Proceedings of the
9th Graz Brain-Computer Interface Conference 2024
10.3217/978-3-99161-014-4-012

PROCESSING OF INCONGRUENT INFORMATION CAN BE DECODED FROM SINGLE-TRIAL EEG: AN AR-STUDY

M. Wimmer[1,2], A. Pepicelli[3], B. Volmer[3], N. ElSayed[1], A. Cunningham[3], B.H. Thomas[3], E.E. Veas[1,4], G.R. Müller-Putz[2,5]

[1]Know-Center GmbH, Graz, Austria
[2]Institute of Neural Engineering, Graz University of Technology, Graz, Austria
[3]Wearable Computer Lab, University of South Australia, Adelaide, SA, Australia
[4]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria
[5]BioTechMed Graz, Graz, Austria

E-mail: eveas@know-center.at, gernot.mueller@tugraz.at

ABSTRACT: Augmented reality (AR) allows users to display additional digital information about their physical environment. We present an interactive AR study, in which participants manipulated a Rubik's cube which served as a physical referent for presented digital information showing the current status of the cube. In 30% of the instances, the presented information did not match its status. We recorded the electroencephalographic data of 19 participants to study their responses to incongruent stimuli and assessed if they could be classified on a single-trial level. We found that the processing of incongruent data in AR elicits both N400 and P600 components. Further, we could classify them in 15 out of 19 participants with accuracies above chance. These results contribute to the design of brain-computer interfaces, as the decoding of such correlates could inform the system about the current mental context of the user.

INTRODUCTION

Augmented reality (AR) allows the integration of virtual content into the real world [1]. With the increasing number of head-mounted displays (HMDs) and other personal electronic devices that can create AR visualizations, the technology has become a widely available tool with manifold applications [2]. One goal of situated AR visualizations is to communicate information about physical objects to users or assist them in specific tasks [3]. To give two practical examples, this information could guide users through procedural tasks via visual cues [4] or support them, e.g., in their purchasing decisions by displaying relevant details directly next to products [5].

Efforts have been made to study users' electroencephalographic (EEG) responses to the presentation of data, e.g., showing anomalous information. A particularly prominent component of the event-related potential (ERP) caused by incongruent stimuli is the N400. It is a negative deflection relative to congruent stimuli that peaks approximately 400 ms after stimulus onset in centro-parietal areas of the scalp [6]. The N400 component has been found in response to numerous types of conflicting stimuli, such as incongruent words in sentences [7], incongruent solutions of simple mathematical problems, [8], pictures [9], or gestures [10]. Hence, Kutas et al. [7] described the N400 as "an electrophysiological sign of the 'reprocessing' of semantically anomalous information". This is relevant in the context of brain-computer interfaces (BCIs) [11], as decoding incongruent stimuli could allow systems to infer information about the user's mental context without making it explicit.

However, only a few studies attempted to decode semantic incongruencies on a single-trial level. Geuze et al. [12] found an N400 effect in a word association task using related and unrelated word pairs and could decode them with accuracies between 54% and 67%. In a similar task, Dijkstra et al. [13] studied the responses to multiple consecutive word stimuli after presenting a target word. They reported similar neural responses and achieved a classification accuracy of 59.5%. Both works used an L_2 regularized logistic regression algorithm for classification. Finally, Tanaka et al. [14] presented semantically correct and incorrect sentences and found both N400 and P600 components. Using a multilayer perceptron, they could correctly identify them in up to 59.5% of the instances. Interested readers are referred to [15] for an overview of N400 for BCIs.

All three above-mentioned papers explored EEG correlates of semantic anomalies in language processing. In this work, we studied the neural responses to incongruent information in an AR scenario. In particular, we investigated the following two research questions (RQs):

RQ1) Can we find EEG responses to the presentation of incongruent information using AR?

RQ2) Can we use these responses to discriminate congruent and incongruent information on a single-trial level?

