Krzysztof MOSKWA, Izabela REJER

West Pomeranian University of Technology Szczecin

doi:10.15199/48.2017.01.24

How human perceive an application error? Error potential study

***Abstract**. The aim of this paper is to present the results of the experiment that was performed to compare the characteristics of brain potentials evoked by subject slips, to that evoked by errors made by the system itself. Experiment has been performed with two subjects. Obtained results show rather clear but unexpectedly opposite averaged patterns of brain potentials. Brain waves of one of the subjects suggest that he perceived the application error like his own, the averaged EEG signal from the second subject clearly shows that he correctly recognized application errors.*

Streszczenie.** Celem artykuły jest przedstawienie rezultatów eksperymentu, który został przeprowadzony, aby porównać cechy sygnałów EEG wywołanych przez pomyłkę badanego obiektu, z cechami sygnałów wywołanymi przez system komputerowy. Eksperyment wykonany został na dwóch ochotnikach.Uzyskane rezultaty pokazują dość wyraźne, choć niespodziewanie różne wzory przebiegów sygnałów. Wyniki pierwszego przypadku sugerują, że badany obiekt rozpoznał błędne działanie systemu jako własne błędy, podczas gdy uśrednione przebiegi EEG w przypadku drugiego badanego obiektu wskazują na to, że dostrzegł on błędne działanie testowej aplikacji. **Jak człowiek odbiera błędy aplikacji

Keywords: EEG, Error Potential, Error Detection.
Słowa kluczowe: EEG, Potencjał Wyzwolony Błędu, Detekcja Błędów.

Introduction

The graphic interface of a computer application is one of the key factors that decides on its success. The incorrectly designed interface, demanded the constant attention of the user and flooded him with an endless stream of possible options is the first step to an application failure. Research on emotion and cognition has shown that better designed products are more attractive. “Having better, easier to learn and use products supposed to be personal and professional best interest”, writes A. Cooper in his business case book [1].

Of course, there are some general rules on how to design a user-friendly application interface, e.g. performing research, collecting requirements, modelling users, goal oriented design based on personas and other methods [2]. These rules, however, do not cover all the conditions that can be met when an interface of a new application is designed. Hence, before the application is released to the market, its interface should be tested with prospective users. During the tests several aspects of the interface are investigated, one of them is vulnerability to errors.

D. Norman defines two fundamental categories of human errors. First of them are mistakes. Mistakes result from the choice of inappropriate goals and conscious deliberations. Slips, which are the second group, arise from automatic human behaviour. For example, “when subconscious actions that are intended to satisfy our goals get waylaid route” [3]. “People make errors every minute. The human brain is a flexible and error tolerant, and makes corrections automatically”, writes D. Norman. Sometimes these are even hardly being noticed, if not pointed out by someone else. Computers are strictly logic items and do not have the same tolerance unless special mechanisms are embedded in their applications.

The most popular methods for exploring the user experience are surveys, talkalouds, focus groups and interviews. Due to the communication barriers such methods, traditionally used for interface design, may fail to capture appropriately issues that last a very short time. For example the post-use interviews rely on our imperfect memory, and talkalouds may interrupt the natural flow of interaction [4]. A potential solution to the mentioned problem might be an extension of the classical techniques of interface design with other methods, allowing more direct ways of assessing the emotional state of the user and his patterns of actions while operating with the software. Electroencephalography (EEG) is one of the potential methods that allow detecting the areas of the interface causing users to make the most slips. J. Escalante and others have shown in [4] the practical approach of using, so called, error potential to identify interface design flows. Error potential (ErrP) is an event-related potential (ERP) appearing after an error trial. It is characterized by two components: a negative wave called error negativity (Ne) and a following broader positive peak called error positivity (Pe) [5]. According to the research performed by Falkenstein et al. [6] Pe is more specific to errors, while a small negativity similar to Ne can also be observed also on the correct trials.

Detecting such signal features may allow finding interaction operational patterns, that are potential sources of user slips. With this information the application designer could correct and redesign the application to make it better adapted to human cognitive capabilities, thus characterized by significantly fewer design flaws than original interface. To be able to detect ErrP and differentiate it from other EEG signal features, it is necessary to collect a number of EEG signal samples and make an attempt to compare their characteristics depending on the different user interface conditions.

