ORIGINAL RESEARCH

MODELLING THE EEG BASED EVENT-RELATED BRAINWAVES USING STATISTICAL TIMES SERIES

Ömer Utku Erzengin[1], Vildan Sümbüloğlu[2], Sirel Karakaş[3]

[1]*Marmara Üniversitesi, Biyoistatistik ABD, İstanbul, Türkiye* [2]*Zonguldak Karaelmas Üniversitesi, Biyoistatistik ABD, Zonguldak, Türkiye* [3]*Hacettepe Üniversitesi, Deneysel Psikoloji Uzmanlık Alanı, Ankara, Türkiye*

ABSTRACT

Objective: The aim of this study was to test whether the alpha oscillation in the auditory event-related activity of the brain can be explained using an Autoregressive Integrated Moving Average (ARIMA) model.

Methods: The sample consisted of 17 (11 males, 6 females), right-handed, healthy, volunteer participants. Event-related activity was recorded under the omitted stimulus paradigm. Alpha response was obtained through digital filtering. The alpha response of each participant was separately analyzed.

Results: The ARIMA(0,2,1) model was found to be applicable to alpha response that was evoked by stimulus omission. ARIMA (0,2,1) shows that this activity is short-lasting and it fluctuates stochastically.

Conclusion: The present study showed that the alpha responses of single participants obtained under the omitted stimulus paradigm were best modeled by the ARIMA (0, 2, 1) type of time series model.

Keywords: Event-related potential, ERP, Digital filter, Alpha response, Oscillation, ARIMA

EEG TEMELLİ OLAY-İLİŞKİ BEYİN DALGALARININ İSTATİSTİKSEL ZAMAN SERİLERİ İLE MODELLENMESİ

ÖZET

Amaç: Çalışmanın amacı beyinin olay ilişki aktivesinin alfa osilasyonlarının Otoregresif Bütünleşik Hareketli Ortalama - ARIMA – modeli ile açıklamaktır.

Yöntem: Çalışan örneklem 17 (11 erkek, 6 kadın) sağ elini kullanan, sağlıklı gönüllülerden oluşmuştur. Olay ilişkili aktivite, atlanmış uyarıcı paradigması altında kaydedilmiştir. Alfa tepkisi sayısal filtreyle elde edilmiştir. Her katılımcının alfa tepksi ayrı ayrı analiz edilmiştir.

Bulgular: Yapılan çalışma göstermektedir ki her bir katılımcının atlanmış uyarıcı paradigması için alfa tepkisi, ARIMA (0,2,1) tipindeki zaman serisiyle en iyi modellenebilmektedir.

Sonuç: Yapılan çalışma atlanmış uyarıcı paradigması altında elde edilmiş, bir katılımcının alfa tepkilerinin en iyi ARIMA (0, 2, 1) zaman serisiyle modellendiğini göstermiştir.

Anahtar Kelimeler: Olay ilişki potansiyel, Sayısal filtre, Alfa tepkisi, Salınım, Otoregresif bütünleşik hareketli ortalama

Corresponding author:
Ömer Utku Erzengin, M.D.
Marmara Üniversitesi, Biyoistatistik ABD, İstanbul, Türkiye
e-mail: ouerzengin@marmara.edu.tr

Marmara Medical Journal 2006;19(1);6-12
Ömer Utku Erzengin, et al
Modelling the EEG based event-related brainwaves using statistical times series

INTRODUCTION

Event-related potentials (ERPs) are time-varying waveforms and they represent the neuroelectric activity of the brain in response to sensory stimulation or to execution of a motor, cognitive, or psychophysiological task. ERPs are time-locked to events. This event may be a sensory stimulus (such as a visual flash or an auditory sound), a mental event (such as recognition of a specified target stimulus), or the omission of a stimulus (expectancy of a stimulus which however is omitted).

The spontaneous activity of the brain, the electroencephalogram (EEG), and the ERPs are continuous over time. Accordingly, they are analog signals. Recording the analog electrical activity into a computerized environment is called digital transformation or digitalization. In digital transformation, electrical activity is recorded in equally spaced time intervals.