For this, we designed an interactive paradigm, in which users were visually instructed on how to manipulate a physical Rubik's cube and presented situated information related to the cube, which could either match the users' expectations or not.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-012

MATERIALS AND METHODS

Participants: Twenty healthy volunteers (20 to 45 years old, 28.2 years on average, 14 male and six female) participated in the study. The study was conducted in accordance with the Declaration of Helsinki (1975) and approved by the ethics committee of the University of South Australia. All participants gave their written consent before conducting the experiment and received vouchers worth 40 AUD as compensation.

AR HMD: AR visualizations were presented using a HoloLens 2 (Microsoft, Redmond, WA, USA) and designed in Unity 2021.1.3[1]. We used the HMD to record gaze signals at 30 Hz during the data presentations.

EEG recordings: EEG signals were acquired using a BrainAmp amplifier (Brain Products, Munich, Germany) at 500 Hz. We positioned 32 electrodes according to the international 10-10 system at AFz, F3, F1, Fz, F2, F4, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P3, P1, Pz, P2, and P4. Reference and ground electrodes were placed at Fpz and the right mastoid, respectively. EEG signals, gaze data, and markers from the experimental paradigm were synchronized utilizing the lab streaming layer (LSL) protocol[2].

Experimental setup: Participants sat at a table such that they could easily reach a tricolor (red, blue, white) Rubik's cube (see Fig. 1). A camera (Canon EOS 200D II, Tokyo, Japan) pointed at the cube to detect the nine colors of its top surface (camera not visible in Fig. 1). Color detection was performed based on the Qbr Rubik's cube solver[3], implemented in Python and OpenCV.

Figure 1: **Experimental setup.** A participant wearing an EEG cap and an HMD. The tricolor Rubik's cube is in its starting position.

Experimental procedure: First, participants took an Ishihara test[4] to assess their color vision [16]. After successful completion, participants performed one training run consisting of ten trials to familiarize themselves with the experimental paradigm. The following experiment consisted of 6 runs of 33 trials (23 congruent, 10 incongruent). The order of the trials was randomized. Between runs, the participants could take short breaks of usually one to five minutes to avoid fatigue.

Experimental paradigm: The timings of one trial of the paradigm are shown in Fig. 2. Each trial started with the presentation of one or two visual cues indicating which one or two cube manipulations the participants should perform. Manipulations included rotating a specific row or column of the cube, or the whole cube in a given direction. Depending on the number of manipulations, the visual cues were presented for 1 or 2 seconds (s). Thereafter, participants were instructed to take the cube, perform the indicated manipulations, and return it to its starting position. At this point, participants should count the number of red, blue, and white squares on the cube's top surface. For example, in Fig. 2, after performing the manipulations the count would be three red, two blue, and four white squares. The participants indicated that they knew the correct count by pressing a physical button on the keyboard. This triggered the presentation of a fixation cross and a frame on the left side of the cube. After 1.3 to 1.7 s (randomized), a congruent or incongruent count was presented inside the frame for .75 s following the order red-blue-white. Incongruent answers deviated strongly from congruent answers, i.e., 1-1-1 or 0-8-0, which are impossible counts per se. Participants were instructed to fixate their gaze on the cross and to avoid gaze shifts during the data presentation as much as possible. Each data presentation was followed by a break of .75 s before a new trial was introduced with a countdown from two to zero (1.5 s).

EEG data preprocessing: The data processing and analysis were performed offline using Matlab R2022a (The MathWorks, MA, USA) incorporating the EEGLAB toolbox (v2022.0) [17].

EEG data were filtered between 1 and 25 Hz using a zero-lag Butterworth filter of order 4. We applied two notch filters at 30 Hz and 50 Hz to suppress noise from the HMD and the power line (zero-lag, Butterworth, second order). After resampling the signals to 125 Hz, we applied the extended Infomax algorithm [18] to perform independent component analysis (ICA) [19]. Based on visual inspection, we rejected components corresponding to eye movements or blinks. Thereafter, we segmented the data into epochs of 1.5 s ([-.5, 1] s relative to the stimulus onset).

