) After peak alignment, the spectra of first session (8) and second session (28) are regressed onto each other. Each marker represents a frequency point in one of the frequency bands δ (blue circles, 2–4 Hz), θ (green squares, 4–8 Hz), α (orange stars, 8–12 Hz) and β (red diamonds, 12–32 Hz). Frequency points in β span the longest stretch on the regression line (dashed). The slope of the regression line is used to calculate the t-value $t_{\text{power}} = 94$.

Note that there is an asymmetry induced by the GLM-model which implies that comparing person k to person j if person k is the reference is not the same as comparing person j to person k if person j is the reference. This procedure amounts to 74 comparisons for each recording session. The total number is then $40 * 74 = 2960$ comparisons in the whole experimental setting.

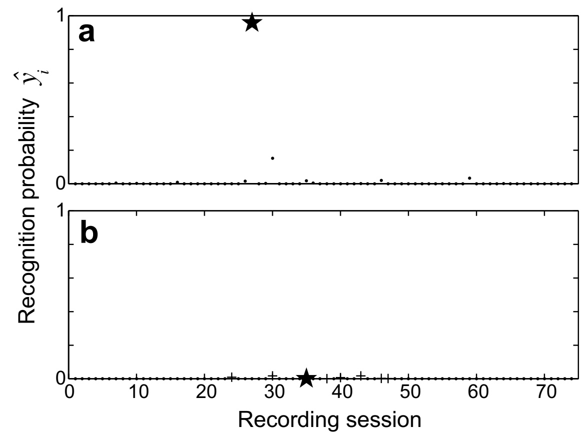

Fig. 4. Person recognition. Recognition probabilities $\hat{y}_i$ are plotted for all comparisons concerning the first recording session of a person (a). For person 8 the highest value occurs for comparison with the second session of person 8 (asterisk), indicating correct recognition. (b) For person 16 (44 years old, test–retest interval 12 months), seven other sessions achieve higher $\hat{y}_i$ (crosses). Thus person 16 is not correctly recognized.

3. Results

The recognition probabilities $\hat{y}_i$ for all pairwise comparisons with the first session of person 8 are shown in Fig. 4a. Among all $\hat{y}_i$ the highest value appears for the comparison between first and second session of person 8 and thus indicates correct recognition. Incorrect recognition is illustrated in Fig. 4b for person 6, where comparisons with other persons achieve higher recognition probabilities.

If we base the performance of the method on the numbers of sessions correctly matched to the correct person, we can estimate 35/40 = 88% with 95% confidence interval [73–96%]. We can also analyse the rate of correct decisions made by our model which can be based on the set of all $40 * 74 = 2960$ pairwise comparisons. Specifically, we assume session k is taken from the 40 retested persons and the model computes scores (probabilities) for each pairwise

Table 1
Input to and output from the GLM

Electrode site	AFz			Cz			Pz		
Observable	t_{power}	Δp	Δf	t_{power}	Δp	Δf	t_{power}	Δp	Δf
j	1	4	7	2	5	8	3	6	9
Number of sessions matched ($n = 40$)	27	4	16	27	7	17	26	3	10

Each electrode site provides the observables t_{power}, Δf and Δp (t-value of spectral shape regression, difference in peak frequency, difference in peak power). These nine observables form the elements $x_i^{(j)}$ of the predictors for the GLM. If each observable is taken individually, the number of sessions matched deteriorates. The best match ($\geqslant$26 persons) is achieved on the basis of the spectral shape of power.

comparison with the 74 remaining sessions. Then person k would be assigned to session i_{k1} in the test-sample if the probability of recognition is maximal. Assume now that session i_{k1} does not correspond to person k. Then we take the second largest probability corresponding to session i_{k2}. If i_{k2} is false again, then we proceed further that way until we get session i_{kn} corresponding to the second measure of person k. All in all we made $n-1$ false decisions out of 74 possible pairwise comparisons. Adding all false comparisons for sessions $1,\ldots,40$ amounts to a total of 14 false recognitions. The recognition rate then amounts to $1-14/2960=99.5\%$ and $35/40=88\%$ as sensitivity and $(2960-35-5-14)/(2960-40)=99.5\%$ as specificity of the test.