Any analogue or digital time domain activity can be represented by sinusoids. Frequency analysis involves the conversion of the data in the time-domain to frequency domain. A widely used technique of frequency analysis, is the fast Fourier Transform (FFT) (Jean Baptiste Joseph Fourier, 1807). The purpose of FFT is to decompose a complex time series with cyclical components into a few underlying sinusoidal (sine and cosine) functions of particular wavelengths (oscillations). The transformation of the data from the time-domain to the frequency-domain finds the frequency components whose superposition produce the time-domain waveform[1]. FFT thus demonstrates that the time-domain waveform is represented with frequency components, which are called event-related oscillations (EROs). In this study one of the brains oscillations called alpha activity is formulated with statistical time series.

The advantage of FFT is the simultaneous demonstration of the frequency components that make up the time-domain waveform[2-8]. A new and/or pioneering method of analysis involves the simultaneous description of the frequency components in the time-frequency domain. Among these are the various forms of wavelet transform, wavelet entropy, Wigner transform, Cohen's class of distributions, short-time Fourier transform and time-frequency component analyzer[9-18]. These show the progression of the oscillations over the time axis[9,19].

When FFT is used, the progression of the oscillations over the time axis is made possible through the application of digital filters. The cut-off frequencies of these filters are obtained in a response-adaptive way since they are determined from the limits of the resonant selectivities in the amplitude characteristics that FFT provides[1,19,20]. Digital filtering uses a function for extracting the sinusoidal components from the analog time-domain waveform. In signal processing, also, filters are used for extracting some parts of the signal, such as those components that occur within a certain frequency range. They are also used for removing unwanted parts of the signal, such as random noise.

Statistical techniques can also be used for studying any time varying activity. According to statistics, a time series is composed of successively appearing data points which are typically spaced apart in uniform time intervals. The main goal of time series analysis is to identify and understand the nature of time-dependent events. There are several statistical techniques whereby a time series can be modeled. One of the most studied statistical time series models is that of Box and Jenkins[21]. This Autoregressive Integrated Moving Average (ARIMA) model employs three sequential processes: the autoregressive (AR) process, integrated (I) process, and moving average (MA) process. The technique is hence called the ARIMA model, a synthetic term that combines the names of the specified subprocesses. The Box and Jenkins model requires stationary time series data for identifying and analysing, which means no change at mean and variance of series according to time.

Autoregressive process (AR) is used on a time series where observations are serially dependent. In such data, one can estimate a coefficient or a set of coefficients that describe consecutive elements of a series from the specific, time-lagged previous elements. Integration is a parameter, which is computed for the series after it is differenced once. Each integration order corresponds to differencing the series being forecast. A first-order integrated component means that the forecasting model is designed for the first difference of the original series. A second- order component corresponds to using second differences. An integration parameter is used for obtaining stationary time series. An integration parameter that equals 1 denotes that the time series is increasing linearly, one that

Marmara Medical Journal 2006;19(1);6-12
Ömer Utku Erzengin, et al
Modelling the EEG based event-related brainwaves using statistical times series

equals to 2 denotes that the time series is fluctuating over time. The moving average process (MA) is independent from the autoregressive process. Each observation in a time series can also be affected by an error or errors in the past. Each observation can thus be estimated from the previous errors.

The brain electricity is a time–varying activity. The aim of this study was to test whether the alpha oscillation (8-12 Hz) which is obtained through the digital filtering of the ERPs that are recorded under the omitted stimulus paradigm can be formulated through ARIMA models.

METHODS

Participants

The sample consisted of 17 (11 males, 6 females), right-handed, voluntary participants. Mean age of participants was 22.5 and the standard deviation was 2.3 (min.: 20, max.: 27). The participants were reportedly healthy. Hearing level, which was investigated using computerized audiometric testing, was found to be within normal limits. One of the participants was called ALGU in short (nickname) and ALGU's data was used as a sample.

Electrophysiological Procedures

Electrophysiological responses were obtained under the repetitive, omitted stimulus paradigm[22]. Auditory stimuli were presented over the ear-phones. The frequency of the auditory stimuli was 1000 Hz. The amplitude and duration of the auditory stimuli were 65 dB and 50 msec. Stimuli were presented at the rate of 2/sec. Stimulation consisted of blocks of five stimuli; in each block, four consecutive stimuli were presented and the fifth was omitted. The task of the participant was to decide on the time of occurrence of the omitted stimulus.