We rejected contaminated epochs through visual inspection and based on amplitude ($\pm$ 35 μV), kurtosis, and joint probability (both 5 times the standard deviation) [20], similar to [21]. To avoid the influence of possible residual eye-related artifacts, we rejected trials with excessive eye movements. For that, we computed the variance of the eye movements of each trial and removed epochs with a z-score outside $\pm$ 3. On average, we rejected 9% of the trials per participant.

One participant could not identify the incongruent stimuli and was subsequently removed from the analysis.

Classification: We performed stimulus-locked classification with two classes (congruent and incongruent). As

[1] https://unity.com
[2] https://github.com/sccn/labstreaminglayer
[3] https://github.com/kkoomen/qbr
[4] https://www.colorblindnesstest.org/ishihara-test/

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-012

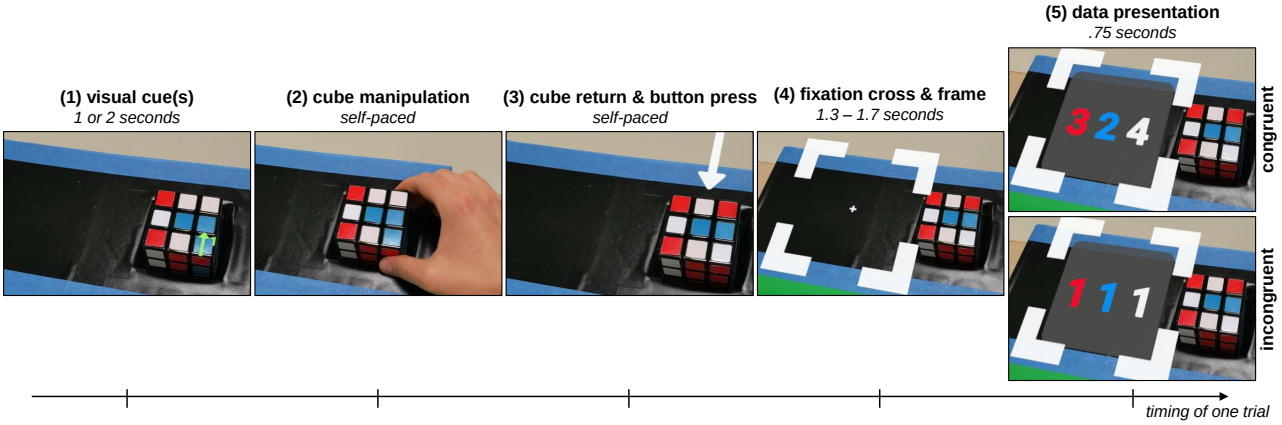

Figure 2: **Timing of one trial. (1)** Each trial started with the presentation of one or two visual cues (green arrow). In the given example, the participant was instructed to rotate the right column upwards. **(2)** After performing the manipulation, participants returned the cube to its starting position and counted the colors on the top surface. In the example, the count is 3 red, 2 blue, and 4 white. **(3)** As soon as the participants knew the correct count, they pressed a physical button. **(4)** Next, a fixation cross and a frame for the following data presentation were presented for 1.3 to 1.7 s. **(5)** Finally, a correct (top) or incorrect (bottom) count was presented for .75 s.

features, for each channel, we computed the average amplitude of overlapping windows of 152 ms in steps of 32 ms, where the first window started at 252 ms after the stimulus and the last window after 844 ms (e.g., 252-404 ms, 284-436 ms, etc.) [22]. These features were z-scored and used to train a shrinkage linear discriminant analysis (sLDA) classifier [23]. For every participant, we trained and tested the personalized classifier ten times in a 5-fold cross-validation approach on balanced datasets by choosing a random sample of congruent trials for each of the ten iterations.

We report the true positive rates (TPR) as the fraction of incongruent trials that were correctly classified. Analogously, the true negative rate (TNR) is the fraction of correctly classified congruent trials. The accuracy is the mean of the TPR and the TNR. TPR, TNR, and accuracy are calculated from the average of the 50 folds.