The aim of this paper is to present the results of the preliminary experiment that was performed to compare the characteristics of ErrP evoked by subject slips, to ErrP evoked by errors made by the system itself. The experiment, carried out with two subjects, consisted of three sessions. In the first session subjects were supposed to make slips on their own (mainly due to the imposed time restriction). In the second session both, the subject and the system could make an error (the system error rate was equal to 20%) but the subject was not informed about the intended machine errors and hence was sure that he was the one mistaking. The only change introduced in the third session was that before it started, the subject had been informed about the possible machine errors. With such a setup it was possible to answer the question whether all three conditions (human error, machine error perceived as a human error, and machine error perceived as a computer error) induced ErrP with similar characteristics. The idea for this experiment has been taken from Ferrez and Millán [7]. They designed a similar experiment, but they analyzed only the ErrP appearing when the human perceived the errors made by a system.

For the purpose of the experiment, a simple synthetic test application was created. This application was constructed to trigger predetermined user behaviour and erroneous reactions. As the test application behaviour is known a priori, so it is possible to determine an exact moment in time, when the specific erroneous situation happened and find potential relationships with ErrP features.

The structure of the paper is as follows. Section 2 presents the experiment setup. It also covers the detailed description of the test environment prepared for the purpose of the experiment. Next section describes the methodology used for EEG signal processing. Section 4 delivers the results of the experiment together with their analysis. Finally, the last section closes the paper.

Experiment Setup

To gather EEG data for ErrP analysis, the test environment of the configuration presented in Fig. 1 was prepared. EEG data was recorded from 19 monopolar channels at a sampling frequency of 256 Hz. The (passive) electrodes, connected to the scalp according to the International 10-20 system [8], were used in the experiment. The reference electrode was placed on the left mastoid and the ground electrode at Fz. The impedance of the electrodes was controlled with BrainMaster Discovery software and was kept below 5 kΩ.

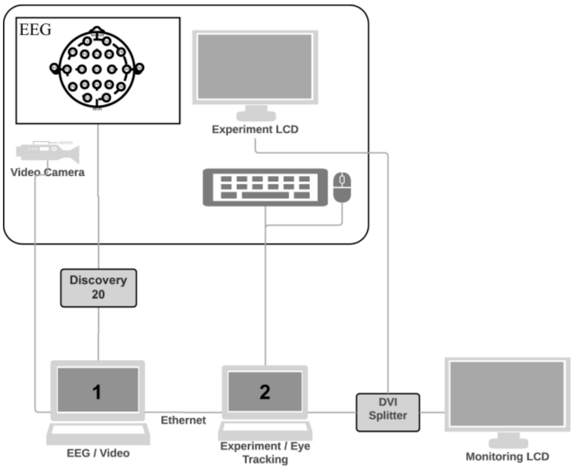

Fig. 1. The configuration of the test environment setup. EEG signal acquired with Discovery 20 device. EEG signal and audio-video recording stored on laptop number 1. Eye-tracking data with video screen capture stored on the number 2 laptop. The experiment performed on laptop 2 - second display. The second screen output splitted to two external monitors. One for subject's operations, another for experimenter monitoring purposes

Since ErrP is characterized by a frontocentral distribution along the midline [6], only Cz channel was selected for signal processing. The remaining channels were stored for further analysis. Together with EEG signal, two types of supplementary materials were also recorded: data from eye-tracking system, showing areas of the screen where a subject looked at during the experiment, and audio-video recordings of his behaviour that allow reviewing the experiment for a better understanding of the subject state of mind at the time of the session. The raw EEG signals were recorded with OpenVibe software [9]. Two file formats were used to store the raw data simultaneously – native OpenVibe OLV format and CSV format. The OpenVibe scenario, prepared for storing the EEG signals, also included components allowing to store stimulus information coming from a test application. The stimulus information was sent over integrated VRPN protocol.

