NeuroImage 43 (2008) 687–693

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Test–retest reliability of EEG spectra during a working memory task

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

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

ARTICLE INFO

Article history:
Received 4 January 2008
Revised 11 August 2008
Accepted 14 August 2008
Available online 4 September 2008

Keywords:
Sternberg
Alpha
Generalized linear model
GLM
Cross-validation

ABSTRACT

The inter-individual variation of EEG spectra is large even for the same cognitive task. We asked whether task-induced EEG spectra remain stable over more than a year.
We recorded EEG in 41 healthy adults who performed a modified Sternberg task. In 20 subjects EEG was recorded in a second session with retest intervals 12–40 months. For electrodes AFz, Cz and Pz peak frequency and peak height were determined. We compared the curve shape of power spectra by regressing spectra pairwise onto each other and calculated a *t*-value. The *t*-value and pairwise differences in peak frequency and peak height between all 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.
Of the 40 sessions, 35 were correctly matched. The shape of power spectra contributed most to recognition. Out of all 2400 pairwise comparisons 99.3% were correct, with sensitivity 87.5% and specificity 99.5%.
The intra-individual stability is high compared to the inter-individual variation. Thus, interleaved EEG-fMRI measurements are valid. Furthermore, longitudinal effects on cognitive EEG can be judged against the intra-individual variability in subjects.

Introduction

The human scalp EEG non-invasively provides direct access to electrical activity of neurons. To assess the validity of EEG measurements, the different sources of variability have to be understood. Obviously, the use of cognitive tasks aims to correlate well-defined cognitive states with the variation of EEG observables. While this can rarely be achieved on a single trial level, average results for groups of subjects are generally accepted. In research on working memory (WM), most studies average over the group and consider some individuals as outliers, e.g. Jensen et al. (2002).

The inter-individual variability of physiological observables has been the focus of a smaller number of studies, e.g. Meltzer et al. (2007) and Michels et al. (2008). In this earlier publication (Michels et al., 2008), we have analyzed several aspects of the EEG activity elicited by WM retention, in particular the workload dependence of theta and alpha power, the time-frequency pattern of activity, and the cortical generators of the EEG activity. The workload dependence showed the largest effects at electrodes AFz and Pz and for alpha power it was positive in some subjects and negative in others. The alpha activity that increased with workload was generated in the cuneus and is seen as functional inhibition of visual brain areas less relevant to the task. The alpha activity that decreased with workload was generated in the precuneus and is seen as release of functional inhibition of these areas to meet the increased attentional demands. The generators of both types of alpha activity seem to be different to that during eyes closed resting. These results indicate that different individuals draw on different brain processes to different extent in their effort to perform well in the task. Furthermore, alpha activity in the EEG can distinguish between these individuals and describe the inter-individual variability.

At the level of the intra-individual variability, the question remains whether spectral EEG observables are reproducible over time. Spectral EEG observables in adults are known to be quite stable for the resting condition (Stassen, 1980; Gasser et al., 1985; Pollock et al., 1991). The genetic contribution to the stability of several spectral (and other) EEG observables has been shown to be significant in a large number of twin and family studies (van Beijsterveldt and van Baal, 2002). In an earlier publication, we have introduced a new method to assess the test–retest reliability of resting EEG spectra (Näpflin et al., 2007). A *t*-value describing the similarity of spectral shape and differences in alpha peak power and frequency were combined in a generalized linear model (GLM) to obtain the recognition probability of individual subjects. The recognition probability is high if the intra-individual variability is low with respect to the inter-individual variability.

The brain state during task performance differs from the resting condition in several aspects. First, the arousal during the task performance desynchronizes the EEG (Moruzzi and Magoun, 1949). Second, the subjects have their eyes open, which reduces alpha power

* Corresponding author. Neurochirurgie, UniversitätsSpital Zürich, CH-8091 Zürich, Switzerland.
E-mail address: johannes.sarnthein@usz.ch (J. Sarnthein).

1053-8119/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.08.028

with respect to the eyes closed resting condition. Third, for the resting condition large amounts of data can be easily acquired. The amount of data, which can be extracted from the correct trials, is in general much smaller. Fourth, during the resting condition subjects follow their individual thoughts which might lead to a large variability in the EEG between subjects and also between sessions of the same subject. In the task condition, subjects focus their mental resources on the task, which might lead to a reduced variability in the EEG between sessions of the same subject, i.e. low intra-individual variability. Since the task is the same for all subjects, this might lead to enhanced similarity between EEG spectra between subjects, i.e. low inter-individual variability.

Our present study was motivated by the above considerations on the task/resting differences and addresses the test–retest reliability of the data obtained during WM task performance (Michels et al., 2008) with the GLM method (Näpflin et al., 2007). The length of the retest interval is longer than that of other studies (Salinsky et al., 1991; McEvoy et al., 2000; Fingelkurts et al., 2006) and is motivated by a clinical study, which found sizeable changes in EEG spectral observables as a response to therapy only after 1 year (Sarnthein et al., 2006). In a longitudinal approach, we quantify the test–retest reliability of the EEG recorded during WM retention and also discuss it with respect to the reliability of the eyes closed resting condition.