Electroencephalographic (EEG) activity was recorded over the Fz recording site (ref: linked earlobes; ground: forehead) of the 10-20 system under eyes-open condition. Bipolar recordings were made of electrooculogram and electromyogram for monitoring eye movement artifacts and muscle artifacts, respectively. Epochs with artifacts were discarded through online rejection (of responses whose amplitudes exceed ±50 mV) and offline rejection (through visual inspection). The EEG was amplified and filtered with a bandpass between 0.16 - 70 Hz (3 dB down, 12 dB /octave). It was recorded with a sampling rate of 512 Hz and a total recording time of 2000 ms. The pre- and post-stimulus interval were each 1000 msec. The single sweeps of each participant (range: 12 - 29 sweeps for each participant) were digitally band-pass filtered in the alpha band range (8-12 Hz)[22].

RESULTS

Calculations for digital filtering were performed under Octave for Linux (GNU license), and statistical analysis was made under R for Linux (GNU license). Graphics were prepared using SPSS 11.5 (licensed for Marmara University, Department of Biostatistics).

Model Fitting Procedures

The time series analysis method of Box and Jenkins involves special procedures. First ofall, the time series must be converged into a stationary form according to time. Afterwards, the converted time series are modeled with the help of autocorrelation function and partial autocorrelation function.

Autocorrelation function (ACF) and partial autocorrelation function (PACF) are the major tools in the Box and Jenkins method. ARIMA process is thus determined by ACF and PACF graphics. Autocorrelation is a kind of correlation coefficient. However, instead of a correlation between two different variables, the correlation is between two values of the same variable at consecutive time lags (e.g. Xt and Xt+k etc.). The partial autocorrelation removes the effect of shorter autocorrelation lags from the correlation estimate at longer lags. The partial autocorrelation at lag k is the autocorrelation between Xt and Xt-k that is not accounted for by lags 1 through k-1.

The first stage of the analysis involved the calculation of unfiltered, single sweep ERPs and the calculation of filtered ERPs (f-ERP) that show the alpha response. In the next stage, the autocorrelations and partial correlations were calculated for single sweeps. The autocorrelation and partial correlation graphics are shown in Figure 1.

The ACF and PACF of each participant showed that the function represented by the sweeps did not have a stationary time series trend. Thus, in the third stage, the time series model that best fits these sweeps was investigated. To this end, autoregressive, moving average, seasonal and

Marmara Medical Journal 2006;19(1);6-12
Ömer Utku Erzengin, et al
Modelling the EEG based event-related brainwaves using statistical times series

exponential smoothing models were tested. None of these models sufficiently explained the ACF and PACF. Thus, in the fourth stage, first-order difference was calculated whereafter the ACF and PACF were recalculated.

Figure 2 shows that, in spite of the first-order difference manipulations, a stationary time series trend could still not be achieved. At the fifth stage, a second-order difference was calculated whereafter the ACF and PACF were again recalculated.

Figure 3 shows that, by taking the second-order difference, the time series gained a stationary character.

This manipulation involves the usage of the Autoregressive Integrated Moving Average Model (ARIMA). The parameters of the ARIMA were (0,2,1). These values indicate, respectively, that the data did not show an autoregressive structure (AR: 0); that is, it fluctuated (I: 2); and the data were described by only one moving average coefficient for the residual (MA: 1) that varied between 0.60 and 0.88. Statistically, the coefficients of the moving average for the residual should be between 0.5 and 0.9.

The integration parameter I is equal to 2, that means there are stochastic fluctuations in the time series. Linear exponential smoothing models (stochastic fluctuations) are ARIMA models, which use two nonseasonal differences in conjunction with MA terms. The second difference of a series Z_t is not simply the difference between Z_t and itself lagged by two periods, but rather it is the first difference of the first difference. A second difference of a discrete function is analogous to a second derivative of a continuous function; it measures the "acceleration" or "curvature" in the function at a given point in time.