The goal of the test application is to trigger the predetermined user behavior and erroneous reactions. The application works as follows. First, a left or right arrow is displayed on the screen (Fig. 2). The arrow direction is chosen randomly. The subject's task is to press the left control button on the connected keyboard when the left arrow appears, and the right control button, when the right arrow appears. The subject has up to half a second to react to this stimulation. The subject is also informed that he should try to react as fast as possible in time lower than 300ms. The time pressure is necessary to ensure that enough trials within one session will end with a subject slip. When the subject presses the button, the information, whether the answer was correct or not, appears on the screen together with the reaction time. To be more specific, when the subject responds correctly, the "OK" text appears on the screen, and the arrow is filled with green colour. When the subject makes a mistake, he gets feedback with "Wrong" text, and the red arrow pointing in the opposite direction. In both cases, the user is informed about the reaction time, typically varied between 250 and 350ms. The entire cycle repeats after a random time between 5 and 10 seconds.

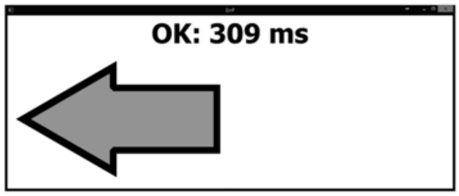

Fig. 2. The test application, displaying randomly left or right arrow and feedback based on subject's reactions.

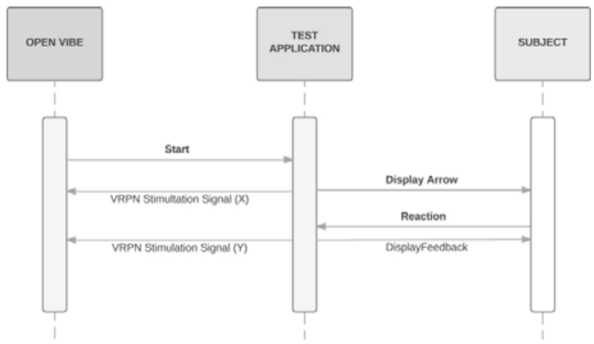

Fig. 3. Workflow of stimulations signals transferred between OpenVibe and test application

Fig. 3 presents the workflow of stimulation signals transferred between OpenVibe and the test application. Five different types of stimulations are recorded during the experiment:

- arrow appears on the screen;
- arrow disappears from the screen;
- subject response is correct;
- subject response is incorrect;
- subject response is correct, but the application shows an incorrect feedback.

The workflow of stimulation signals is as follows. OpenVibe sends information when to start the next cycle of the application scenario. Test application sends stimulation signal to OpenVibe informing that the left or right arrow is displaying on the screen and at the same time presents an arrow to the subject. At the moment when subject reacts to the arrow that is displayed on the screen, the application sends another type of stimulation signal to OpenVibe and also displays the feedback to the subject. The time of recorded stimulation signal is the same as the time of the feedback and is synchronized with EEG signal.

Two healthy volunteers (male, aged 29 and 32) participated in the experiment. Both subjects were righthanded and had normal or corrected-to-normal vision. None of them reported any previous mental disorders. Three slightly different sessions were performed with each subject. The aim of the first session was to evaluate the average error potential in the classic settings, it is when the subject unintentionally made slips. The second session was similar to the first one, with one exception – in 20% of correct subject responses, the application reacted incorrectly. In this session, the subject was not aware of the erroneous application behaviour. Before the third session, the subject was informed that the application could react incorrectly. Each session took about 30 minutes. To minimize the subject fatigue, each session was followed by 10 minutes break.

Methods

The EEG data, recorded during the experiment, together with the stimulation information were imported into MatLab environment as CSV files for processing and analysis. First, the set of stimulation signals received from VRPN communication channel was taken into consideration. While the signals generated by OpenVibe did not need any preprocessing, some of those generated by subjects reactions were redundant and had to be removed from the stimulation file. The excessive stimulation signals occurred mainly in the second sessions when the subject was exposed to the unexpected wrong responses from the test application. In such situation, both subjects tried to correct themselves by pressing the key once again. The cases when they really slipped also triggered the correction reaction but not so often as the application errors. The results from the preprocessing stage had been stored in stimulation array.