Methods

Participants

A group of 20 healthy volunteers (age 25–76 years at first session, median 58.5 years, 7 women, 13 men) participated in two EEG sessions with a minimum retest interval of 1 year (retest interval 12–40 months, median 15 months). The variability in the test–retest interval was introduced to test for a systematic influence of interval length on test–retest reliability. To document the specificity of the method, we included an additional group of 21 subjects where EEG was recorded only once, leading to a total population of 41 subjects (19–76 years, median 44 years, 16 women, 25 men). The study was approved by the Kanton Zürich ethics committee. All subjects were informed about the aim and the scope of the study and gave written informed consent according to the declaration of Helsinki. All subjects were screened for health problems using a detailed health questionnaire. The subjects had no current or previous history of relevant physical illness and they were not currently taking drugs or medication known to affect their EEG.

Design of the cognitive task

We used a modified Sternberg task where encoding of memory items, retention and recall are temporally separated (Fig. 1) (Jensen et al., 2002). Each trial started with the word “Blink”, encouraging the subjects to make eye blinks to reduce artefacts later in the trial. After the subjects triggered the advancement of the trial by button press, a fixation dot was presented for 1 s. After that, a set of consonants (stimulus) was presented at the center of the screen for 2 s and had to be retained in memory for 3 s. All stimuli contained eight white consonants on a black screen. The screen size was 32.5 cm*30 cm. The stimulus size was 10 cm*25 cm, which was divided equally between the eight letters of the stimulus. Subjects viewed the screen from a distance of 1 m. Of the eight consonants the middle four, six or eight letters were the memory items. For memory setsize (ss) 4 and 6, the outer positions were filled with ‘X’, which was never a memory item. Thus, the physical size and the visual content of the stimulus was the same, irrespective of the size of the memory set. During both stimulus presentation and retention interval, a fixation dot was presented at the center of the screen. After the retention interval, a probe letter replaced the fixation dot. There was a 50% probability that the probe letter was part of the memory stimulus or not. Subjects indicated whether the probe was part of the stimulus by button press on a joystick (‘yes/no’ procedure). Since there was no significant difference in our results for correct ‘yes’ and correct ‘no’ trials, data was pooled across these trial types. Subjects were instructed to aim not for fast but for correct responses and only correct trials entered the final analysis.

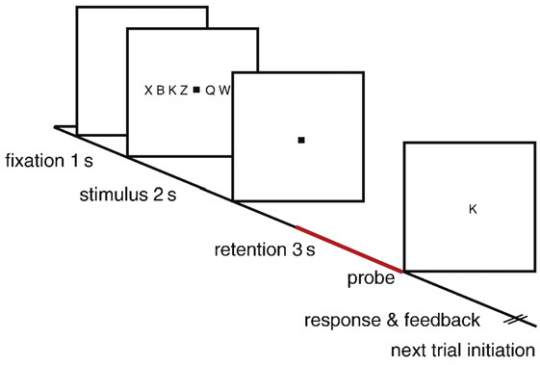

Fig. 1. Working memory task. Sets of consonants (stimulus) had to be retained in memory for 3 s. The setsize (ss 4, 6, or 8 letters) determines the memory workload. For ss 4 and ss 6, four and two positions on the screen were filled with the letter ‘X’ to keep the number of letters constant on the screen. After the retention interval, a probe letter was shown. Analysis focused on the last 2 s of the retention interval (red line). Subjects were asked to decide whether the probe was part of the stimulus by pressing a button.

EEG recording sessions

Subjects were seated in a dimly lit room shielded against sound and stray electric fields and were video monitored. The recording sessions were performed between 9 to 12 h in order to exclude an impact of circadian factors on the EEG. Subjects abstained from caffeinated beverages on the day of recording to avoid the caffeine-induced theta decrease in the EEG (Landolt et al., 2004). Before each recording, subjects were instructed to assume a comfortable position in a chair and were free to place their head on a chin-rest.

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 43 recording sessions, we used passive Ag/AgCl electrodes (Easycap, Herrsching, Germany). For the remaining 18 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 Hz–100 Hz, −12 dB/octave) and continuously viewed on PC monitor.

Data preprocessing and editing

Data were analyzed offline in Matlab (The Mathworks, Natick, MA) using EEGLAB (http://sccn.ucsd.edu/eeglab/index.html) (Delorme and Makeig, 2004), and custom scripts. First, 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). Data was then high-pass filtered with a filter of 0.5 Hz to remove linear trends that would negatively affect the independent component analysis (ICA). Only correct trials were used for the analysis. In this publication we analyze data only from the last 2 s of the retention interval (Fig. 1). This avoids interference from visual evoked responses and eye-blink artefact as well as the somewhat arbitrary choice of a baseline to subtract. The