

Application of modern tests for stationarity to single-trial MEG data

Transferring powerful statistical tools from econometrics to neuroscience

Lech Kipiński · Reinhard König · Cezary Sielużycki ·
Wojciech Kordecki

Received: 14 June 2010 / Accepted: 30 August 2011 / Published online: 18 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Stationarity is a crucial yet rarely questioned assumption in the analysis of time series of magneto- (MEG) or electroencephalography (EEG). One key drawback of the commonly used tests for stationarity of encephalographic time series is the fact that conclusions on stationarity are

only indirectly inferred either from the Gaussianity (e.g. the Shapiro–Wilk test or Kolmogorov–Smirnov test) or the randomness of the time series and the absence of trend using very simple time-series models (e.g. the sign and trend tests by Bendat and Piersol). We present a novel approach to the analysis of the stationarity of MEG and EEG time series by applying modern statistical methods which were specifically developed in econometrics to verify the hypothesis that a time series is stationary. We report our findings of the application of three different tests of stationarity—the Kwiatkowski–Phillips–Schmidt–Schin (KPSS) test for trend or mean stationarity, the Phillips–Perron (PP) test for the presence of a unit root and the White test for homoscedasticity—on an illustrative set of MEG data. For five stimulation sessions, we found already for short epochs of duration of 250 and 500 ms that, although the majority of the studied epochs of single MEG trials were usually mean-stationary (KPSS test and PP test), they were classified as nonstationary due to their heteroscedasticity (White test). We also observed that the presence of external auditory stimulation did not significantly affect the findings regarding the stationarity of the data. We conclude that the combination of these tests allows a refined analysis of the stationarity of MEG and EEG time series.

L. Kipiński and W. Kordecki formerly at Wrocław University of Technology.

Present Address:

L. Kipiński (✉)
Department of Pathophysiology, Wrocław Medical University,
ul. K. Marcinkowskiego 1, 50-368 Wrocław, Poland
e-mail: lech.kipinski@patfiz.am.wroc.pl

L. Kipiński
Department of Neurology, T. Marciniak Memory Lower
Silesia Specialist Hospital—Medical Emergency Centre,
ul. Traugutta 116, 50-420 Wrocław, Poland

L. Kipiński · W. Kordecki
Institute of Biomedical Engineering and Instrumentation, Wrocław
University of Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland

R. König · C. Sielużycki
Special Lab Non-invasive Brain Imaging, Leibniz Institute
for Neurobiology, Brennekestr. 6, 39118 Magdeburg, Germany
e-mail: reinhard.koenig@lin-magdeburg.de

C. Sielużycki
e-mail: cezary.sieluzycki@lin-magdeburg.de

W. Kordecki
Department of Economics and Accountancy, University
of Business in Wrocław, ul. Ostrowskiego 22,
53-238 Wrocław, Poland
e-mail: wojciech.kordecki@handlowa.eu

Keywords Magnetoencephalography (MEG) ·
Nonstationarity · Heteroscedasticity · KPSS test · PP test ·
White test

1 Introduction

Magnetoencephalography (MEG) and electroencephalography (EEG) directly record the activity of neural cell

assemblies and allow tracking of brain dynamics with millisecond temporal resolution. Any analysis of the time course of the acquired magnetic fields or electric potentials faces the problem that the measured signal is a superposition of contributions from different sources. Beside a component evoked, for example, by an external sensory stimulus which results in a response time-locked to the stimulus onset, and which reveals a certain inter-trial variability, the ongoing brain activity consists of oscillations and of transient structures characteristic to different behavioural or even pathological cases. Above all, there is noise originating from the acquisition electronics, or from magnetically or electrically active parts of the human body, such as the heart or the eyes. In view of these numerous signal components, and taking into account their complexity, it seems rather unlikely that the tacit, yet common, assumption of signal stationarity is always justified.

From the statistical point of view, evoked magnetic fields (EFs) or electric potentials (EPs) can be considered as stochastic processes, with the single-trial signals acquired during repeated stimulation of a subject being their realizations. The characterization of encephalograms by means of stochastic time series is widely used (see, for example, Gersch 1970; Wright et al. 1990; Gasser and Molinari 1996, and references therein), and recently, some authors (see, for example, Krystal et al. 1999; Wong et al. 2005) applied methods of time-series analysis to model the time-varying structure of the statistical properties of brain signals. However, one has to be aware of the fact that most of the existing methods of time-series analysis have been developed to describe stationary rather than nonstationary data. Therefore, the application of these methods to the analysis of MEG or EEG signals is restricted to relatively short fragments of recordings (Barlow 1985). The duration of a so-called quasi-stationary interval of continuous EEG recordings is expected not to exceed 2–4 s (Kaplan 1999), but some authors found much longer fragments of even 12 s (Cohen and Sances 1977), 25 s (Kawabata 1976) or 40–60 s (McEwen and Anderson 1975) to be approximately stationary.

Our motivation for this study was to test encephalographic brain signals for stationarity by using advanced tests which were specifically developed to test stationarity of time series. These tests have been so far applied in econometrics, not in brain research. We report our findings of applying three different tests of stationarity—the Kwiatkowski–Phillips–Schmidt–Schin (KPSS) test, the Phillips–Perron (PP) test and the White test—to a set of MEG data acquired from an exemplary subject, who was exposed to a specific auditory stimulation protocol. We do not intend to provide a general proof of the stationarity or nonstationarity of MEG/EEG signals; rather, we approach the problem of testing stationarity from a methodological point of view.

2 Methods

2.1 Stationarity and nonstationarity

2.1.1 Definition of stationarity

Let us consider a time series Z_t with a finite-dimensional joint distribution function

$$F(x_1, x_2, \dots, x_n) = P(Z_{t_1} \leq x_1, Z_{t_2} \leq x_2, \dots, Z_{t_n} \leq x_n). \quad (1)$$

The strict stationarity of a time series $\{Z_t, t = -1, 0, 1, \dots\}$ is defined by the condition that (Z_1, Z_2, \dots, Z_n) and $(Z_{1+h}, Z_{2+h}, \dots, Z_{n+h})$ have the same joint distributions for all integers h and $n > 0$. Unfortunately, the multidimensional distribution function usually cannot be formulated easily, and this definition of stationarity is too strong for most applications. Therefore, in many practical applications, the notion of weak stationarity, which imposes conditions only on the first and the second moment of $\{Z_t\}$, is used. Clearly, a proper definition of these two moments is then required (see, for example, Box and Jenkins 1970; Shumway and Stoffer 2000; Brockwell and Davis 2002).

The stochastic process $\{Z_t\}$ is weakly stationary if the mean value

$$E(Z_t) = \mu_Z(t) = \text{const} = \mu, \quad (2)$$

and the autocovariance function (ACVF) is invariant with respect to translations:

$$\gamma(h) = \text{Cov}(Z_t, Z_{t+h}) \quad (3)$$

for $t = 0, \pm 1, \dots, h = 0, \pm 1, \dots$, and depends only on h , not on t .

A wide class of stationary time series is addressed by the autoregressive moving-average (ARMA) model, defined for every t as

$$z_t = \sum_{i=1}^p \varphi_i z_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + e_t, \quad (4)$$

where $\{e_t\}$ is the white noise process $\text{WN}(0, \sigma^2)$ and the two polynomials have no common factors.