The ARIMA (0,2,1) model calculated for the participant, "ALGU", could be expressed in the form of the following mathematical equation:

$$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6012a_{t-1} + a_t$$

where Z_t is last observed time series value, Z_{t-1} is only one lag before observed time series value, Z_{t-2} is two lag before observed time series value, a_{t-1} is a residual term of Z_{t-2} and Z_{t-1}, and at is a residual term of Z_{t-1} and Z_t. Moreover the ARIMA (0,2,1) coefficient (0.6012) is significant at the $p<0.01$ level. The calculated coefficients for the data of the present study were, accordingly, optimally obtainable values (Table I).
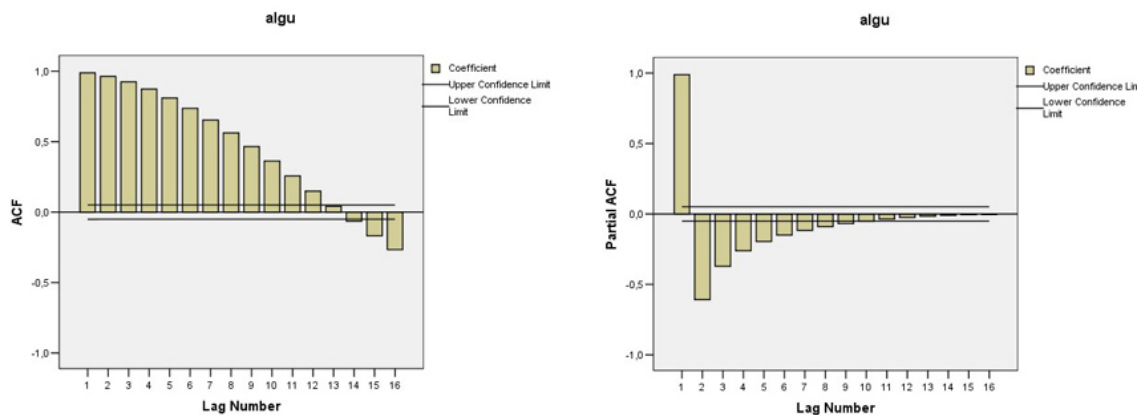

Fig. 1: The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the alpha response (8-12 Hz) of a typical participant that was obtained in response to omitted stimuli.
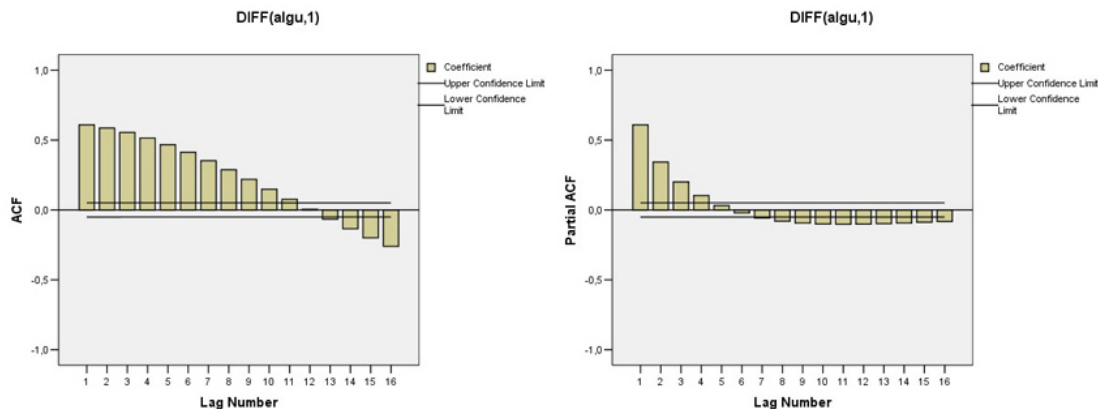

Fig. 2: First-Order Difference of ACF and PACF for the alpha response (8-12 Hz) of a typical participant that was obtained in response to the omitted stimuli.