The number of sessions correctly matched depended on the choice of EEG processing parameters. As a test, we first varied the length of the EEG data, i.e. the number of EEG sweeps with 5 s length used to calculate the spectra. On the basis of all available data, 35 of the 40 sessions were correctly matched (88%). For reduced data sets of 20, 30 and 40 sweeps, we obtained 27, 28 and 33 correct matches, respectively. The mismatch occurred mainly because the power spectral parameter Δf showed more variability. This suggests that the resting EEG parameters considered here have better validity for a data length of more than 200 seconds. We also analysed all available data without ICA preprocessing and found the number of matches (32 sessions, 80%) to be slightly reduced. Among the other parameters we optimised were: sweep length 5 s; NFFT length 10 s; multi-taper parameters k, nw; frequency range entering the regression (Fig. 3b). In five of the retested persons, passive electrodes were used in the first session and active electrodes in the second session. Nine of the 10 sessions were matched correctly, which indicates that the data of the two recording setups can be directly compared.

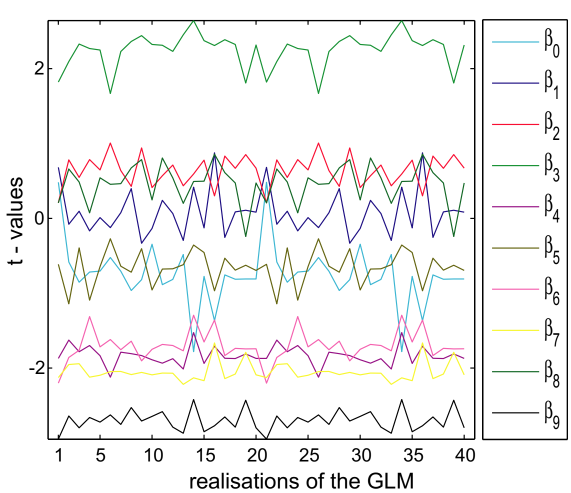

Fig. 5. Stability of the GLM. To cross-validate the recognition probabilities, one GLM is fitted for each session k under exclusion of the data of session k. Among all 40 realizations of the GLM, the t-values show little variation.

The different observables entering the GLM contribute to the recognition probability with varying extent. The marginal effect may be quantified by the number of persons recognized correctly on the basis of each observable alone (Table 1). These recognition rates are below that of the GLM, indicating that the GLM collects and processes information from all three variables efficiently.

The stability of the GLM is related to the explanatory power of its coefficients. In order to check the model structure, t-values were obtained by leaving one session k ($k=1,\ldots,40$) out of the estimation process. The stability of the obtained t-values can be seen in Fig. 5. Obviously the model is resistant against the exclusion of sessions. The stability of t-values confirms that the group of sessions (number of observations for the GLM) is sufficiently large.

4. Discussion

4.1. Selection criteria of observables

When exploring the observables entering the model we aimed to maximise the recognition rate of individuals while keeping the number of parameters in the GLM low. A larger number of parameters would make the GLM-equation unnecessarily complicated and thus introduce deleterious overfitting. The latter would immediately deteriorate our true out-of-sample performance measure. Therefore we selected a subset of only three electrode sites to avoid overfitting, as would have been a risk with all 60 electrodes, and provide adequate robustness of the method, which requires data from more than one electrode site. Among other observables we explored were the standard deviation of the spectral estimation (Stassen, 1980) and regression between coherence spectra. The recognition rate of the method appears as a useful criterion to select EEG observables which faithfully describe the EEG of an individual.

4.2. The final set of observables

We chose the midline sites AFz, Cz and Pz (Fig. 1) because they show little contamination with muscle artefact and display high dominant peaks. The high data quality may explain why the recognition rate is similar with or without ICA artefact rejection. The choice of electrode sites accounts also for spatial variation of the EEG on the scalp because the dominant frequency at frontal sites is generally lower than at parietal sites (Niedermeyer and Lopes da Silva, 1999) and thus the signals recorded from the three sites may reflect different aspects of brain function.

The parameters peak height and peak frequency describe the dominant peak in the EEG of persons with eyes closed and thereby put weight on the frequencies below 12 Hz. The height of the dominant peak is given in units of absolute spectral power, thereby supporting earlier publications which recommend absolute rather than relative spectral power for recognition (Pollock et al., 1991; Kondacs and Szabo, 1999).