2.1.2 Types of nonstationarity

Nonstationarity can be defined as a property of a time series where the criteria set by Eqs 2 or 3 are not fulfilled. Possible reasons why the statistical properties of a time series vary in time include the following cases:

1. The process has a deterministic trend,
2. The process has a single or multiple unit roots,

3. The process is heteroscedastic, i.e., it is nonstationary in variance,
4. The autocorrelation function of the process is time dependent.

With regard to the first case, let us consider a time series

$$z_t = \alpha + \beta b_t + e_t, \tag{5}$$

where e_t denotes a time series of independent and identically distributed random variables (IID). α and β are some constants, and b_t is a deterministic function which describes the variability of the mean of Z_t in time. This variability may be represented by a deterministic function from a large family of functions (for example, linear, exponential, logarithmic and others), where the values of time series will differ from such a function randomly. Low-frequency oscillatory behaviour of b_t can be described by few-parameter regression models.

Regarding the second case, let us consider the time-series model given by Eq. 4. Its characteristic polynomial is defined by

$$m^p - m^{p-1}\varphi_1 - \dots - \varphi_p = 0. \tag{6}$$

A unit root is a root $m = 1$ of Eq. 6. Stochastic processes which have unit roots are described by autoregressive integrated moving-average (ARIMA) models. Such processes become stationary after differencing d times, where d is the multiplicity of a unit root. A theoretical account of these models was introduced by [Box and Jenkins \(1970\)](#). In our study, we consider the simplest case of such kind of nonstationarity, i.e., when at most one single unit root ($d = 1$) is present in the time series.

Nonstationarity can also be due to the variation of either the mean or the variance in time (case 3). Many time series with such properties can be described by means of the generalized autoregressive conditional heteroscedasticity (GARCH) models proposed by [Bollerslev \(1986\)](#).

Let us consider a time series z_t with the following covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & \\ & \dots & \\ & & \sigma_N^2 \end{bmatrix} \mathbb{I}, \tag{7}$$

where \mathbb{I} is an identity matrix. If all the elements of the leading diagonal of Σ are equal, the time series has constant variance

$$\text{Var}(z) = E(z_i^2) = \sigma^2 \tag{8}$$

and is homoscedastic. Otherwise,

$$\text{Var}(z) = E(z_i^2) \neq \text{const}, \tag{9}$$

and the time series is heteroscedastic. Note that, for a Gaussian multidimensional distribution, random variables are independent if they are uncorrelated. Of course, in more general

situations, as e.g. described by GARCH models, the existence of an autocorrelation function is not assumed, or Σ does not need to be diagonal.

Case 4 requires the estimation of the parameters of a time-dependent autocorrelation function, which we do not discuss here, since this topic is beyond the scope of this study.

2.2 Testing encephalographic time series for stationarity: a brief survey of the literature

A commonly used approach to test short EEG time series for stationarity is based on the sign (or the run) test and the trend test proposed by [Bendat and Piersol \(1986\)](#). These methods were introduced for testing the properties of a series of means and variances calculated for segments of a signal. They can also be applied to power spectra of MEG or EEG signals, provided that the signals are split into equally long segments, and that each segment is considered independent from all other segments ([Kawabata 1976](#)). Note, however, that the sign test is a test of randomness widely used for testing random-number generators, whereas the trend test detects the regression slope in a data set and is usually performed under a specific model for the residuals. Therefore, neither test explicitly tests for stationarity.

Alternatively, stationarity of encephalographic signals is also inferred from the results of tests for Gaussianity. The Shapiro–Wilk test ([Bender et al. 1992](#)) or the Kolmogorov–Smirnov test ([Weiss 2005](#)) are often used, but these tests have their limitations. The Kolmogorov–Smirnov test, for instance, requires reduction of the effective sampling rate by discarding some samples, because the method requires consecutive samples to be independent. This is achieved by increasing the distance between samples, i.e., subsampling. Above all, these tests are only indirectly linked to the investigation of stationarity.

The problem of stationarity of MEG and EEG brain signals is also closely related to the trial-to-trial variability of evoked responses. It is possible to infer about the stationarity of long-time MEG/EEG recordings from the statistical homogeneity of single-trial responses. [Möcks et al. \(1984a,b\)](#) introduced special tests for the homogeneity of stimulus-related EEG data.

2.3 Tests for the stationarity of time series

The problem of the evaluation of stationarity of time series has been extensively discussed in econometrics, resulting in the development of the Kwiatkowski–Phillips–Schmidt–Schin (KPSS) test for trend- or mean-stationarity ([Kwiatkowski et al. 1992](#)), the class of tests for the presence of unit roots—the Dickey–Fuller (DF) test, the augmented Dickey–Fuller (ADF) test ([Dickey and Fuller 1979](#); [Said and Dickey 1984](#)), or the Phillips–Perron (PP) test ([Phillips and](#)

Perron 1988)—and the White test for variance-stationarity (White 1980). We studied the stationarity of MEG time series by means of these statistical tests. They were used to investigate short epochs of brain signals in the time domain and the trial-to-trial variability of selected components of the frequency spectra.

2.3.1 The Kwiatkowski–Phillips–Schmidt–Schin test for mean-stationarity

The KPSS test is a powerful statistical test for the null hypothesis that a time series is mean- or trend-stationary (Kwiatkowski et al. 1992). The time series is expressed as a sum of a deterministic trend ξt , a random walk component r_t , and a stationary error e_t :

$$z_t = \xi t + r_t + e_t. \tag{10}$$

The test is in fact the Lagrange-multiplier (LM) test (see Tanaka 1983; Saikkonen 1993) of the hypothesis that the random walk has zero variance. The test statistic is given by

$$\hat{\eta} = T^{-2} \sum S_t^2 / s^2(l), \tag{11}$$

where $t = 1, 2, \dots, T$. S_t is the partial sum of the residuals e_t from the regression of z on an intercept and on the time trend, l is the lag truncation parameter, and s^2 is the Newey–West estimator of the long-run variance σ^2 (Kwiatkowski et al. 1992)

$$s^2(l) = T^{-1} \sum_{t=1}^T e_t^2 + 2T^{-1} \sum_{r=1}^l w(r, l) \sum_{t=r+1}^T e_t e_{t-r} \tag{12}$$

with

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} E(S_T^2) \approx \frac{1}{T} \sum_{i=1}^T \sigma_i^2. \tag{13}$$

Here, $w(r, l) = 1 - r / (l + 1)$ is an optional Bartlett window which serves (among other similar weighting functions) to ensure positivity of the variance estimator. Note that the Newey–West estimator is heteroscedasticity and autocorrelation consistent (Newey and West 1987).

The asymptotic distribution of the statistic is rather complicated, expressible by an integral of a squared Brownian bridge, and the critical values have been determined by simulation or, alternatively, by numerical inversion of its characteristic function (for details, see Kwiatkowski et al. 1992). It is derived under the null hypothesis in two versions, mean-stationarity ($\xi = 0$) or trend-stationarity ($\xi \neq 0$), and under the alternative hypothesis that the time series is non-stationary due to the presence of unit root. Interpretation of the results of the test is based on comparison of the value of the test statistic with the critical value for a chosen significance level. If the test statistic has a value smaller than the critical value, there is no reason to reject the null hypothesis.

2.3.2 The Phillips–Perron test for unit root

In statistics, particular emphasis in testing time series for stationarity is dedicated to tests for unit roots. Let us assume that the time series has at most one unit root. The well-established tests to verify such a hypothesis are the Dickey–Fuller class tests—the DF and the ADF test introduced by Dickey and Fuller (1979) and by Said and Dickey (1984), respectively. They are based on the assumption that the residuals in the time-series model represent a white-noise process $WN(0, \sigma^2)$, which, however, is a simplified assumption.

