How Does the Toolbox Choice Affect ERP Analysis?

Andrés Quintero-Zea[1](✉) , Mónica Rodríguez[1], María Isabel Cano[1],
Karen Melisa Pava[1], Manuela Suaza[1], Natalia Trujillo[2,3],
and José David López[1]

[1] SISTEMIC, Engineering Faculty, Universidad de Antioquia UDEA,
Calle 70, No. 52-21, Medellín, Colombia
andres.quintero@udea.edu.co
[2] GISAME, Facultad Nacional de Salud Pública, Universidad de Antioquia UDEA,
Calle 62 No. 52-59, Medellín, Colombia
[3] Neuroscience Group, Universidad de Antioquia UDEA,
Calle 62 No. 52-59, Medellín, Colombia

Abstract. Event-related potentials (ERP) help understanding neural activity related to both sensory and cognitive processes. But due to their low SNR, EEG signals must be processed to obtain the ERP waveform. Such a processing can be carried using a number of toolboxes that may provide different results on further analyses. Here, we present an experimental design that quantitatively evaluates the effect of choosing a particular toolbox in the further ERP analysis. We select three widely used toolboxes: EEGLAB, SPM12, and Fieldtrip to process EEG data acquired from a Flanker-like task with a Biosemi Active-Two device. Results show that although there is not a significant difference between ERP obtained from each toolbox, the choice of a specific toolbox may have subtle effects in the resulting ERP waveforms.

Keywords: EEG · ERP · Flanker task · Toolbox · EEGLAB
SPM12 · Fieldtrip

1 Introduction

Electroencephalography (EEG) is a noninvasive technique that records the electrical activity of the brain [1]. EEG provides an excellent medium to understand cognition and brain function. Furthermore, time-locked EEG activity or event-related potential (ERP) allow researchers to analyze human brain activity associated with presentation of specific stimuli [2].

ERP records neural responses of task-related events with high temporal resolution, and constitutes a convenient method to explore dynamics of brain behavior. ERP technique has been used for decades to answer questions about sensory,

---

A. Quintero-Zea and M. Rodríguez—These authors contributed equally to the work.

cognitive, motor, and emotion-related processes in clinical disorders such as mild cognitive impairment [3], dementia [4], or emotional processing [5,6].

An ERP waveform is described according to latency and amplitude [2], and it can be split in two categories: the early exogenous components peaking within the first 100 ms after stimulus; and the later endogenous components that reflect how the subject evaluates the stimulus. Among these later components is the N200 (or N2), which is associated with conflict detection. N2 is the negative deflection peaking at about 200 ms after stimulus presentation. It is evoked during tasks in which two or more incompatible response tendencies are activated simultaneously, such as go/no-go or Flanker tasks [5].

To obtain the ERP waveform, the recorded EEG activity has to be processed off-line by means of a computational program. There are a number of freely available software packages, commonly known as toolboxes, to analyze these data. Despite the large number of studies investigating brain function by means of ERP, there is not a systematic effort to examine how different software packages can affect the findings and their associated neurophysiological interpretation. It is reasonable to expect that the results obtained with different toolboxes may differ, but it would be necessary a proper statistical analysis to determine if they are significant.

The aim of this paper is to evaluate the impact of the toolbox choice on the ERP components. To this end, we selected three widely used toolboxes to obtain the ERP waveforms for a Flanker task: EEGLAB [7], SPM12 [8,9], and Fieldtrip [10]. Further, we applied a repeated measures experimental design on the N2 component latencies and peak amplitudes to quantitatively evaluate differences among them.

2   Experimental Data

EEG data were acquired by the mental health (GISAME) research group of Universidad de Antioquia (Medellín, Colombia). Participants were 20 adults with mean age of 35 years and standard deviation of 9 years. This convenience sample was 100% Colombian, and included 15 men and five women. Volunteers that informed having psychiatric and neurological disorders were excluded from the study. All subjects participated voluntarily and signed an informed consent in agreement with the Helsinki declaration. The research protocol was approved by ethical committee of University of Antioquia (Medellín, Colombia).

EEG registers were acquired with a 64-electrode BIOSEMI EEG ActiveTwo system [11] at a sampling rate of 2048 Hz and 24-bit resolution. The electrodes were placed according to the international 10–20 system [12].