Statistics: We performed Wilcoxon signed-rank tests for a sample-wise comparison of the ERPs following congruent and incongruent stimuli (Fig. 3). To correct for multiple comparisons, we applied the false discovery rate (FDR) procedure (α = .05). To assess if the classification accuracies are significantly above chance [24], we computed the 95% confidence interval of the chance level using a cumulative binomial distribution [25]. We calculated the significance level (SL) for each participant individually (Tab. 1).

RESULTS

Figure 3 shows the grand average EEG results, i.e., the mean of the 19 participant averages. The grand average ERPs at CPz are depicted for both classes (mean $\pm$ 2 times the standard error of the mean (SEM)) as well as their difference computed by subtracting congruent from incongruent (Fig. 3a). This difference signal has a negative peak with a maximum amplitude of -2.53 µV at t = .47 s and a positive peak with a maximum of 1.82 µV at t = .68 s, relative to the stimulus onset. Congruent and incongruent responses differ significantly ($p < .05$) in the intervals [.40, .52] s, [.64, .72] s, and [.76, .77] s. For the first two, we show the mean topographical distributions of the intervals, revealing mainly centro-parietal responses (Fig. 3b).

Table 1 summarizes the classification accuracies for each participant, including the TPR, TNR, and SL. On average, 63.3% (TPR = 62.3%, TNR = 64.3%) of the trials were correctly classified, which exceeds the average SL by 5.3%. In 15 out of 19 participants (79%), incongruent stimuli could be distinguished from congruent ones above chance, five achieved accuracies of 70% or higher.

Table 1: **Classification results.** Participants with accuracies exceeding the SL are marked with '*'.

Participant	TPR	TNR	Accuracy	SL
	%	%	%	%
P1*	76.0	76.8	76.4	58.0
P2	51.6	56.8	54.2	57.6
P3*	59.6	60.6	60.1	57.6
P4*	60.0	56.1	58.1	58.0
P5	47.6	52.4	50.0	58.2
P6*	63.8	64.4	64.1	58.0
P7	55.3	54.1	54.7	58.0
P8*	56.5	60.3	58.4	57.5
P9*	67.1	72.9	70.0	58.2
P10*	59.1	61.8	60.4	57.6
P11*	57.5	63.9	60.7	58.3
P12*	73.5	77.3	75.4	58.3
P13	57.5	55.3	56.4	58.0
P14*	63.3	63.6	63.5	58.0
P15*	72.5	70.3	71.4	58.0
P16*	59.8	61.8	60.8	58.2
P17*	63.4	63.8	63.6	58.0
P18*	56.8	61.6	59.2	57.8
P19*	82.4	88.0	85.2	58.2
Average	62.3	64.3	63.3	58.0