Phillips and Perron have developed a more comprehensive theory of unit-root nonstationarity. Their test, the Phillips–Perron (PP) test (Phillips 1987; Perron 1988; Phillips and Perron 1988), is a nonparametric equivalent of the ADF test, but incorporates automatic correction to the DF procedure to allow for autocorrelated residuals. For a time series $\{z_t\}$ consider the following regression model:

$$\Delta z_t = \mu + \beta z_{t-1} + e_t, \tag{14}$$

where $t = 1, 2, \dots, T$, which is equivalent to the first-order autoregressive process AR(1) with zero-mean error e_t . The test statistic is defined by

$$\tau_{PP} = t_\beta \frac{\gamma_0^{1/2}}{\hat{\omega}} - \frac{T}{2} s_\beta \frac{(\hat{\omega}^2 - \gamma_0)}{\hat{\omega} \hat{\sigma}}. \tag{15}$$

Here, t_β denotes the regression t -statistic, s_β is the error of estimation of the parameter β , $\hat{\sigma}^2$ is the variance estimator for e_t , and $\hat{\omega}^2$ is the Newey–West long-run variance estimator given by

$$\begin{aligned} \hat{\omega}^2 &= \gamma_0 + 2 \sum_{j=1}^m w(i, j) \gamma_j \\ &= \frac{1}{T} \sum_{t=1}^T \hat{e}_t^2 + \frac{2}{T} \sum_{j=1}^m w(j, m) \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j}, \end{aligned} \tag{16}$$

where $\gamma_j = 1/T \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j}$ for $j = 0, 1, 2, 3, \dots$, and the kernel is given by the Bartlett weights $w(j, m) = 1 - j / (m + 1)$ for m truncation lags. The null hypothesis H_0 is that $\beta = 0$, which is equivalent to the presence of unit root.

2.3.3 The White test for variance-stationarity

The White test (White 1980) is a direct test for homoscedasticity. It uses the classical model of linear regression

$$z_t = \beta x_t + e_t. \tag{17}$$

To introduce the idea of this test, let us define the consistent ordinary least-squares (OLS) estimator of a $K \times 1$ vector β of unknown parameters

$$\hat{\beta}_n = (x'x)^{-1} x'z, \tag{18}$$

where x is an $n \times K$ matrix with rows x_t , and z is an $n \times 1$ vector with elements z_t . Unfortunately, e_t is not observable; however, it can be estimated as

$$\hat{e}_{tn} = z_t - \hat{\beta}_n x_t.$$

The artificial regression for the α parameters estimated as $\hat{\beta}$ in Eq. 18 is defined by

$$\hat{e}_{tn}^2 = \alpha_0 + \sum_{i=1}^K \sum_{j=1}^K \alpha_s x_{ti} x_{tj}, \tag{19}$$

and the White heteroscedasticity consistent estimator is given by

$$\text{Var}(\hat{\beta}) = n (x'x)^{-1} \hat{V}_n n (x'x)^{-1}, \tag{20}$$

where

$$\hat{V}_n = \frac{1}{n} \sum_{i=1}^n \hat{e}_{tn}^2 x_t x_t'. \tag{21}$$

In the fixed regressor case, this amounts to replacing the t th diagonal element σ_t^2 of Σ given by Eq. 7 with \hat{e}_{tn}^2 , the t th squared residual (White 1980). The null hypothesis of the White test is that the series is homoscedastic (i.e. it has a time-independent variance), which is asymptotically equivalent to testing the joint hypothesis that $\alpha_1 = \alpha_2 = \dots = \alpha_{K(K+1)/2} = 0$ using the standard R^2 statistic from the regression (Eq. 19). The LM statistic equal to nR^2 has asymptotic χ^2 distribution with $K(K + 1)/2$ degrees of freedom. If the statistic is small, we have reason to reject the null hypothesis. In our case, the number of observations n is constant and K is fixed.

It is important that the proposed parameter-covariance-matrix estimator is consistent even when the disturbances of the linear regression model are heteroscedastic. This estimator does not depend on the formal model of the structure of heteroscedasticity.

2.4 Testing the trial-to-trial variability of power spectra

We used an algorithm similar to the one introduced by Bendat and Piersol (1966) and later applied by, for example, Kawabata (1976), but we focussed on the properties of the variability of power spectra only. In accordance with that study, we applied a three-step procedure which includes: (1) signal segmentation, (2) estimation of parameters and (3) testing for stationarity, but we introduced some important modifications. Kawabata (1976) divided the sample record into T time intervals of equal duration, the data of which might be assumed as mutually independent. We also divided the measurement signal into T epochs of equal length, but each epoch was time-locked to the stimulus onset, or, in those measurements which were performed without any external

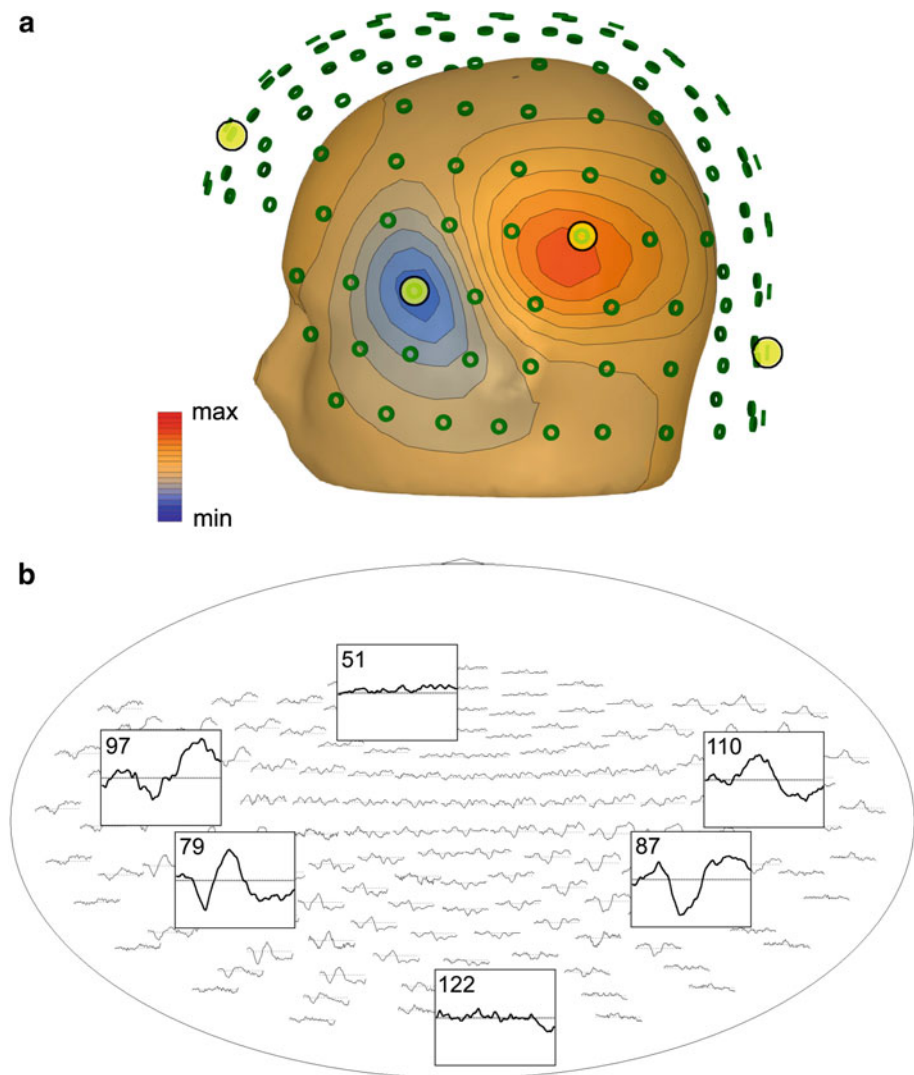
stimulation, to a well-defined trigger point. Due to the constant stimulus-onset interval (SOI) of the experimental sessions used in our study (Sect. 3.1), neighbouring epochs were separated from each other by the same constant time interval, which was determined by the difference between the overall duration of a trial and the duration of an epoch, and which was large enough to assume that the data of such subsequent epochs were mutually independent. Subsequently, we computed the power spectrum for each epoch, and derived a sequence of selected coefficients from these spectra, $\{P_{i1}, P_{i2}, P_{i3}, \dots, P_{iT}\}$, where $i = 1, 2, 3, \dots, N$ indicates the selected coefficients. Like Kawabata (1976), we estimated the power spectra by using the fast Fourier transform (FFT). Note that only one spectral coefficient for each segment was selected at a time. Contrary to Kawabata (1976), where also the sequences of mean and variance were tested with regard to underlying trends or unexpected variations, we did not need to compute cross-sectional means and variances because in the final testing for stationarity we applied the KPSS test, the PP test and the White test for stationarity, which is a new and innovative approach in the analysis of encephalographic time series in brain research.