ERPs were recorded using a Flanker task [13]. Participants were seated in a comfortable chair in front of a computer monitor at a distance of 60 cm. Participants were asked to try not to blink, move, nor speak while performing the task. The impedances were maintained below 10 kΩ to obtain an adequate conductivity between scalp and electrodes.

The Flanker attentional emotional task involves violent and neutral situations. The stimuli were 60 real violent images, 60 neutral real images and as distraction 60 drawings of animate and inanimate objects. The participants were specifically instructed that the monitor would screen a central stimulus that could contain either a real picture or a black and white drawing. When a real image appeared on the center should be classified between a violent or neutral image, or if a drawing appeared in the center position should be classified as animate or inanimate. Four events result per each stimuli: threatening periphery (TP), threatening center (TC), neutral periphery (NP), and neutral center (NC) depending on the position of the real images. Further details about the experiment protocol have been previously described in [13].

3   Methods

We selected three toolboxes to process the Flanker-task related EEG data: EEGLAB[1] [7], FieldTrip[2] [10], and SPM12[3] [8,9]. We have performed a similar data processing with each toolbox and obtained their respective ERPs. The differences on data processing obey to actual differences on the software packages, as they not always offer the same methods for specific stages. Once we obtain the ERPs with each software, we perform two analyses aiming to find variations on typical ERP parameters: latency and peak intensity of N2, and an ANOVA test over the same component.

3.1   Data Processing

The stages of data processing are not standard but procedures do not largely vary from those depicted in [14,15]. Most variations on these stages are due to specific requirements of the task or due to the acquisition device characteristics. For the task and device used for experimentation, we perform off-line the following stages: Downsampling, filtering, bad channels rejection, re-referencing, artifact rejection, epoching, baseline correction, and visual noise rejection. These stages are presented below:

Downsampling: The high temporal resolution of most EEG devices is not necessary for most studies, but it may cause high computational burden; then, most authors reduce the frequency sample to around 200–500 Hz. The term downsampling refers to the process of reducing the sampling rate of a signal. In this case from 2048 Hz to 500 Hz. This stage was equally implemented in the three toolboxes.

---

[1] http://sccn.ucsd.edu/eeglab/.
[2] http://fieldtrip.fcdonders.nl/.
[3] http://www.fil.ion.ucl.ac.uk/spm/.

Filtering: To reduce environmental artifacts (such as power line noise) in the EEG data and to extract specific frequency bands associated with human cognition, it is necessary to filter the signals. A band-pass IIR digital filter was applied in all toolboxes. The cutoff frequencies were 0.5 and 30 Hz to elicit the typical frequency band of interest for ERP studies.

Bad Channels Rejection: Once the signal is filtered, it is desired to remove and interpolate those channels with low recording SNR. For EEGLAB case, bad channels are detected using `findNoisyChannels()` function from PREP pipeline library [16] and further interpolated using the spherical interpolation function `eeg_interp()`. In FieldTrip, this step is performed with `ft_channelrepair()` function, which finds the time series of the missing or damaged channel using a weighted average of its neighbors. SPM12 only offers the option of setting a channel as bad without interpolation for further processing. For such a reason, interpolation was not performed in SPM12.

Off-line Re-referencing: Next step consists on re-referencing the signals using a common reference for all channels. In this case, EEG data was re-referenced to the average of all electrodes.

Artifact Rejection: EEG signals are known to be contaminated with noise artifacts. The most common physiological artifacts are perhaps those generated by muscles. This includes eyes blinking, eye movements (EOG), muscular contractions (EMG), cardiac signals (ECG), and pulsations [17]. In addition, breathing and body movement can cause alterations in EEG signals. There are additional artifacts caused by the skin-electrode connection; if there are deformities, such as scars, they can change the impedance. Each toolbox offers different algorithms to reduce the impact of artifacts.

In SPM12, we used visual artifact rejection `spm_eeg_ft_artefact_visual()`. This tool is a FieldTrip function which is included in the SPM12 toolbox. This function allows browsing through the large amount of data in a MATLAB GUI by showing a summary of all channels and trials. The user visually identifies the trials or data segments that are contaminated, and selects those to be removed from the data.