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-012

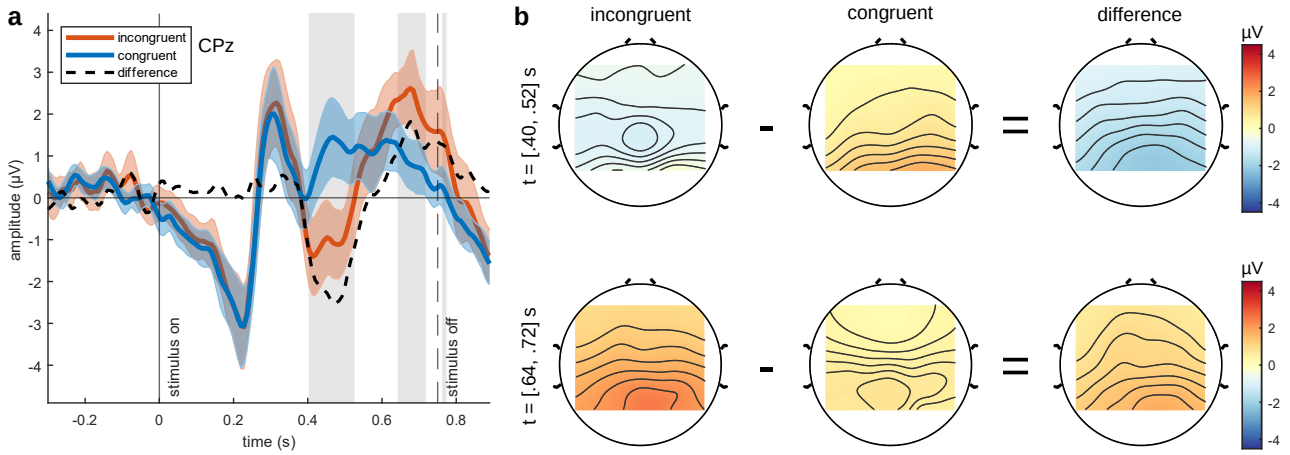

Figure 3: **Grand average EEG results.** **(a)** ERPs at CPz following the presentation of congruent (blue) and incongruent (orange) stimuli, and their difference *incongruent* − *congruent* (dashed black). Shaded areas indicate ± 2 · SEM. Significant differences ($p < .05$) are highlighted in gray. **(b)** Average topographical distributions of the significant intervals [.40, .52] s (top row) and [.64, .72] s (bottom row) for incongruent (left column), congruent (middle column), and their difference (right column).

DISCUSSION

We developed an interactive AR task incorporating situated information related to a Rubik's cube. In situ data presentation is an application area for AR technologies that offers complementary information about the user's physical environment [26]. The information is usually derived from available data and may not match the expectations of the users, given their mental context in their current situations.

Subsequently, to answer **RQ1**, we studied the participants' neurophysiological responses following incongruent stimuli and found centro-parietal N400 and P600 effects. The N400 has previously been linked to the processing of incongruent information. Since the current experiment yielded very similar patterns, we conclude that participants perceived the erroneous counts as incongruent with their semantic context, i.e., the Rubik's cube. For instance, both the N400 and P600 components have been reported for arithmetic incongruencies, found after simple multiplication errors, e.g., "$7 \cdot 8 = 54$" [8]. Judged by its scalp distribution, the authors hypothesized that the positive peak belongs to the family of P300 effects and subsequently reflects the participants' surprise following improbable stimuli, which is likely to be the case in our work too. Similarities in time course, topography, and polarity have already been suggested earlier [27]. This aligns with Coulsen et al. [28], whose experiments revealed the sensitivity of the P600 amplitude to the frequency of improbable stimuli, similar to the P3b, a subcomponent of the P300. The increased latency was explained by differences in the stimulus complexity [29]. However, counterevidence was provided when different neural generators were found to play crucial roles in the modulation of the P600 and the P300 [30]. The debate on whether the P600 is a form of a P3b is still ongoing, we refer to Leckey and Federmeier [31] for an overview.

For **RQ2**, we studied the feasibility of distinguishing congruent and incongruent responses on a single-trial basis. Using personalized classifiers, i.e., trained and tested with data of the same participant, we could decode incongruent stimuli with accuracies between 50% and 85%. Given the relevance of the N400 and P600 in neuroscience research, these components have not yet been granted much attention in the BCI community. To our best knowledge, only three works have attempted time-locked classification of incongruent information using linguistic stimuli, achieving accuracies of nearly 60% [12–14]. Our classification results are in a similar range, exceeding their reported accuracies slightly by about 4%. However, we provide first evidence that decoding the processing of incongruent in situ information in AR is feasible.

CONCLUSION

In this work, we demonstrated that neural correlates of the processing of incongruent information can be measured in AR scenarios. These correlates are consistent with the existing literature that focuses on monitor-based tasks. Further, we showed that the classification of incongruent trials on a single-trial level above chance is possible for most participants. This can be relevant for the design of BCIs since these correlates could allow to infer active mental concepts of users.