3 Experimental procedures

3.1 Stimuli and experimental conditions

The statistical tests described in Sect. 2.3 were applied to MEG recordings from a healthy subject who was exposed to a specifically designed stimulation protocol. The experiment consisted of five consecutive sessions. During the first session, spontaneous brain activity was acquired, i.e., the subject was not exposed to any external stimulation. The stimulus of session 2 was a series of sinusoidal 1-kHz tones (duration 500 ms, SOI 2 s). This session was followed by another session without any external stimulation (session 3). During sessions 4 and 5, the same set of frequency-modulated (FM) sweeps (duration 500 ms, SOI 2 s) was presented, in two different experimental conditions. In session 4, referred to as FM/P, the subject was passively listening to the FM sweeps. In session 5, abbreviated as FM/T, the subject was instructed to categorize the sweeps with regard to the direction of their frequency modulation (upward or downward), and to indicate the respective category by pressing one of two response keys. The use of stimuli of different complexity (pure tones, FM sweeps) and different conditions within a stimulus class (passive listening to FM sweeps and the FM categorization task) allowed studying the impact of brain responses resulting from different experimental sessions on the stationarity of the MEG time series. Detailed description of the stimulus material is given in König et al. (2008).

Fig. 1 **a** Schematic of the arrangement of the 148 magnetic field sensors (*black rings*) around a standard head. The two sensors displaying the strongest positive and negative signals over the left hemisphere, as well as the selected frontal and occipital sensors (see text), are marked with semitransparent *large circles*. Iso-magnetic field lines reconstructed from all sensor signals at a point in time near the peak of the M100 are also shown. **b** Projection of the 148 magnetometers onto a plane. The sensors show averaged auditory evoked magnetic fields recorded during session 2 (stimulation with 1-kHz tones). Zoomed channels [two sensors above the auditory cortices of the left (#79, #97) and the right (#87, #110) hemisphere, one frontal sensor (#51) and one occipital sensor (#122)] were selected for further analysis. Ordinate: $B(t)$, abscissa: t



For each condition with auditory stimulation, 96 stimuli were presented to the subject. Data in sessions 1 and 3 were continuously acquired over a period of time, the duration of which equalled the duration of the sessions with auditory stimulation ($96 \times 2 \text{ s} = 192 \text{ s}$). MEG data were acquired with a 148-channel whole-head system (Magnes 2500WH; 4-D Neuroimaging, San Diego, USA), using a sampling rate of 1,017.25 Hz. The lower and upper frequencies of the band-pass filter during acquisition were set to 0.1 and 100 Hz.

3.2 Data processing

Rejection of artefacts (e.g. due to eye blinks) led to a reduction of the number of trials by approximately 15%. To have an equal number of artefact-free trials among all five sessions, we chose 79 trials for all sessions for further analysis, which was the smallest number of artefact-free trials across the five sessions. Since testing for stationarity of the

variability of trial-to-trial power spectra required equal time intervals between consecutive observations, that particular analysis was performed on all measured trials.

We selected some characteristic MEG sensors for the detailed analysis as follows. We first computed the average auditory evoked magnetic field (AEF) across all artefact-free trials separately for each stimulus, and picked, for each hemisphere, those two channels above the auditory cortices that revealed the strongest MEG signal at a latency of about 100 ms. This signal is known as the auditory M100 peak (Näätänen and Picton 1987), see Fig. 1. The resulting spatial distribution, i.e. the bipolarity, of the corresponding magnetic field pattern as shown in Fig. 1 is typical for magnetometers as MEG sensors, and is expected from a source located underneath the midpoint of a line connecting the two extrema. To study the impact of the auditory stimulation, we compared the findings from the 2×2 channels exhibiting the strongest M100 peak (for the pos-