The EEG signals from all 19 channels were filtered with 4th-order bandpass Butterworth filter with low and high cutoff frequencies set to 1 and 10Hz respectively. The filtering operation was performed using zero-phase transfer functions as described in [10]. Next, data frames (epochs) representing signals surrounding each subject-generated stimulation had to be extracted from continuous EEG signal. This step was performed on the basis of data stored in the stimulation array. The EEG data was divided into epochs of 900 ms, starting 300 ms before the stimulus onset (feedback appearing on the screen). Taking into consideration sampling frequency 256Hz and epoch length equal 900 ms, each epoch contained 230 samples. DC offset was removed from each epoch by subtracting mean signal amplitude from the filtered signal. To reduce the number of artifacts, all epochs with amplitude exceeding 100 μV were discarded from the analysis. Finally, to calculate the error potential, the EEG epochs were averaged separately for each stimulation type [11]. Then the differential signal between the mean from the correct trials and mean from incorrect trials was calculated and subjected to the analysis.

Results and discussion

Figure 4 presents the difference signal (solid line) between the averaged EEG signals received when the subject reacted correctly (dashed line), and the averaged EEG signal received when the subject made a slip (dotted line). Both averages were calculated over the trials from all three sessions. The left chart shows the results for the first subject (named S1 in further analysis), the right one – for the second subject (S2). The characteristics of waveforms from the right chart are similar to the typical Error Potential described in [4]. About 120ms after stimulus onset the averaged signal forms a negative peak (potential below -10.5μV), while 300ms after stimulus onset it goes up to +10μV and forms a positive peak. The negative peak depicts specific error negativity (Ne) while the following positive peak shows typical error positivity (Pe). Graph located on the left side of Fig. 4 does not show such distinctive results. However, signal slope around 80ms with -1.38μV can be interpreted as weak Ne, while the peak at 260ms with +2.9 μV can be construed as Pe.

To prepare waveforms presented in Fig. 5 only data from correctly responded trials recorded during the second session was taken into consideration. The figure compares EEG signal averaged over trials with correct feedback from the application (dashed line) to that averaged over trials with feedback suggested an incorrect subject response (dashed-dot line). As in can be noticed in the figure, both averaged signals are virtually similar before the stimulus onset. The amplitude of the difference waveform for both subjects does not exceed 3μV. However, after the feedback is presented to the subject (time 0), both waveforms start to differ. This phenomenon is more visible in the right chart (subject S2), where differences between signals exceed 9.9μV around the time of 100 ms after feedback.

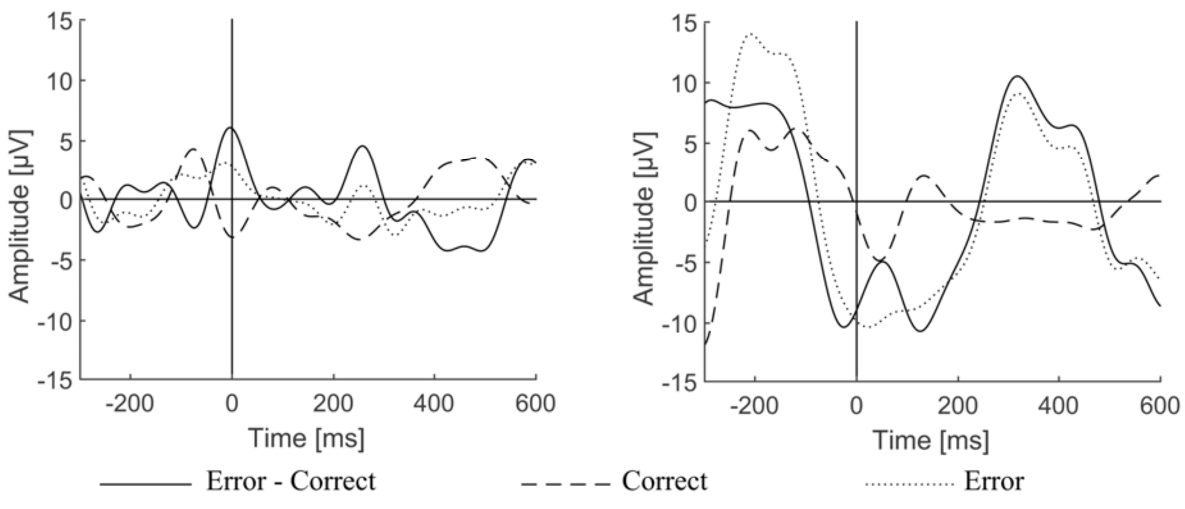

Fig. 4. The signal average of the correct responses (dashed line), wrong responses (dotted line) and their difference (solid line); subject S1 on the left, subject S2 on the right