To account for spectral shape, we regressed spectral power of two recordings and calculated one t_{power}-value for each pair (Fig. 3). Normalizing spectral power to relative values before the regression would not affect the t-value. Alternatively, one could have divided the spectrum into several frequency bands and calculated several comparisons. However, this would lengthen the coefficient vector β of the GLM and possibly lead to overfitting. Since the number of frequency points above 12 Hz is relatively large among all frequency points entering the regression (Fig. 5), t_{power} is strongly influenced by EEG power above 12 Hz. In this frequency range, recognition of spectra is facilitated by the reduced variance achieved by the multi-taper spectral analysis.

4.3. Efficiency of the statistical model

The GLM provides a method to combine the three spectral parameters and the observations from the three recording sites for each recording session. The high recognition rate indicates that a well-founded choice of spectral characteristics of the EEG can be encoded in a GLM with high statistical efficiency. The results of the statistical model were cross-validated with the leave-one-out method (out-of-sample testing). Therefore the results obtained are not only valid for the sample of persons in this study but also for the basic population and thereby achieve high explanatory power. Taken together, the good performance of the statistical apparatus indicates that the test–retest reliability can be used to select spectral observables of the EEG which are a valid signature of a person.

Acknowledgements

We cordially thank V. Rousson for fruitful discussions and D. Jeanmonod for local support. We thank the Swiss National Science foundation for financial support (SNF Grant 3200B0-11028) and three anonymous reviewers for their comments on an earlier version of the manuscript.

References

Buckelmüller J, Landolt HP, Stassen HH, Achermann P. Trait-like individual differences in the human sleep electroencephalogram. Neuroscience 2006;138:351–6.

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21.

Dustman RE, Shearer DE, Emmerson RY. Life-span changes in EEG spectral amplitude, amplitude variability and mean frequency. Clin Neurophysiol 1999;110:1399–409.

Gasser T, Bacher P, Steinberg H. Test–retest reliability of spectral parameters of the EEG. Electroencephalogr Clin Neurophysiol 1985;60:312–9.

Kondacs A, Szabo M. Long-term intra-individual variability of the background EEG in normals. Clin Neurophysiol 1999;110:1708–16.

Landolt HP, Retey JV, Tonz K, Gottselig JM, Khatami R, Buckelmuller I, et al. Caffeine attenuates waking and sleep electroencephalographic markers of sleep homeostasis in humans. Neuropsychopharmacology 2004;29:1933–9.

Maltez J, Hyllienmark L, Nikulin VV, Brismar T. Time course and variability of power in different frequency bands of EEG during resting conditions. Neurophysiol Clin 2004;34:195–202.

Niedermeyer E, Lopes da Silva FH. Electroencephalography: basic principles, clinical applications, and related fields. Philadephia: Lippincott Williams & Wilkins; 1999.

Nunez PL, Wingeier BM, Silberstein RB. Spatial–temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks. Hum Brain Mapp 2001;13:125–64.

Percival DB, Walden AT. Spectral analysis for physical applications. Cambridge: Cambridge University Press; 1993.

Pollock VE, Schneider LS, Lyness SA. Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. Electroencephalogr Clin Neurophysiol 1991;79:20–6.

Poulos M, Rangoussi M, Alexandris N, Evangelou A. Person identification from the EEG using nonlinear signal classification. Methods Inf Med 2002;41:64–75.

Salinsky MC, Oken BS, Morehead L. Test–retest reliability in EEG frequency analysis. Electroencephalogr Clin Neurophysiol 1991;79:382–92.

Sarnthein J, Stern J, Aufenberg C, Rousson V, Jeanmonod D. Increased EEG power and slowed dominant frequency in patients with neurogenic pain. Brain 2006;129:55–64.

Stassen HH. Computerized recognition of persons by EEG spectral patterns. Electroencephalogr Clin Neurophysiol 1980;49:190–4.

Vuga M, Fox NA, Cohn JF, George CJ, Levenstein RM, Kovacs M. Long-term stability of frontal electroencephalographic asymmetry in adults with a history of depression and controls. Int J Psychophysiol 2006;59:107–15.