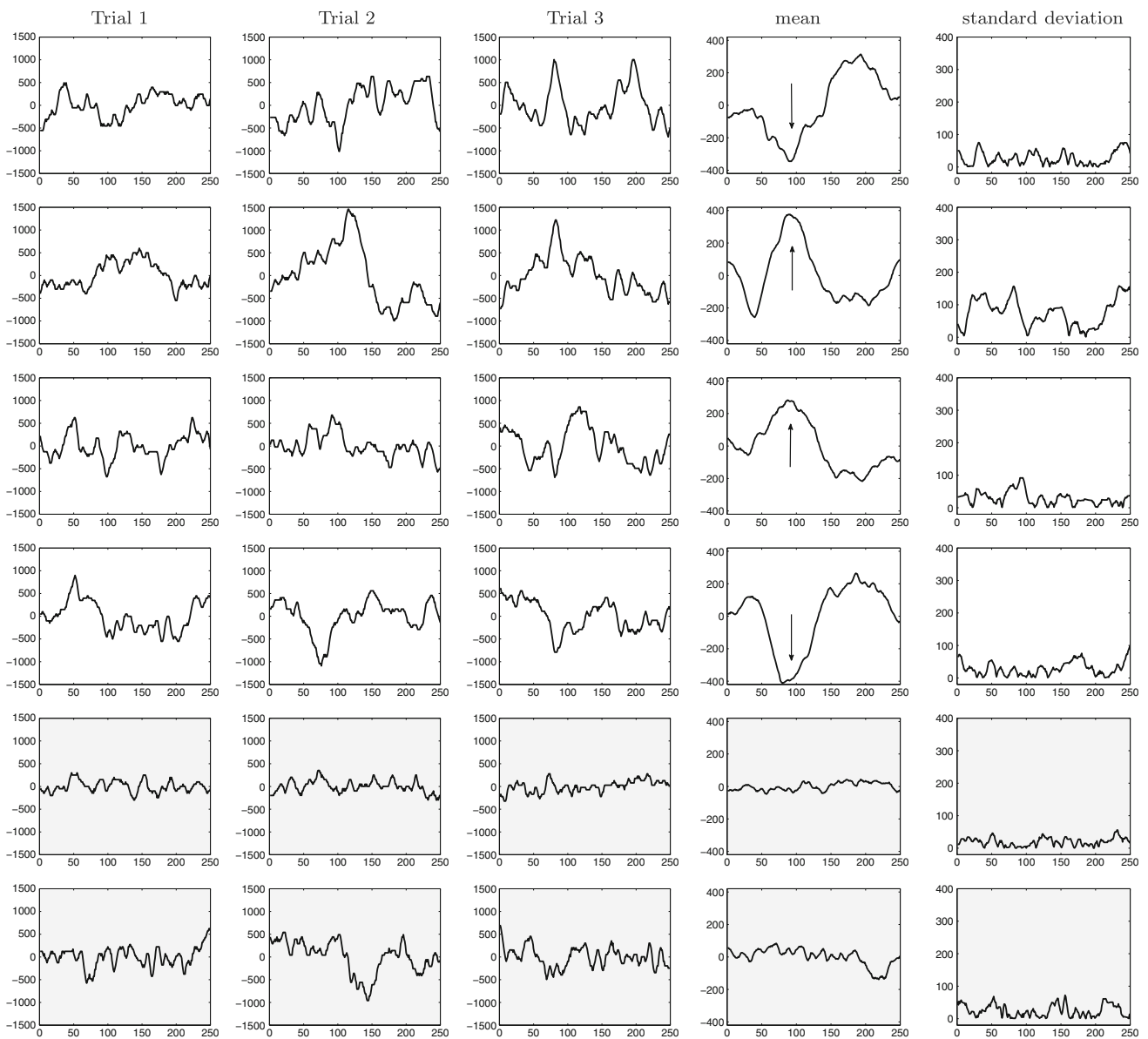


Fig. 2 MEG signals from the measurement with the 1-kHz stimulus. Ordinate: [fT], abscissa: [ms]. The auditory evoked fields were detected by the four sensors with the strongest mean M100 peak (*arrows*) over the left (rows 1, 2) and the right (rows 3, 4) hemisphere. Columns 1–3 show the first three single trials of the measurement. Column 4 displays

the mean signal, and column 5 the standard deviation across all 79 trials. The panels with the *light-grey* background show the corresponding signals of the two reference sensors in the frontal (row 5) and occipital (row 6) area, for which no prominent stimulus-related activity was observed at about 100 ms.

itive and the negative polarity of two hemispheres) with those for two additionally selected “reference” channels, one located above the frontal and one above the occipital brain region, that did not reveal any stimulus-related signal. The results presented were obtained from the separately grouped auditory and reference channels, not from the individual channels.

Figure 2 shows, as an example, the MEG signals of the first 250 ms after stimulus onset, taken from the measurement where the subject was exposed to the 1-kHz stimulus.

The AEFs shown were detected by the four MEG sensors with the strongest mean M100 peak over the left (rows 1, 2) and the right (rows 3, 4) hemisphere. Columns 1–3 show the first three single trials of the measurement. Column 4 displays the mean signal and column 5 the standard deviation across all 79 trials. The prominent maxima or minima indicated by an arrow in column 4 at a latency of about 100 ms refer to the M100. Rows 5 and 6 display the first three single trials, the average signal, and the standard deviation of the selected frontal (row 5) and occipital (row 6) sensor. Note

that the mean signal of these sensors did not reveal any significant stimulus-related features at a latency of about 100 ms (M100).

In order to study the influence of the epoch duration and the pre-stimulus baseline period on the (non)stationarity of the signal, we examined three different cases in which the epochs lasted (1) from stimulus onset to 250 ms, (2) from stimulus onset to 500 ms and (3) from 150 ms prior to stimulus onset to 350 ms after stimulus onset.

4 Results

All calculations were performed using the R Project environment with the freely available “tseries” package (Time Series Analysis and Computational Finance, the R Project; [Trapletti and Hornik 2011](#)). The significance level applied in all tests was $\alpha = 0.05$.

4.1 Testing for mean-stationarity: the KPSS-mean test, the KPSS-trend test and the PP test

Figure 3 shows, on the left ordinate, the percentage of rejection of the null hypothesis for the consecutive stimulus sessions obtained from the analysis using the KPSS-mean test (H_0 : mean-stationarity; circles) and the KPSS-trend test (H_0 : trend-stationarity; squares). To allow a direct comparison of these findings with the PP test (H_0 : unit root; H_1 : mean-stationarity; triangles), the right ordinate shows the percentage of acceptance of the null hypothesis of the PP test. Separately for auditory and reference channels, the subplots of Fig. 3 display the findings for the three different time intervals investigated: the 250-ms interval (Fig. 3a), the 500-ms interval which starts at stimulus onset (labelled 500a, Fig. 3b), and the 500-ms interval which starts 150 ms prior to stimulus onset (labelled 500b, Fig. 3c). A common result obtained from all three tests is that there are only small differences between the findings for the auditory (closed symbols) and the reference (open symbols) channels. These differences appear to be slightly more pronounced for the PP test, yet without any clear systematic tendency. Figure 4 shows the corresponding dependence of the percentage of rejection [KPSS-mean (a) and KPSS-trend test (b)] or acceptance [PP test (c)] for all stimulus sessions on the three different time intervals for the auditory channels only.

4.1.1 KPSS tests for mean- and trend-stationarity

The null hypothesis of mean- or trend-stationarity was rejected only in a small percentage of trials, typically between 5% and 15% (Fig. 3). This observation holds for all five experimental conditions, although there was a certain