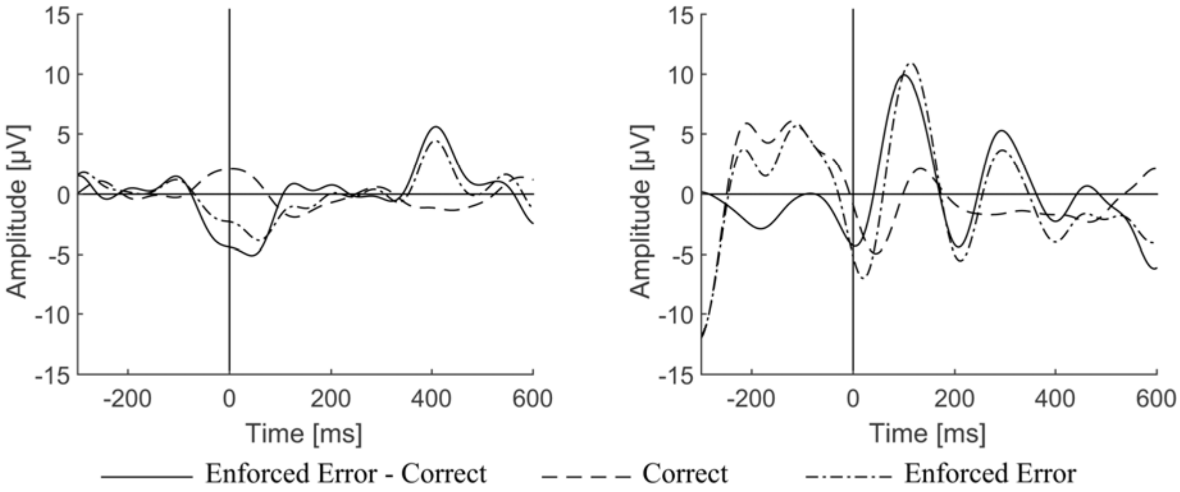

Fig. 5. The signal average of the correct subject responses with the correct application feedback (dashed line), correct subject responses with wrong application feedback (dashed-dot line) and the difference of both waveforms (solid line); data from the second session; subject S1 on the left, subject S2 on the right

The similarity of both waveforms before stimulus onset was expected, since in both cases subject's responses belonged to the same "correct response" class. Something unexpected was observed after the stimulus onset. While the averaged signal of the correct subject responses with correct machine feedback have more or less the same shape for both subjects, the shape of averaged waveforms of correct subject responses with wrong machine feedback and the difference waveforms differ significantly for both subjects. The difference waveform for subject S1 is similar to the correct wave. It looks like subject S1 correctly recognized application error and did not perceive it as his own. On the contrary, the difference waveform of subject S2 is very similar to the error potential from Fig. 4. That might mean that subject S2 recognized the application error as his own (classic Pe at 300 ms).

Before attending session 3, the subjects were informed about the possible incorrect application behaviour and hence knew that the reverse feedback did not mean their error. As it can be noticed in Fig. 6, in the case of subject S1 the application error waveform and the correct waveform once again are very similar, like in Fig. 5. However, while in Fig. 5 there was a positive peak at 400 ms that could be interpreted as a very late Pe, the difference waveform in Fig. 6 does not present anything interesting at all. The results obtained for subject S2 are not so obvious. There is no Ne visible in the right chart but the positive peak of the difference waveform can be found at about 380 ms. This peak can be interpreted as late Pe. It looks like subject S2 had problems with recognizing the application errors even if he knew that they might appear.

The last figure (Fig. 7) depicts a comparison of the EEG signal averaged over trials with incorrect subject responses (dotted line) and the EEG signal averaged over trials with the correct subject responses but wrong application feedback (dashed-dot line). As expected, since subject S1 did not perceive the application error as his own, the difference between both waveforms (solid line) for subject S1 is similar to the difference waveform observed in Fig. 4 (the difference between correct and incorrect subject responses). The main difference between both figures is a small late Ne potential at about 190 ms after stimulus onset. The difference waveform obtained for subject S2 is difficult to analyze since both errors (subject error and application error) were perceived by the subject similarly, but the perception of application error was weaker and slightly delayed in time.

