Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-42

HETEROGENEITY OF EVENT-RELATED POTENTIALS IN A SCREEN-FREE BRAIN-COMPUTER INTERFACE

Henrich Kolkhorst[1,2], Joseline Veit[1], Wolfram Burgard[2], Michael Tangermann[1,2]

[1]Brain State Decoding Lab, Cluster of Excellence BrainLinks-BrainTools
[2]Autonomous Intelligent Systems Lab, Department of Computer Science
University of Freiburg, Freiburg, Germany

E-mail: kolkhorst@informatik.uni-freiburg.de
michael.tangermann@blbt.uni-freiburg.de

ABSTRACT: Interacting with the environment using a brain-computer interface involves mapping a decoded user command to a desired real-world action, which is typically achieved using screen-based user interfaces. This indirection can increase cognitive workload of the user. Recently, we have proposed a screen-free interaction approach utilizing visual *in-the-scene* stimuli. The sequential highlighting of object surfaces in the user's environment using a laser allows to, e.g., select these object to be fetched by an assistive robot.
In this paper, we investigate the influence of stimulus subclasses—differing surfaces between objects as well as stimulus position within a sequence—on the electrophysiological response and the decodability of visual event-related responses. We find that evoked responses differ depending on the subclasses. Additionally, we show that in the presence of ample data subclass-specific classifiers can be a feasible approach to address the heterogeneity of responses.

INTRODUCTION

In the context of assistive robotics, brain-computer interfaces (BCIs) can offer impaired users means to supervise or control the robot. Brain signals can be used to directly move the robot [1] or for high-level commands in a shared-control setting [2]. While user goals in robotic applications mostly correspond to objects or positions in the environment (e.g., which object to fetch and deliver to the user), most existing BCI setups require an auxiliary interface for stimulus presentation.

This interface is typically implemented using a graphical user interface on a screen, but auditory or haptic interfaces are also possible [3–5]. Actions in the world cannot be executed directly using the BCI but need to have a representation in this interface. In BCI paradigms based on visual event-related potentials (ERPs) [6], on visual steady-state evoked potentials [7], motion-onset visual evoked potentials [8] or visual noise codes [9], the screen has been considered an enabling, indispensable building block of the experimental paradigms. But even in BCI paradigms based on mental imagery tasks, the translation of these tasks into actions requires a supporting interface. The indirection step introduced by the interface, however, comes at a price. First, the user is required to switch attention between the screen representation and the real world in order to deliver commands. Second, the interface creates additional mental workload, as the user has to carefully plan a mapping from a desired real-world action into a sequence of potentially unrelated tasks.

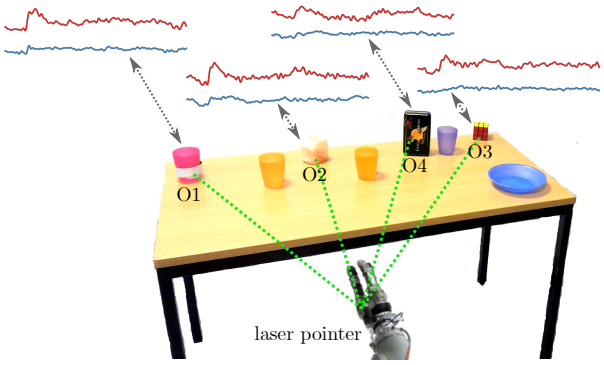

Figure 1: Overview of the experimental setup. In each condition, four objects were highlighted repeatedly with a laser pointer. Annotated objects and average responses correspond to condition C_{250}, the remaining objects correspond to condition C_{500}.

To address both issues, we recently proposed to select objects based on visual ERPs in a novel screen-free interaction [10]. We use a robotic arm to highlight objects *in the scene*. We attach a laser pointer to the arm to generate the visual stimuli. This novel screen-free approach circumvents the aforementioned indirection step, avoids frequent attention switches and may therefore reduce cognitive workload. In short, it allows for a more "intuitive" BCI usage.

While we could show the feasibility of the novel approach [10], the screen-free paradigm introduces other challenges. First, we found that decoding performance requires improvement specifically in the presence of heterogeneous candidate objects and surfaces (see Figure 1).