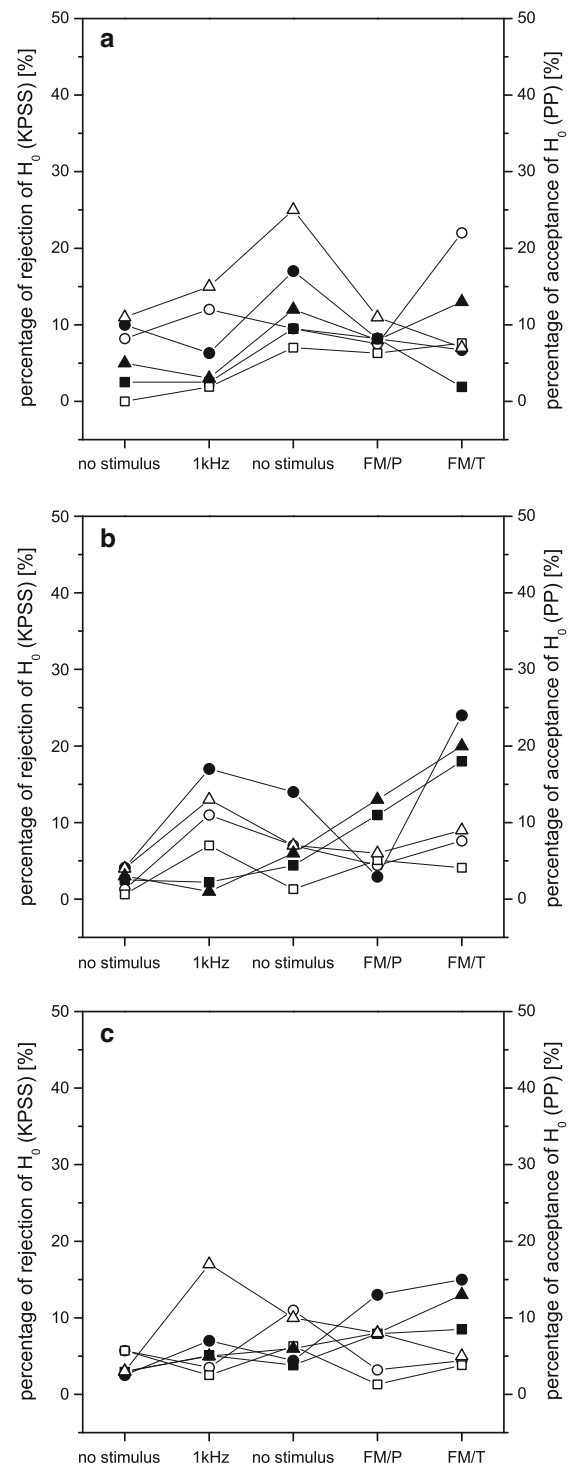


Fig. 3 Percentage of rejection (left ordinate, KPSS tests) or of acceptance (right ordinate, PP test) of the null hypothesis (of mean-stationarity in the KPSS tests and mean-nonstationarity in the PP test, respectively) for the five experimental sessions for the 250-ms interval (**a**), the 500-ms interval starting at stimulus onset (**b**) and the 500-ms interval starting 150 ms prior to stimulus onset (**c**). Solid symbols display results obtained for the channels located above the auditory cortices, open symbols for the reference channels. Circles KPSS-mean test; squares KPSS-trend test; triangles PP test.

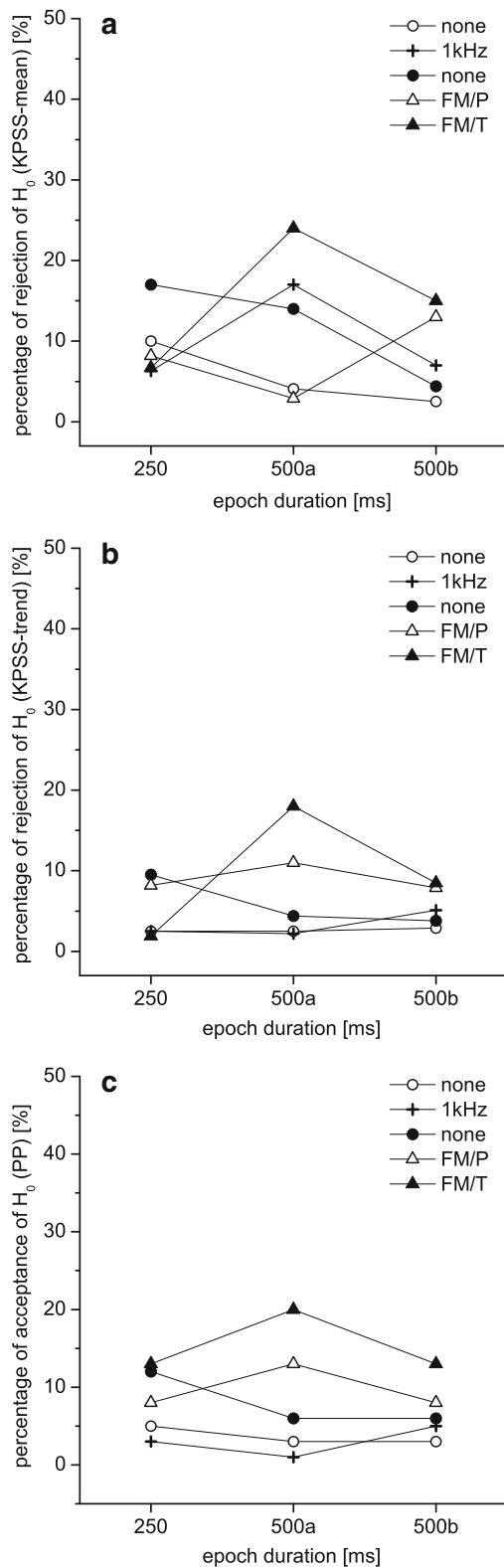


Fig. 4 Percentage of rejection (a KPSS-mean test, b KPSS-trend test) or of acceptance (c PP test) of the null hypothesis of mean-stationarity (KPSS tests) or mean-nonstationarity (PP test) for three different time intervals. See Sect. 4.1 for details on epoch duration

Table 1 Results for the mean-stationarity

Channel	Test	250 ms	500 ms (a)	500 ms (b)
Auditory	KPSS mean	9.7	12	8.4
	KPSS trend	4.9	7.6	5.6
	PP	92	91	93
Reference	KPSS mean	12	6.2	5.5
	KPSS trend	4.6	6.3	5.6
	PP	86	92	91

The table shows the percentage of rejection of the null hypothesis (stationarity of epochs for the two KPSS tests; presence of a unit root for the PP test) averaged across all five stimulus sessions for the auditory and the reference channels (a) and (b) refer to different starting points of the 500-ms intervals (see Sect. 4.1)

variability among the results. Table 1 shows a comparison of the mean percentage of rejection (i.e. data were averaged across the five stimulus sessions). The results obtained from the two KPSS tests lead to the conclusion that the majority of the studied epochs of the single trials were stationary.

For the auditory channels, the mean percentage of rejected trials was somewhat smaller in the case of the test for trend-stationarity, compared with the findings of the test for mean-stationarity (Table 1). This observation may indicate the presence of a linear trend. The findings regarding the dependence of the percentage of rejection on the epoch duration do not allow unambiguous conclusions to be drawn (Fig. 4a, b). The most noticeable observation for the auditory channels is the increase in the percentage of rejection for the FM/T condition (solid triangles) with increasing epoch duration from 250 to 500 ms for both the mean (a) and the trend (b) test. Furthermore, for FM/T, we observed a larger percentage value for the 500-ms interval which starts at stimulus onset (500a) compared with the interval which starts 150 ms prior to stimulus onset (500b). The percentage of rejection for the first session without auditory stimulus (session 1, open circles) is, partly even clearly, smaller for all three epochs compared with the second stimulus-free session 3 (solid circles) for both tests. The results for the 1-kHz tones did not reveal any clear systematic pattern.

4.1.2 PP test for unit roots

The percentage of rejection of the null hypothesis using the PP test (H_0 : unit root; H_1 : mean-stationarity) was, with only two exceptions, always larger than 80% for all sessions and intervals, which corresponds to a percentage of acceptance of 20% or less (Fig. 3c). This observation leads to the conclusion that most of the investigated epochs were mean-stationary, thus confirming the findings of the corresponding KPSS test. Furthermore, we observed a similar dependence

of the percentage of acceptance on the duration of the epoch (Fig. 4c) for the FM/P and the FM/T condition, with larger values for the task condition.

The results obtained for the two sessions without stimulation confirmed the corresponding findings of the two KPSS tests: for all three epochs, the percentage of acceptance for the first session was smaller than for the third session. This observation means that in the first session more epochs were identified as mean-stationary, which is in line with the smaller rejection percentage obtained from the KPSS tests. Moreover, the percentage of acceptance decreased from the 250- to the 500-ms intervals for the two non-auditory sessions.