ACKNOWLEDGEMENTS

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-012

REFERENCES

[1] Azuma RT. A survey of augmented reality. Presence: Teleoperators & Virtual Environments. 1997;6(4):355–385.
[2] Marriott K *et al.* Immersive analytics. Springer (2018).
[3] Kalkofen D, Sandor C, White S, Schmalstieg D. Visualization techniques for augmented reality. In: Handbook of Augmented Reality. Springer, 2011, 65–98.
[4] Volmer B, Liu JS, Matthews B, Bornkessel-Schlesewsky I, Feiner S, Thomas BH. Multi-level precues for guiding tasks within and between workspaces in spatial augmented reality. IEEE Transactions on Visualization and Computer Graphics. 2023;29(11):4449–4459.
[5] ElSayed N, Thomas B, Marriott K, Piantadosi J, Smith R. Situated analytics. In: Proc. BDVA. 2015, 1–8.
[6] Kutas M, Federmeier KD. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology. 2011;62:621–647.
[7] Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science. 1980;207(4427):203–205.
[8] Niedeggen M, Rösler F, Jost K. Processing of incongruous mental calculation problems: Evidence for an arithmetic N400 effect. Psychophysiology. 1999;36(3):307–324.
[9] Barrett SE, Rugg MD. Event-related potentials and the semantic matching of pictures. Brain and Cognition. 1990;14(2):201–212.
[10] Wu YC, Coulson S. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology. 2005;42(6):654–667.
[11] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clinical Neurophysiology. 2002;113(6):767–791.
[12] Geuze J, Gerven MA van, Farquhar J, Desain P. Detecting semantic priming at the single-trial level. PloS One. 2013;8(4):e60377.
[13] Dijkstra K, Farquhar J, Desain P. Electrophysiological responses of relatedness to consecutive word stimuli in relation to an actively recollected target word. Scientific Reports. 2019;9(1):14514.
[14] Tanaka H, Watanabe H, Maki H, Sakriani S, Nakamura S. Electroencephalogram-based single-trial detection of language expectation violations in listening to speech. Frontiers in Computational Neuroscience. 2019;13:15.
[15] Dijkstra K, Farquhar J, Desain P. The N400 for brain computer interfacing: Complexities and opportunities. Journal of Neural Engineering. 2020;17(2):022001.
[16] Birch J, McKeever LM. Survey of the accuracy of new pseudoisochromatic plates. Ophthalmic and Physiological Optics. 1993;13(1):35–40.
[17] Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004;134(1):9–21.
[18] Lee TW, Girolami M, Sejnowski TJ. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation. 1999;11(2):417–441.
[19] Makeig S, Bell A, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems. 1995;8.
[20] Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage. 2007;34(4):1443–1449.
[21] Wimmer M, Weidinger N, Veas E, Müller-Putz GR. Multimodal decoding of error processing in a virtual reality flight simulation. Scientific Reports. 2024;14(1):9221.
[22] Yasemin M, Cruz A, Nunes UJ, Pires G. Single trial detection of error-related potentials in brain–machine interfaces: A survey and comparison of methods. Journal of Neural Engineering. 2023;20(1):016015.
[23] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components—a tutorial. NeuroImage. 2011;56(2):814–825.
[24] Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random: A closer look on BCI results. International Journal of Bioelectromagnetism. 2008;10:52–55.
[25] Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. Journal of Neuroscience Methods. 2015;250:126–136.
[26] Veas E, Grasset R, Ferencik I, Grünewald T, Schmalstieg D. Mobile augmented reality for environmental monitoring. Personal and Ubiquitous Computing. 2013;17:1515–1531.
[27] Osterhout L, Holcomb PJ. Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language. 1992;31(6):785–806.
[28] Coulson S, King JW, Kutas M. Expect the unexpected: Event-related brain response to morphosyntactic violations. Language and Cognitive Processes. 1998;13(1):21–58.
[29] Kutas M, McCarthy G, Donchin E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science. 1977;197(4305):792–795.
[30] Frisch S, Kotz SA, Von Cramon DY, Friederici AD. Why the P600 is not just a P300: The role of the basal ganglia. Clinical Neurophysiology. 2003;114(2):336–340.
[31] Leckey M, Federmeier KD. The P3b and P600(s): Positive contributions to language comprehension. Psychophysiology. 2020;57(7):e13351.

Published by
Verlag der Technischen Universität Graz