Second, as the novel paradigm interacts with the real world, it entails different experimental constraints than

 DOI: 10.3217/978-3-85125-682-6-42

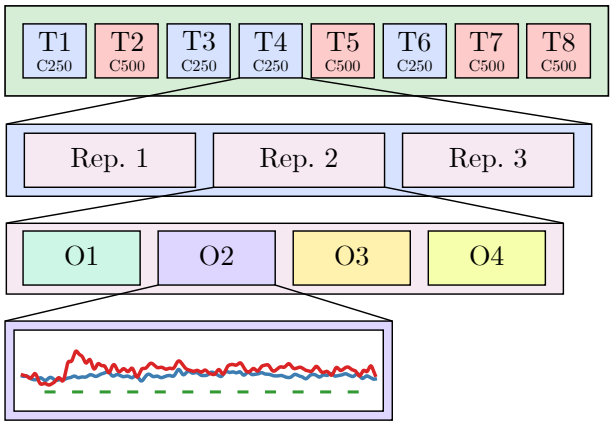

Figure 2: Structure of a single experiment block. Trials were alternated between two conditions (C_{250} and C_{500}) (row 1) in a pseudo-randomized manner. Each trial consisted of three repetitions (row 2) in which each object (row 3) was highlighted for a single sequence of 6 or 12 stimuli (row 4).

traditional screen-based ERP paradigms. Most importantly, the repositioning of the robotic arm between highlighting different objects as well as differing optical properties of stimulus objects necessitates a careful consideration of non-i.i.d. effects in the obtained ERP data.

Previously, the class-wise distributions of auditory ERPs with regard to individual stimuli within a sequence have been found to differ based on the position in the sequence, duration as well as target-to-target intervals [11]. Investigating the influence of subclasses (e.g., stimulus identity), Höhne and colleagues have proposed to incorporate subclass information into the decoding approach by selecting different shrinkage targets for Linear Discriminant Analysis (LDA) [12].

In the following sections, we characterize the influence of *within-sequence stimulus position* and *object surface properties* on the resulting evoked responses using data from six subjects who participated in an in-the-scene object selection paradigm. We present corresponding results in the form of decoding performances using a classification pipeline utilizing Riemannian geometry in differing data size regimes and discuss implications for usage in a BCI.

MATERIALS AND METHODS

Experimental Paradigm and Data Collection:

We recorded the electroencephalogram (EEG) data from six healthy subjects (aged 26 ± 3 years), who each participated in a single session of an experiment. None of the participants had previously used a BCI. Following the declaration of Helsinki, we received approval by the local ethics committee for this study and obtained written informed consent from participants prior to the session.

Subjects were seated in front of a table on which eight objects, denoted O1 to O8, were placed. A Kuka iiwa robotic arm outfitted with a laser pointer next to the end effector was positioned in front of the table and used to highlight objects (c.f., Figure 1). Each object was highlighted either six or twelve times before the robotic arm was repositioned to highlight the next object. Switching the arm position after each single highlighting event was prohibitive considering the delay introduced by the movement time of the robotic arm.

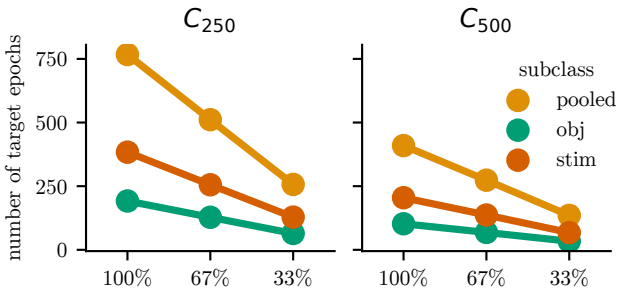

Figure 3: Number of target epochs per subclass based on condition (C_{250} and C_{500}) and data sizes (100 % to 33 %).

The surface types—and therefore, optical properties—of objects O1 to O4 differ from each other, while objects O5 to O8 are made of the same material (see Figure 1). Highlighting O1 resulted in a diffuse and partly specular reflection (away from the subject), whereas highlighting the other objects resulted in diffuse reflections only. Subjectively, O2 showed the strongest reflection among these three objects, the reflection intensity of O3 was reduced and the reflection intensity of O4 was weakest.

We acquired the brain signals using a cap holding $n_s = 31$ Ag/AgCl gel-based passive EEG electrodes positioned according to the extended 10-20 system with a nose reference. We kept channel impedances below 20 kΩ. The amplifier sampled the EEG signals at 1 kHz.

The experiment consisted of six blocks, each containing eight trials. Within a trial, we defined one out of four candidate objects as the target object. As depicted in Figure 2, in each of three repetitions per trial all four candidate objects were highlighted with a stimulus sequence that lasted 3 s per object. For each highlighting, the laser pointer was activated for 100 ms.

Trials were divided into two *conditions*: In condition C_{250}, the candidate objects consisted of O1 to O4 (i.e., they are heterogeneous) and highlighting stimuli were presented with a stimulus-onset asynchrony (SOA) of 250 ms (i.e., 12 stimuli per sequence). In the condition C_{500}, O5 to O8 were the (homogeneous) candidate objects and stimuli were presented with an SOA of 500 ms (i.e., 6 stimuli per sequence). The latter condition directly corresponds to the experiments performed in [10].