Marmara Medical Journal 2006;19(1);6-12
Ömer Utku Erzengin, et al
Modelling the EEG based event-related brainwaves using statistical times series
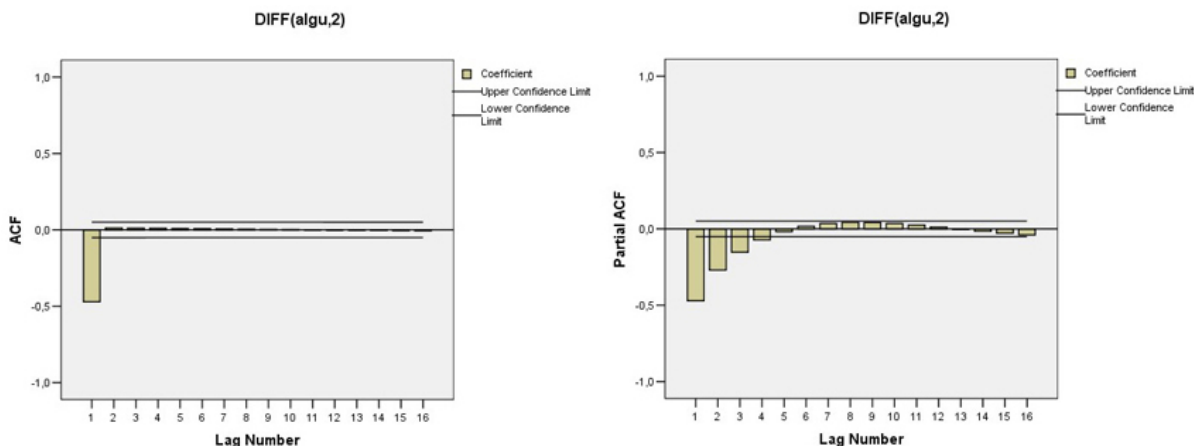
Fig. 3: Second-Order Difference of ACF and PACF for the alpha response (8-12 Hz) of a typical participant that was obtained in response to the omitted stimuli.

Table I: Time Series Equations for the Responses of Participants Obtained under the Omitted Stimulus Paradigm and the Portmanteau Statistics of the ARIMA (0,2,1) Models for Each Participant.

| Participants	ARIMA (0,2,1) Model	T-Ratio	p-Value	Q-Statistics
ALGU	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6012a_{t-1} + a_t$	29.566	0.00000	P = 0.28
AYBA	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.7264 a_{t-1} + a_t$	47.934	0.00000	P = 0.65
BIES	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.7045 a_{t-1} + a_t$	58.528	0.00000	P = 0.60
EMYI	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6876a_{t-1} + a_t$	56.865	0.00000	P = 0.29
FEBE	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6776a_{t-1} + a_t$	47.927	0.00000	P = 0.81
GOUS	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6522a_{t-1} + a_t$	47.834	0.00000	P = 0.50
HAKA	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6822a_{t-1} + a_t$	56.042	0.00000	P = 0.44
HIGU	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.7129a_{t-1} + a_t$	57.682	0.00000	P = 0.46
MUKU	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6853a_{t-1} + a_t$	51.217	0.00000	P = 0.71
MUSU	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6762a_{t-1} + a_t$	52.081	0.00000	P = 0.66
NASA	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6903a_{t-1} + a_t$	48.409	0.00000	P = 0.40
ORGU	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6132a_{t-1} + a_t$	44.892	0.00000	P = 0.33
OZBE	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.8820a_{t-1} + a_t$	109.94	0.00000	P = 0.28
OZOZ	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.7461a_{t-1} + a_t$	96.761	0.00000	P = 0.65
TATA	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6102a_{t-1} + a_t$	47.041	0.00000	P = 0.60
TAUL	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6777a_{t-1} + a_t$	54.347	0.00000	P = 0.29
YEUN	$Z_t = 2Z_{t-1} - Z_{t-2} - 0.6237a_{t-1} + a_t$	47.927	0.00000	P = 0.81
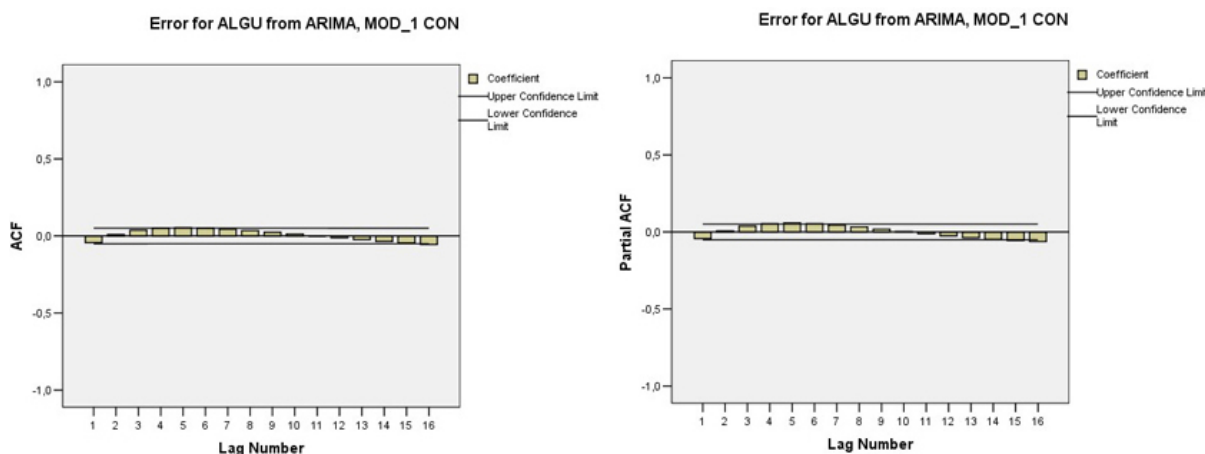
Fig. 4: Left panel: The difference between the actual data and the data simulated by the ARIMA (0, 2, 1) model: The error of prediction. Right panel: The autocorrelation and the partial autocorrelation of the error term of the ARIMA (0, 2, 1) model were found to be within the confidence interval limits. This shows that the model does not produce any systematic error.