In EEGLAB and FieldTrip, we used the Independent Component Analysis (ICA) [18] algorithm for artifact rejection. This methodology is widely used in EEG because ICA allows decomposing the signals into different independent components (in terms of variance). Some of these components are expected to be sources of artifacts. In this case, all components are presented as images to the user, whose must manually remove those considered as noise based.

Epoching, Baseline Correction and Visual Noise Rejection: After identifying and removing artifacts, the registers are segmented from 200 ms and 800 ms prior and after the stimulus, respectively. Each type of stimulus described in Sect. 2 is known as condition and constitutes a kind of epoch or trial. In our experimental design, we have four conditions leading to four epoched data types: threatening

periphery (TP), threatening center (TC), neutral periphery (NP), and neutral center (NC).

Once epoched, we are able to remove very low frequency noise that may affect the zero level among trials. Trials are then baseline corrected by determining the trend of the baseline before the stimulus (time window -200 to $0\,\mathrm{ms}$, being $0\,\mathrm{ms}$ the stimulus trigger time), and then removing this trend of the rest of the window (0 to $800\,\mathrm{ms}$). Each trial is inspected for leftover noise to make sure that only clean segments go forward for later analysis. Finally, epoched averaged data per condition of all participants are combined in a 3D matrix (channels × time points × trials) which forms the basis for all further ERP analysis.

3.2   Data Analysis

To evaluate differences in ERPs, we focused data analysis within a time window of 180 to 240 ms to obtain the peak amplitude and latency of the N2 component. This time window is adopted after a visual inspection of the grand averages, and it is similar to those reported in previous studies (e.g. [19]). Only two electrodes (F3, PO3) are used for the successive statistical analysis.

The statistical analysis consists on a one-way repeated-measures ANOVA aimed to compare variations due to the toolbox used on amplitude and latency of the ERP-N2 component. This analysis is performed using the Statistical Package for Social Sciences (IBM SPSS version 23.0 for Windows).

4   Results

In this section, we present differences on peak amplitudes and latencies of N2 component between the waveforms obtained with the three toolboxes. From the topographic maps shown in Fig. 1 maps seem roughly similar. Such likelihood is not present in the map obtained with EEGLAB for the threatening center stimulus. Smaller differences were observed between FieldTrip and SPM12.

The grand average ERPs at selected electrodes are depicted in Figs. 2 and 3. Note that waveforms obtained from EEGLAB exhibited lower N2 peak amplitudes, being more notorious at the central condition. Figure 2 also shows visual differences on the EEGLAB ERP in the late positive potential (LPP) component for latencies above 300 ms. This difference is extended to the three toolboxes in the central condition of Fig. 3. Although these differences in LPP are not part of the window of interest and are consistent among conditions and sensors (i.e., they should not affect posterior analysis within single software), these results demonstrate that there exist confidence issues for performing analyses on this window. No latency variations are observable.