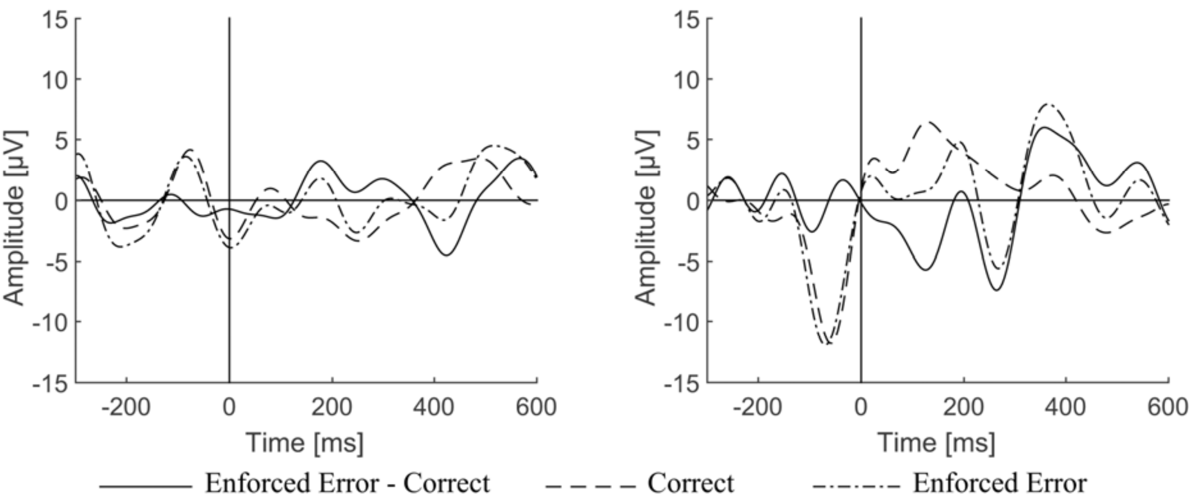

Fig. 6. The signal average of the correct subject responses with the correct application feedback (dashed line), correct subject responses with wrong application feedback (dashed-dot line) and the difference of both waveforms (solid line); data from the third session; subject S1 on the left, subject S2 on the right

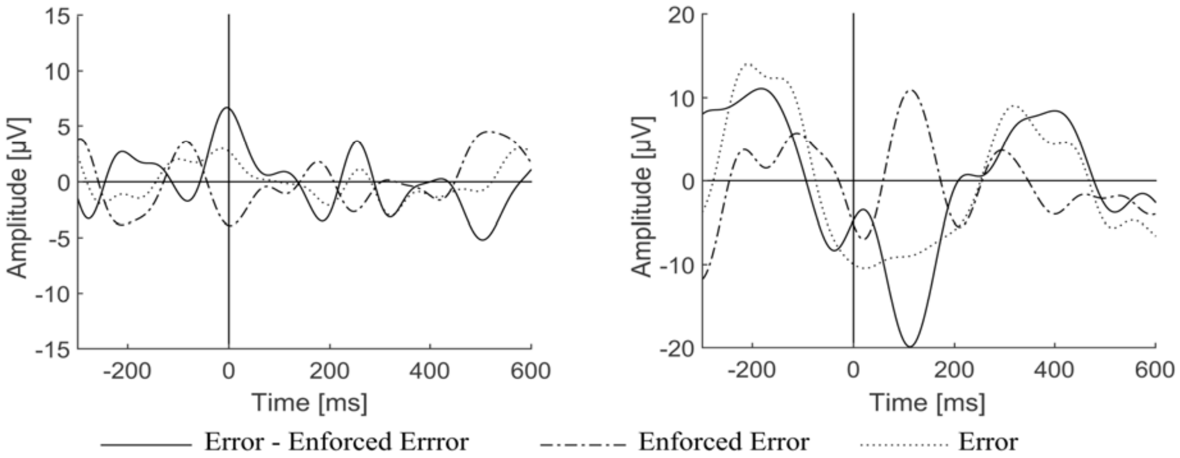

Fig. 7. The signal average of the correct subject responses with wrong application feedback (dashed-dot line), incorrect subject responses (dotted line) and the difference of both waveforms (solid line); subject S1 on the left, subject S2 on the right