Marmara Medical Journal 2006;19(1);6-12
Ömer Utku Erzengin, et al
Modelling the EEG based event-related brainwaves using statistical times series

Model Adequacy

Statistical modeling techniques such as those used in analyzing time series require model checking. Model checking is a test of the model adequacy.

In the time series methodology of Box and Jenkins, model adequacy is checked by analyzing the residuals of time series. Residuals represent the difference between the values that areestimated from the model and those observed in the time series. The residuals have to be in the form of white noise. Furthermore, the values for the autocorrelation function and the partial autocorrelation function have to be nonsignificant at all lags. This would show that the residual (white noise) is a stationary time series or a stationary random process with zero autocorrelation and zero partial autocorrelation. This procedure is realized using Portmanteau test statistics. A portmanteau (most of the time called Q-Statistics) test is a test for serial correlation up to the specified lag (order). In the present study, the graphic representation of the systematic residual in Fig. 4 was calculated mathematically with also the Portmanteau Statistics[21].

The null hypothesis of Portmanteau test statistics is that all autocorrelation values are equal to zero. A p-value smaller than 0.05 justifies the rejection of the null hypothesis and shows that the time series is not in the form of white noise. According to Table I, the p-values are larger than 0.05, this shows that the difference between observed values and those estimated from the ARIMA model are nonsignificant. Accordingly, the residuals in the present data are in the form of white noise.

DISCUSSION

ARIMA (0,2,1) models describe a short-term process that ends rapidly. It is not possible to estimate the future processes with an ARIMA (0,2,1) model; such a model adequately describes only the specific time interval for which the model is designed. The present study showed that, among the time series models, ARIMA (0, 2, 1) best fitted the omitted stimulus alpha response of single participants. As is pointed out below, this finding is in line with the phenomenology of the omitted stimulus response[22].

After fitting a time series model, the model adequacy was checked with Portmanteau test statistics. Portmanteau test statistics show that after fitting the model, residuals for alpha activity oscillations were nonsignificant.

• In the ARIMA model, the coefficient for AR was "0"; this shows that the data does not show an autoregressive structure. According to Box and Jenkins, ARIMA(0,2,1) fits short-lasting data of sudden appearance[21]. The event-related alpha oscillation in fact terminates right after the decision of the participant with respect to the temporal point at which the omission occurred.

• The coefficient for I was "2". This shows that the data has a fluctuating character. The event-related alpha response is a fluctuating brain signal.

• The coefficient for MA was "1". This indicates the existence of a stochastic process that can be described by a weighted sum of white noise error and the white noise error in the preceding periods. In the digitally filtered alpha response, each data point on the alpha waveform is determined with the previous ones. Furthermore, brain neuroelectricity is a complex phenomenon and no recording is devoid of methodological and technical contamination, and the contamination of brain activity that unrelated to that under experimental investigation.

The foregoing coefficients of the ARIMA model fits in with the electrophysiological representation of expectancy in humans. The process of expecting an event terminates with its appearance (thus, AR=0) and the alpha response, the best electrophysiological representation of expectancy, characteristically oscillates within the 8-12 Hz frequency range (thus, I=2). Finally, mankind is not a perfect machine; its responses are not strictly precise and not 100 % reliable; thus "error" of recording is to be expected in human responses (thus MA=1). These points show the statistical ARIMA model can represent human electrophysiological responses in cognitive processing.