Descriptive statistics for peak amplitudes and latencies are summarized in Table 1. In terms of amplitude, a consistent trend is observed: in all cases EEGLAB presented the lower amplitudes, followed by FieldTrip and then by SPM12. However, the variance was close among toolboxes and in all cases larger than mean variations. PO3 presented larger variance than F3, which is expected as the sensor is farther from the source of neural activity.
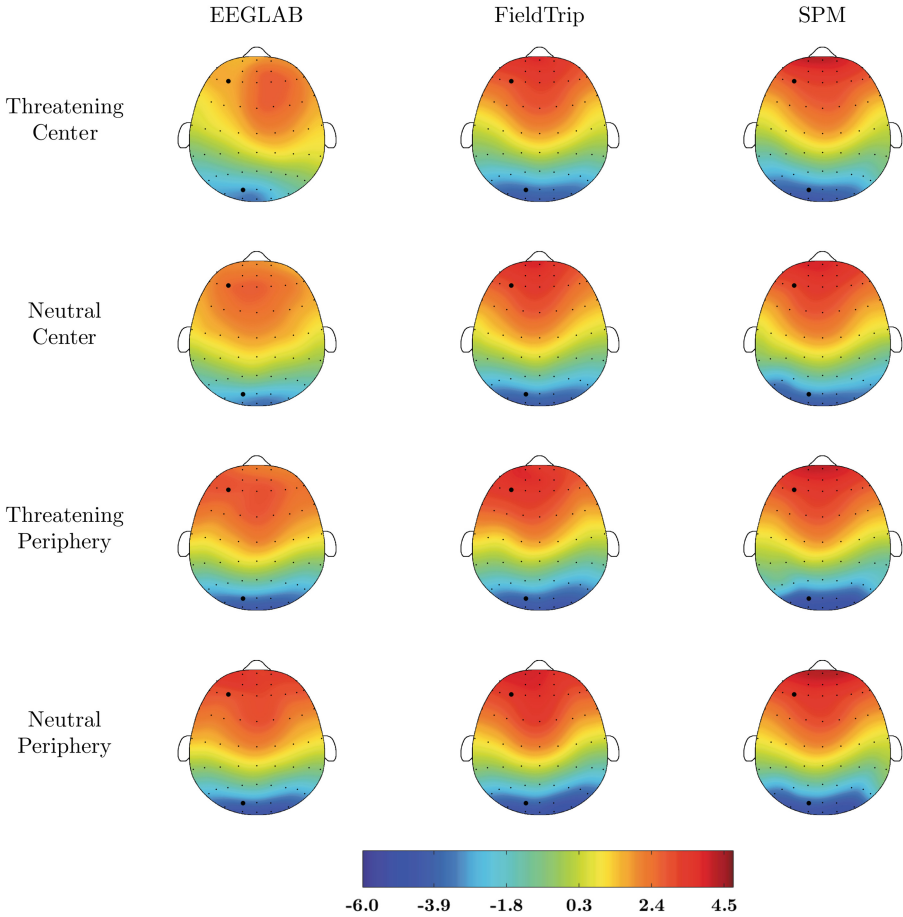
Fig. 1. Topographic maps of the averaged EEG amplitude (in μV) within the 180 to 240 ms window. Big black dots represent the selected electrodes F3 and PO3.

Regarding latency there are not clear trends among toolboxes. They are indeed close to each other, being 6.9 ms the largest variation for a single condition (between FieldTrip and EEGLAB on TP-PO3). Their variance is consistent too.

Table 2 shows results of the one-way repeated-measures ANOVA. These results show that there was not a significant main effect of the toolbox on the average peak amplitude nor latency in any of the conditions (TC, TP, NC, NP). This result is expected because as observed in Figs. 2 and 3 the separation of the ERP obtained with EEGLAB was not outside the confidence region. Besides, as presented in Table 1, these smaller amplitude values were consistent among conditions and sensors, and there were not observable latency variations.
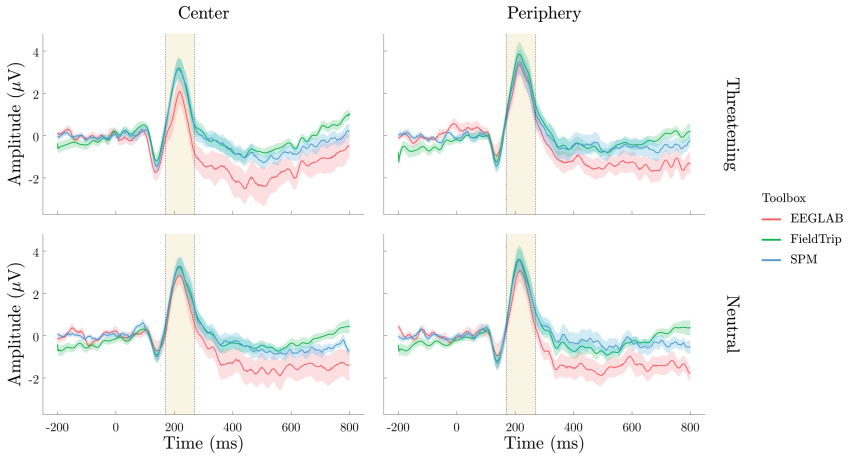
Fig. 2. Grand average ERPs recorded at F3 electrode. The waveforms obtained with the different toolboxes are overlaid for the threatening center stimuli (left-top), the threatening periphery stimuli (right-top), the neutral center stimuli (left-bottom), and the neutral periphery stimuli (right-bottom). Solid lines depict the mean value, and the shaded backgrounds show the standard error of the mean. Yellow-shaded areas show the 180 to 240 ms window used to calculate the N2 component.
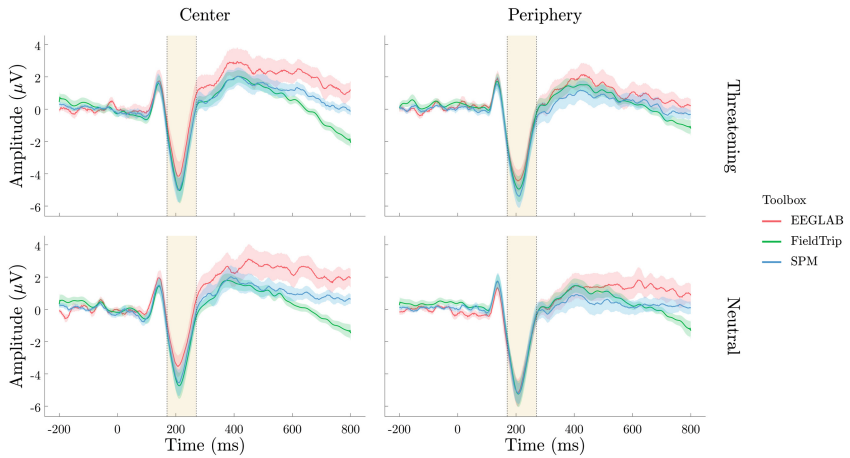