Conclusions

Using the test environment described in the paper, it was possible to carry out the experiment aimed at comparing the characteristics of ErrP evoked by subject slips, to ErrP evoked by errors made by the system itself. The results obtained in the experiment are in agreement with other research in the field [4,6,7]. Although our research was carried out with only two volunteers, we were able to reproduce the signal features described in literature in terms of Pe and Ne. However, sessions with more subjects are needed to confirm recurrence of the obtained results. In addition to confirming the issues known from previous research, we performed further analysis of the recorded signals, learning typical patterns and dependences between all three combinations of average signals collected for: the correct subject responses, the incorrect subject responses, and correct subject responses with incorrect application feedback.

The main, rather unexpected, result obtained in our research was that the brain potentials appearing after perceiving an application error were so much different for both subjects. This suggests that people did not react in the same way on the application error and, hence that it might be difficult to differentiate the application error from the subject error only on the base of error potential.

Our finding is in contradiction to that reported by Ferrez and Millán in [7] where also the brain reaction on the system error was analyzed. They analyzed data from three participants and found Pe at 350-450 ms after the feedback for all three subjects (Ne was observed in two of them with a latency about 270 ms).

The aim of this paper was to deliver the detailed description of our test environment and to present results of the preliminary experiments performed in this environment. With only two subjects participated in the experiment, it is not possible to draw any general conclusion. Since the averaged brain potentials for both subjects were so much different, the experiment did not deliver the answer to the question whether the error potential is a proper tool to improve the application interface design. In order to answer this question much more research has to be done.

Authors: *mgr. inż. Krzysztof Moskwa, Zachodniopomorski Uniwersytet Technologiczny, Wydział Informatyki, ul. Żołnierska 49 71-210 Szczecin, E-Mail: krzysztof.moskwa@zut.edu.pl; dr hab. Izabela Rejer, Zachodniopomorski Uniwersytet Technologiczny, Wydział Informatyki, ul. Żołnierska 49, 71-210 Szczecin, E-mail: irejer@wi.zut.edu.pl*

REFERENCES

[1] Cooper, A. (2004). The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams Publishing

[2] Cooper, A., Reimann, R. (2007). About Face 3: The Essentials of Interaction Design. Wiley Publishing, Inc.

[3] Norman, D. A. (1988). The Design of Everyday Things. Basic Books

[4] Escalante, J., Butcher, S., Costa, M. R., Hirshfield, L. M. (2013). Using the EEG Error Potential to Identify Interface Design Flaws.Foundations of Augmented Cognition Volume 8027 of the series Lecture Notes in Computer Science pp 289-298; 7th International Conference, AC 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013. Proceedings

[5] Blankertz, B., Schäfer, C., Dornhege, G., & Curio, G. (2002) Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates. In Artificial Neural Networks—ICANN 2002 (pp. 1137-1143). Springer Berlin Heidelberg.

[6] Falkenstein, M., Hoormann, J., Christ, S., and Hohnsbein, J. (2000) ERP components on reaction errors and their functional significance: a tutorial. Biological Psychology, 87-100.

[7] Ferrez, P.W., Millán, J.R. (2005) You Are Wrong! - Automatic Detection of Interaction Errors from Brain Waves. Proceedings of IJCAI'2005.1413-1418.

[8] Jasper H. H. (1958), The ten-twenty electrode system of the international federation in electroencephalography and clinical neurophysiology, EEG Journal, (1958) Vol. 10, 371–375

[9] Renard Y., Lotte F., Gibert G., Congedo M., Maby E., Delannoy V., Bertrand O., Lécuyer A. (2010) OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments, Presence: teleoperators and virtual environments, (2010), Vol. 19, No 1

[10] Mitra, Sanjit K. Digital Signal Processing. 2nd Ed. New York: McGraw-Hill, (2001), secs. 4.4.2 and 8.2.5.

[11] Drongelen, W. (2006). Signal Processing for Neuroscientists. Academic Press, pp. 55-70