Target objects were balanced, i.e., each object was a target once per block. To assure this, a cue denoting the target object was presented to the user before each trial. We instructed users to attend the target object for the duration of the trial without mandating visual focus. After each trial, feedback on the decoded object was given (c.f., [10]). In a post-experiment questionnaire, subjects gave feedback on the experiment task using visual analogue scales.

 DOI: 10.3217/978-3-85125-682-6-42

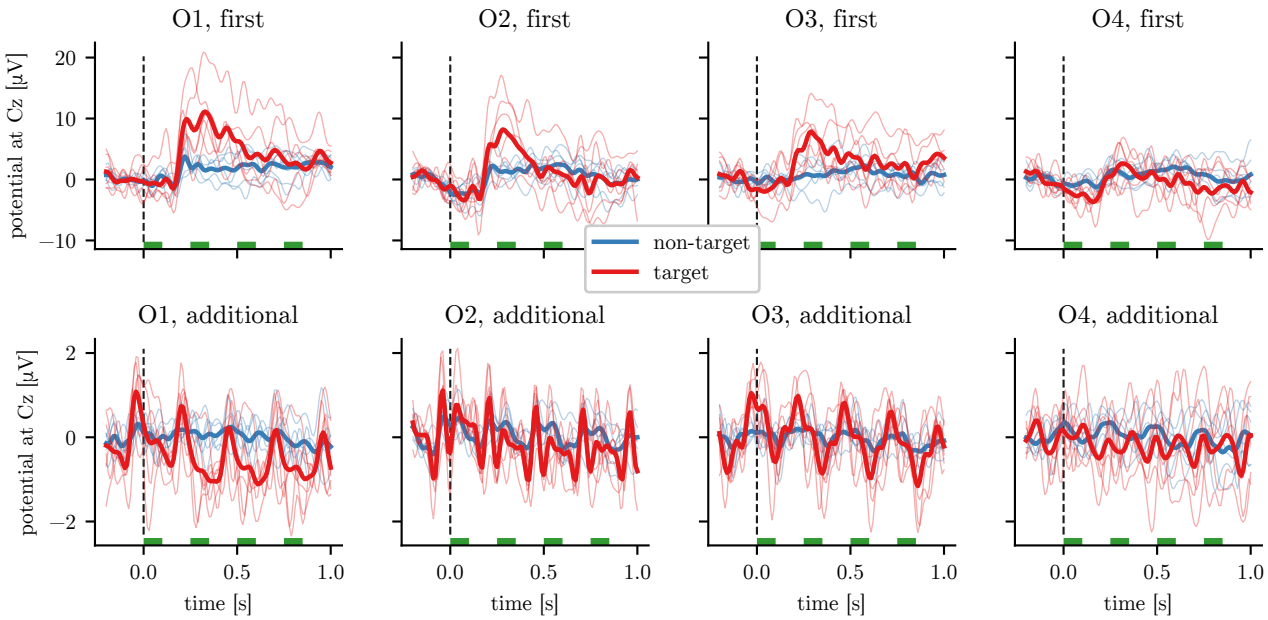

Figure 4: Evoked responses in condition C_{250} by object type (O1–O4) and stimulus position, i.e., first in sequence or additional (stimuli 2–12). The thin lines represent the class-wise average for each subject and the thick lines the grand average over all six subjects. Note the differing y-axis limits between rows. In the top row, each response is averaged based on 108 target and 324 non-target epochs, respectively. In the bottom row, averages correspond to 1,188 and 3,564 epochs. Stimulation intervals are marked with green bars.

Data Analysis:

We analyzed the data separately for each subject and each condition (SOA) in an offline manner. We filtered data to a band of 0.50 Hz to 16 Hz using a FIR filter before downsampling to 100 Hz.

We extracted epochs from -0.20 s to 1.00 s relative to each stimulus onset (leading to $n_s = 121$ samples per epoch) and subsequently corrected them separately for signal drifts using the first 0.20 s as a baseline. Hence, each epoch X_i can be associated with a class $y(i) \in \{target, non\text{-}target\}$, SOA $s(i) \in \{250, 500\}$, object $o(i) \in \{1, \dots 8\}$ as well as the position within the stimulus sequence $p(i) \in 1, \dots 12$. Since we expect most variation between the first and subsequent stimuli [11], stimulus positions can be aggregated to $\tilde{p}$ with $\tilde{p}(i) = first$ if $p(i) = 1$ and $\tilde{p}(i) = additional$ for $p(i) > 1$. We rejected epochs in which the peak-to-peak amplitude exceeded 100 µV in any channel.