Fig. 3. Grand average ERPs recorded at PO3 electrode. Same conditions of Fig. 2. N2 activity was closer on all toolboxes than at F3. However, LPP presented larger differences on all of them at Center condition.

Table 1. Mean and standard deviation for peak latencies and amplitudes of N2 at F3 and PO3.

	F3		PO3	
	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)
TC				
EEGLAB	3.59 ± 2.54	210.66 ± 21.11	5.30 ± 3.89	205.26 ± 20.58
FieldTrip	3.85 ± 2.01	212.46 ± 19.76	5.68 ± 3.06	208.86 ± 16.32
SPM12	3.89 ± 2.23	216.16 ± 18.81	5.99 ± 3.10	210.86 ± 15.90
TP				
EEGLAB	4.03 ± 2.25	210.16 ± 17.50	5.48 ± 3.07	209.06 ± 21.96
FieldTrip	4.44 ± 2.41	209.06 ± 19.80	6.00 ± 3.41	202.16 ± 20.22
SPM12	4.39 ± 2.24	214.36 ± 19.13	6.42 ± 3.29	206.66 ± 17.80
NC				
EEGLAB	3.52 ± 1.95	210.76 ± 18.75	4.57 ± 2.91	203.96 ± 20.05
FieldTrip	3.75 ± 1.84	212.56 ± 18.80	5.60 ± 3.16	206.56 ± 18.56
SPM12	3.97 ± 2.04	217.56 ± 17.98	5.75 ± 3.16	206.86 ± 17.92
NP				
EEGLAB	3.84 ± 2.48	212.96 ± 18.13	5.89 ± 4.02	205.26 ± 17.63
FieldTrip	4.13 ± 2.82	213.46 ± 17.19	5.84 ± 3.62	204.36 ± 17.58
SPM12	4.16 ± 2.32	215.36 ± 19.01	6.14 ± 3.11	204.96 ± 16.94

Table 2. Results of the one-way repeated-measures ANOVA for peak latencies and amplitudes of N2 at F3 and PO3.

Electrode	Metric	Condition	$F(2,38)$	p	η_p^2
F3	Latency	TC	0.510	0.604	0.026
		TP	1.064	0.355	0.053
		NC	1.338	0.274	0.066
		NP	0.142	0.868	0.007
	Amplitude	TC	0.150	0.861	0.008
		TP	0.338	0.715	0.017
		NC	0.356	0.703	0.018
		NP	0.153	0.858	0.008
PO3	Latency	TC	0.497	0.612	0.025
		TP	0.841	0.439	0.042
		NC	0.154	0.857	0.008
		NP	0.019	0.981	0.001
	Amplitude	TC	0.277	0.760	0.014
		TP	0.638	0.534	0.032
		NC	1.284	0.289	0.063
		NP	0.061	0.941	0.003

5 Conclusion

The present study investigated the effect of using a specific toolbox to process EEG data intended to ERP analysis. In summary, regarding to the N2 component we did not find significant differences between data extracted with the three tested toolboxes: EEGLAB, FieldTrip and SPM12. Although there are not significant differences, results showed that peak amplitude data extracted using EEGLAB exhibited lower average values, but these were consistent among conditions. Then, we do not expect differences in contrasts tests due to this issue. Further work could include a detailed investigation of which steps in the processing contribute most to this variation.