REFERENCES

1. Karakaş S, Erzengin ÖU, Başar E. A new strategy involving multiple cognitive paradigms demonstrates that ERP components are determined by the superposition of oscillatory responses. Clin Neurophysiol 2000; 111: 1719-1732
2. Başar E. EEG-Brain Dynamics: Relation between EEG and Brain Evoked-Potentials. Amsterdam: Elsevier Publication,1980.
3. Brandt ME, Jansen BH. The relationship between prestimulus alpha amplitude and visual evoked potential amplitude. Int J Neurosci 1991; 61: 261-8.
4. Jervis BW, Nichols MJ, Johnson TE, Allen, Hudson NR. A fundamental investigation of the composition of

auditory evoked potentials. IEEE Trans Biomed Eng 1983; 30: 43-49.

5. Kolev V, Yordanova J. Analysis of phase-locking is informative for studying event-related EEG activity. Biol Cybern 1997; 96: 229-235.

6. Parvin C, Torres F, Johnson E. Synchronization of single evoked response components: Estimation and interrelation of reproducibility measures. In Rhythmic EEG Activities and Cortical Functioning. Amsterdam: Elsevier, 1980.

7. Röschke J, Mann K, Riemann D, Frank C, Fell J. Sequential analysis of the brain's transfer properties during consecutive REM periods. Electroencephalogr Clin Neurophysiol 1995; 96: 390-397.

8. Solodovnikov VV. Introduction to the Statistical Dynamics of Automatic Control Systems. New York: Dover Publications 1960.

9. Başar E. Brain Function and Oscillations I: Brain Oscillations. Principles and Approaches Heidelberg: Springer-Verlag 1998.

10. Başar E. Brain Function and Oscillations II: Integrative Brain Function. Neurophysiology and Cognitive Processes. Heidelberg: Springer-Verlag 1999.

11. Cohen L. Time-frequency distributions: A review. Proc IEEE 1989; 77 (7): 941-81.

12. Demiralp T, Ademoğlu A, Schürmann M, Başar E. Wavelet analysis of brain waves. In: Başar E, editor. Brain Function and Oscillations: I. Brain Oscillations. Principles and Approaches. Heidelberg: Springer-Verlag 1998.

13. Demiralp T, Ademoğlu A, Schürmann M, Başar-Eroğlu C, Başar E. Detection of P300 in single trials by the wavelet transform (WT). Brain Lang 1999; 66: 108-128.

14. Özdemir AK., Karakaş S, Çakmak ED, Tüfekçi Dİ., Arıkan O. Time-frequency component analyser and its application to brain oscillatory activity. Journal of Neuroscience Methods 2005; 145: 107-125.

15. Rosso OA, Blanco S, Yordanova J, Kolev V, Figliola A, Schürmann M, Başar E. Wavelet entropy: A new tool for analysis of short-duration brain electrical signals. J Neurosci Methods 2001; 105: 65-75.

16. Samar VJ, Bobardikar A, Rao R, Swartz K. Wavelet analysis of neuroelectric waveforms: A conceptual tutorial. Brain and Lang 1999; 66: 7-60.

17. Tağluk, M.E., Çakmak, E.D., Karakaş, S. Analysis of time-varying energy of brain responses to an oddball paradigm using short-term smoothed Wigner-Ville distribution. Journal of Neuroscience Methods 2005; 143: 197-208.

18. Yordonova J, Kolev V, Rosso OA, et al. Wavelet entropy analysis of event-related potentials indicates modality of independent theta dominance. J Neurosci Methods 2002; 117: 99-109.

19. Farwell LA, Martinerie JM, Bashore TR, Rapp PE, Goddard PH. Optimal digital filters for long-latency components of the event-related brain potential. Psychophysiology 1993; 30: 306-315.

20. Cook III. EW, Miller GA. Digital filtering: Background and tutorial for psychophysiologists. Psychophysiology 1992; 29: 350-367.

21. Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis, Forecasting and Control, 3rd ed. NJ: Prentice Hall 1994.

22. Demiralp T, Başar E. Theta rhytmicities following expected visual and auditory targets. Int J Psychophysiol 1992; 13: 147-160