4.2 Testing for variance-stationarity: The White test

The most systematic results were obtained for the White test of variance-stationarity (H_0 : homoscedasticity; H_1 : heteroscedasticity). Figure 5 shows the corresponding results of the percentage of rejection of the null hypothesis obtained for the auditory channels for the different time intervals. The large rejection values clearly indicate that the majority of the epochs of the single trials were heteroscedastic, i.e. the variance changed over time. Furthermore, the longer the duration of the analysed epoch, the larger the number of rejected epochs, i.e. the more epochs were heteroscedastic. In addition, we also found a dependence of the percentage of rejection of the null hypothesis on the experimental session. For the shortest epoch (250 ms), considerably fewer epochs of

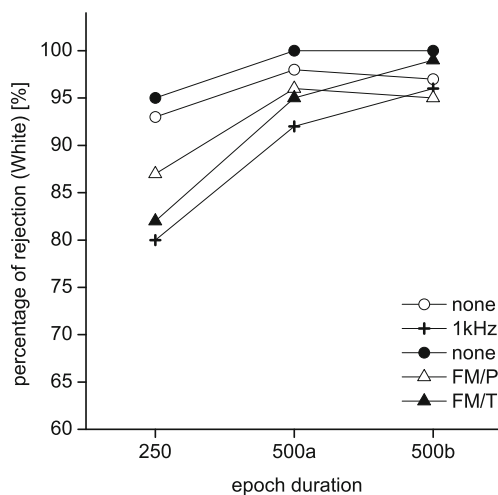


Fig. 5 The percentage of rejection of the null hypothesis of homoscedasticity of the White test for the signals recorded with the channels located above the auditory cortices. Large percentage values reflect heteroscedasticity, i.e. the variance within the epochs changes with time. Note the systematic dependence of heteroscedasticity on the experimental session, especially for the 250-ms epoch. See Sect. 4.1 for details on epoch duration

Table 2 Results for the variance-stationarity obtained from the White test

Channel	250 ms	500 ms (a)	500 ms (b)
Auditory	87	96	97
Reference	89	91	93

The table shows the percentage of rejection of the null hypothesis (homoscedasticity of epochs) averaged across all five stimulus sessions for the auditory and the reference channels (a) and (b) refer to different starting points of the 500-ms intervals (see Sect. 4.1)

the sessions with auditory stimulation were heteroscedastic compared with the epochs containing the subject's spontaneous brain activity only. This difference became less evident for the two 500-ms intervals. The percentages of rejection averaged across all five sessions are summarized in Table 2.

4.3 Testing for stationarity of the trial-to-trial variability of power spectra

The three aforementioned tests were also applied in search for information on the stochastic nature of the sequence of selected coefficients of the power spectra of single trials. Given the inherent limitation of the Fourier transform in its frequency resolution, which equalled 4 Hz ($1/250 \text{ ms}^{-1}$) and 2 Hz ($1/500 \text{ ms}^{-1}$), respectively, we found two pronounced peaks—at 8 Hz and at 12 Hz—in the FT spectra of the averaged 250-ms epochs, and three peaks—at 8, 10 and 12 Hz—in the spectra of the averaged 500-ms epochs. We performed time-series analysis of the trial-to-trial variability of the power spectra at those frequencies, having six series (corresponding to the number of selected MEG channels) for each type of stimulation for the 250-ms epochs, and, analogously, 12 series for the two kinds of epochs of 500-ms duration.

Similar to the results described in Sect. 4.1, the null hypothesis of the KPSS test was not rejected for the majority of the trials. We also found that, for the PP test, the percentage of acceptance of H_1 , i.e. of mean-stationarity, was not fully compatible with the results obtained with the corresponding KPSS-mean test. This inconsistency may reflect the different power of the individual tests. Since the KPSS test is a test of high power, we assume its results are more objective. There were also no reasons to reject H_0 of the White test.

The series of the Fourier coefficients of trial-to-trial variability of power spectra was, in the majority of the cases, stationary in the wide sense. This property seemed to be independent of the duration of the analysed epoch; however, it is important to note that the influence of the (poor) frequency resolution remains unknown. No significant differences between the results of the examined frequencies were observed.

5 Discussion

We have introduced three specific statistical tests developed in econometrics—the KPSS test (Kwiatkowski et al. 1992), the PP test (Phillips and Perron 1988) and the White test (White 1980)—to analyse the stationarity of encephalographic time series using an exemplary set of MEG data. Unlike the commonly used tests (Bendat and Piersol 1986), which test randomness or presence of a trend rather than stationarity, these tests were particularly designed to infer about some specific aspects of stationarity of time series.

Our study revealed that the majority of the investigated MEG signals were variance-nonstationary and mean-stationary. We would like to emphasize that the testing procedure applied in the KPSS test and the PP test verifies the hypothesis of the presence or the absence of a unit root under the assumption that there is at most one unit root. To exclude the possibility of the presence of multiple unit roots, it is advisable to consider the procedure proposed by Pantula (1989). Moreover, note that we have considered only a linear deterministic term (see Eq. 10) and, from an experimental point of view, trends might have reflected the presence of a slowly varying quasi-periodic component in the series.

In contrast to the findings of McEwen and Anderson (1975), Kawabata (1976) and Cohen and Sances (1977), who evaluated relatively long (of the order of seconds) epochs of EEG data as being stationary, the application of the aforementioned tests showed that the majority of trials even as short as 250 ms were nonstationary; to be more precise, variance-nonstationary (Fig. 5). Furthermore, our observation of a systematic increase of heteroscedasticity from the 250-ms epoch to the 500-ms epoch (Fig. 5) is in accordance with theoretical considerations, according to which stationarity is expected to disappear with increasing epoch duration. From a neuroscientific point of view, this result is conceivable, since longer time intervals may contain more complex features related to stimulus processing, cognitive processes, changes in spontaneous activity, etc. One possible way to model the heteroscedastic character of brain signals is to use the parametric stochastic models for variance regression, which use Kalman filtering, and especially state-space GARCH modelling. Examples of such an approach can be found in the literature (see Galka et al. 2004a; Wong et al. 2005; Galka et al. 2004b; Ozaki et al. 2006).

The presence and the kind of external sensory stimulation did not significantly affect the findings as to the stationarity of the analysed MEG time series. This may be due to the fact that the evoked activity is relatively weak with respect to the spontaneous (background) activity, and therefore does not significantly alter the stochastic properties of the acquired signal. We thus conclude that our findings may reflect some general characteristic properties of MEG (and most likely also of EEG) time series, regardless of whether those sig-

nals carry information on the background activity only or, in addition, on evoked brain responses too.

The application of the tests to the power spectra of the MEG data showed a weakly stationary character at several selected frequencies. Here, the results did not depend on the presence (or kind) of stimulation, either. This observation is similar to the results obtained by Kawabata (1976) or Cohen and Sances (1977), but our conclusions are somewhat different. These authors found the evoked brain signals stationary. Our results obtained from the White test showed that the single-trial brain signals were variance-nonstationary (Fig. 5), whereas time series composed from selected Fourier coefficients were both mean- and variance-stationary.