There are visual differences between the ERP waveforms for later potentials (LPP). This could be inconvenient for emotion regulation research, given that LPP reflects facilitated attention to emotional stimuli. Further investigation must be carried to establish how a particular toolbox can affect contrasts among conditions and later group analysis.

References

1. Gevins, A., Leong, H., Smith, M.E., Le, J., Du, R.: Mapping cognitive brain function with modern high-resolution electroencephalography. Trends Neurosci. **18**(10), 429–436 (1995). https://doi.org/10.1016/0166-2236(95)94489-R
2. Sur, S., Sinha, V.: Event-related potential: an overview. Ind. Psychiatry J. **18**(1), 70–73 (2009). https://doi.org/10.4103/0972-6748.57865
3. Gozke, E., Tomrukcu, S., Erdal, N.: Visual event-related potentials in patients with mild cognitive impairment. Int. J. Gerontol. **10**(4), 190–192 (2016). https://doi.org/10.1016/j.ijge.2013.03.006
4. Hirata, K., et al.: Abnormal information processing in dementia of alzheimer type. A study using the event-related potential's field. Eur. Arch. Psychiatry Clin. Neurosci. **250**(3), 152–155 (2000). https://doi.org/10.1007/s004060070033
5. Wauthia, E., Rossignol, M.: Emotional processing and attention control impairments in children with anxiety: an integrative review of event-related potentials findings. Front. Psychol. **7**, 562 (2016). https://doi.org/10.3389/fpsyg.2016.00562. https://www.frontiersin.org/article/10.3389/fpsyg.2016.00562
6. Trujillo, S.P., et al.: Atypical modulations of N170 component during emotional processing and their links to social behaviors in ex-combatants. Front. Human Neurosci. **11**(May), 1–12 (2017). https://doi.org/10.3389/fnhum.2017.00244
7. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**(1), 9–21 (2004). https://doi.org/10.1016/j.jneumeth.2003.10.009
8. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W.: Statistical Parametric Mapping (1994)
9. Litvak, V., et al.: EEG and MEG data analysis in SPM8. Comput. Intell. Neurosci. **2011**, 32 (2011)

10. Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M.: FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. **2011**, 156,869 (2011). https://doi.org/10.1155/2011/156869

11. BioSemi, B.: BioSemi ActiveTwo [EEG system]. BioSemi, Amsterdam (2011)

12. Jurcak, V., Tsuzuki, D., Dan, I.: 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. Neuroimage **34**(4), 1600–1611 (2007)

13. Parra Rodríguez, M., Sánchez Cuéllar, M., Valencia, S., Trujillo, N.: Attentional bias during emotional processing: evidence from an emotional flanker task using IAPS. Cogn. Emotion 1–11 (2017). https://doi.org/10.1080/02699931.2017.1298994

14. Picton, T., Lins, O., Scherg, M.: The recording and analysis of event-related potentials. In: Handbook of Neurophysiology, pp. 4–73. Elsevier Science (1995)

15. Picton, T., et al.: Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology **37**, 127–152 (2000)

16. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.M., Robbins, K.A.: The prep pipeline: standardized preprocessing for large-scale eeg analysis. Front. Neuroinform. **9**, 16 (2015). https://doi.org/10.3389/fninf.2015.00016

17. Teplan, M.: Fundamentals of EEG measurement. Measur. Sci. Rev. **2**(2), 1–11 (2002)

18. Hyviirinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, Hoboken (2001)

19. Balconi, M., Pozzoli, U.: Event-related oscillations (ERO) and event-related potentials (ERP) in emotional face recognition. Int. J. Neurosci. **118**(10), 1412–1424 (2008). https://doi.org/10.1080/00207450601047119