One possible explanation for why the variance-nonstationarity disappeared once the data were transformed into the frequency domain is the estimation method used in the analyses. Like other authors, we used the fast Fourier transform (FFT), which is designed to efficiently handle stationary data. The temporal resolution of the FFT is limited by the length of the time window to be transformed, since the FFT provides only one power spectrum for such a window. The frequency resolution is limited too, since the lowest observable frequency is given by the inverse of the length of the time window. Therefore, the intrinsic limitations of the FFT might have biased our results. This could potentially explain why Kawabata (1976) observed stationarity of EEG power spectra. We applied the same procedure (based on FFT), but the statistical tests used in our analysis were more accurate. The results are similar, but the conclusions drawn from the overall analysis are different. Kawabata (1976) claimed that short fragments of encephalographic time series (EEG) are stationary. We conclude that the calculation of power spectra using FFT is responsible for the observation of the stationarity of time series. One should keep in mind the properties of the Fourier transform, in which contributions of the components of the analysed signal to the resulting spectrum are integrated across the entire time range. Hence, in a stationary series of Fourier coefficients, they do not necessarily correspond to the same phenomenon. Therefore, there is a strong uncertainty as to how one might translate the statistical properties of such a series into the signal from which it was derived by, for example, computing the Fourier spectrum for each of the subsequent/consecutive short epochs of MEG/EEG.

Bearing the disadvantages of the aforementioned nonparametric estimation in mind, one might consider the parametric approach with autoregressive models, where the limitations in resolution, and hence the related uncertainty, do not occur. Unfortunately, however, AR models are suitable for stationary processes, and hence, in general, not suitable for the analysis of varying MEG/EEG spectra. Therefore, an alternative solution might be, in analogy to the spectral approach described above, to examine a series of trial/epoch-dependent amplitudes of Gabor functions—which have the

best properties in terms of the trade-off between time and frequency resolution—obtained from the multivariate adaptive approximation with the matching pursuit algorithm (Sielużycski et al. 2009a,b) instead of examining a series of Fourier coefficients. In contrast to the Fourier transform, the matching pursuit with stochastic dictionaries provides fits of Gabor functions to the local structures present in the signal (Durka 2007).

The combination of distinct tests for mean-, trend- and variance-stationarity and for the presence of unit roots allows a considerably refined analysis of the stationarity of MEG and EEG time series as well as of the corresponding spectra. The findings obtained from the application of these tests have two important consequences: (1) they question the applicability of the commonly used tests, which were not developed to test stationarity per se; (2) they support deciding whether or not the application of analysis methods, which specifically require stationarity in the first place, is justified. Finally, one should bear in mind that MEG or EEG time series have a highly individual, i.e. subject specific, character. Therefore, the application of the KPSS test and the PP test, which in the presented study revealed mean-stationarity, as well as the White test, which revealed variance-nonstationarity, to data acquired on a population of subjects will be required in order to draw more generally valid conclusions.

Acknowledgements We would like to thank Norman Zacharias for the realization of the MEG measurement.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Barlow JS (1985) Methods of analysis of nonstationary EEGs, with emphasis on segmentation techniques: a comparative review. *J Clin Neurophysiol* 2:267–304
- Bendat JS, Piersol AG (1966) Measurement and analysis of random data. Wiley, New York
- Bendat JS, Piersol AG (1986) Random data: analysis and measurement procedures. Wiley, New York
- Bender R, Schultz B, Schultz A, Pichlmayr I (1992) Testing the gaussianity of the human EEG during anesthesia. *Methods Inf Med* 31:56–59
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31:307–327
- Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden-Day, Oakland
- Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting. Springer, New York
- Cohen BA, Sances A (1977) Stationarity of the human electroencephalogram. *Med Biol Eng Comput* 15:513–518
- Dickey D, Fuller W (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74:427–431
- Durka PJ (2007) Matching pursuit and unification in EEG analysis. Engineering in medicine and biology. Artech House, Boston
- Galka A, Yamashita O, Ozaki T (2004a) Garch modelling of covariance in dynamical estimation of inverse solutions. *Phys Lett A* 333:261–268
- Galka A, Yamashita O, Ozaki T, Biscay R, Valdes-Sosa P (2004b) A solution to the dynamical inverse problem of EEG generation using spatiotemporal kalman filtering. *Neuroimage* 23:435–453
- Gasser T, Molinari L (1996) The analysis of the EEG. *Stat Methods Med Res* 5:67–99
- Gersch W (1970) Spectral analysis of EEGs by autoregressive decomposition of time series. *Math Biosci* 7:205–222
- Kaplan AY (1999) The problem of segmental description of human electroencephalogram. *Human Physiol* 25:107–114
- Kawabata N (1976) Test of statistical stability of the electroencephalogram. *Biol Cybern* 22:235–238
- König R, Sielużycski C, Simserides C, Heil P, Scheich H (2008) Effects of the task of categorizing FM direction on auditory evoked magnetic fields in the human auditory cortex. *Brain Res* 1220:102–117
- Krystal A, Prado R, West M (1999) New methods of time series analysis of non-stationary EEG data: Eigenstructure decomposition of time varying autoregressions. <http://citeseer.ist.psu.edu/krystal99new.html>
- Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *J Econom* 54:159–178
- McEwen JA, Anderson GB (1975) Modeling the stationarity and gaussianity of spontaneous electroencephalographic activity. *IEEE Trans Biomed Eng* 22:361–369
- Möcks J, Pham DT, Gasser T (1984) Testing for homogeneity of noisy signals evoked by repeated stimuli. *Ann Stat* 12:193–209
- Möcks J, Gasser T, Pham DT (1984) Variability of single visual evoked potentials, evaluated by two new statistical tests. *Electroencephalogr Clin Neurophysiol* 57:571–580
- Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24(4):375–425
- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708
- Ozaki T, Wong K, Galka A, John R (2006) Recent surge of time series methods in neuroscience—causal modeling and time-varying innovation variances. http://www.taniguchi.sci.waseda.ac.jp/Waseda_Workshop_2006/Proceedings/Ozaki/Ozaki_1.pdf
- Pantula SG (1989) Testing for unit roots in time series data. *Econom Theor* 5(2):256–271
- Perron P (1988) Trends and random walks in macroeconomic time series. *J Econom Dyn Control* 12:297–332
- Phillips PCB (1987) Time series regression with a unit root. *Econometrica* 55:277–301
- Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. *Biometrika* 75:335–346
- Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71:599–608
- Saikkonen P (1993) A note on a Lagrange multiplier test for testing an autoregressive unit root. *Econom Theor* 9:494–498
- Shumway RH, Stoffer DS (2000) Time series analysis and its applications. Springer, New York
- Sielużycski C, König R, Matysiak A, Kuś R, Durka PJ (2009a) Single-trial evoked brain responses modeled by multivariate matching pursuit. *IEEE Trans Biomed Eng* 56(1):74–82
- Sielużycski C, Kuś R, Matysiak A, Durka PJ, König R (2009b) Multivariate matching pursuit in the analysis of single-trial latency of the auditory M100 acquired with MEG. *Int J Bioelectromagn* 11(4):155–160

- Tanaka K (1983) Non-normality of the Lagrange multiplier statistic for testing the constancy of regression coefficients. *Econometrica* 51:1577–1582
- Trapletti A, Hornik K (2011) tseries: Time series analysis and computational finance. R package version 0.10-25. <http://CRAN.R-project.org/package=tseries>
- Weiss M (2005) Testing correlated EEG-like data for normality using a modified Kolmogorov–Smirnov statistic. *J Econom* 128: 195–213
- White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for homoskedasticity. *Econometrica* 48:817–838
- Wong KF, Galka A, Yamashita O, Ozaki T (2005) Modelling non-stationary variance in EEG time series by state space garch model. *Comput Biol Med* 36:1327–1335
- Wright JJ, Kydd RR, Sergejew AA (1990) Autoregression models of EEG—results compared with expectations for a multilinear near-equilibrium biophysical process. *Biol Cybern* 62: 201–210