For the epoch-wise decoding of the target/non-target response, we used a covariance-based feature representation [13, 14]. On training data, we extracted n_x xDAWN-filters [15] for both target and non-target epochs. Epoch data was projected using these filters and subsequently augmented with prototype responses based on the xDAWN-projected class means in the training data. We calculated the Ledoit-Wolf-regularized covariance of each $4n_x \times n_s$-dimensional augmented epoch separately. These covariances are projected into the Riemannian tangent space at the mean of the training data and subsequently classified using Logistic Regression.

We investigate three different subclass-specific regimes per condition: *pooled*, *obj* and *stim*. In all regimes we performed a chronological 5-fold cross-validation. For *pooled*, we train a single classifier per cross-validation fold. For *obj*, we trained four separate classifiers in each fold based on a partition of the data into objects $\{X_i | o(i) = k\}, k \in \{1, \dots, 4\}$ and classified each test epoch with its object-specific classifier. Similarly, for *stim*, we partitioned the data into $\{X_i | \tilde{p}(i) = l\}, l \in \{first, additional\}$ and trained two different classifiers per fold.

We used $n_x = 3$ components per class for the *pooled* classification and $n_x = 2$ components per class and subclass for *stim* and *obj* classification, resulting in 6 total components for *pooled*, 8 total components for *stim* and 16 total components for *obj*.

Note that this implies different amounts of training epochs per classifier in the three settings. To investigate the general performance degradation based on reduced training data, we performed the above analysis also with 67 % and 33 % of the data by only keeping the first two or only the first repetition of each trial, respectively. The size of the resulting data sets is depicted in Figure 3. We evaluate classification performance on an epoch level (i.e., keeping proportions between data partitions) using the area under the receiver-operating characteristic (AUC).

RESULTS

User Feedback:

All of the subjects were able to attend the stimuli and none reported difficulties perceiving the laser highlighting. On visual analogue scales (“easy” to “demanding”),

 DOI: 10.3217/978-3-85125-682-6-42

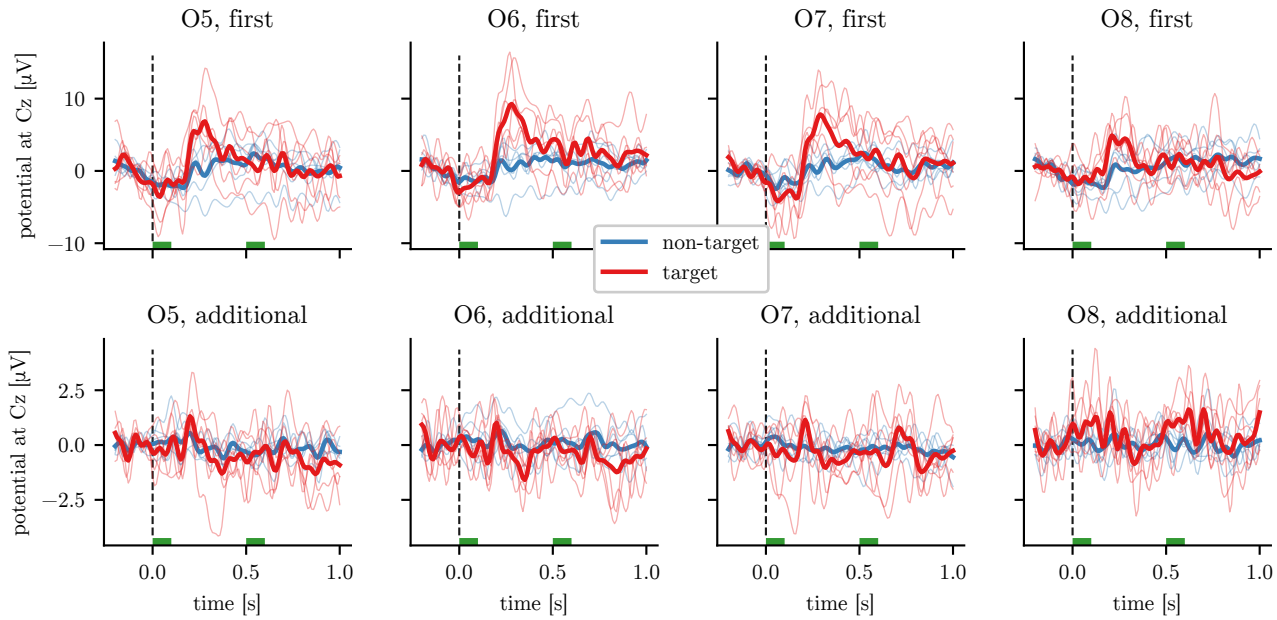

Figure 5: Evoked responses in condition C_{500} by object (O5–O8) and stimulus position, i.e., first in sequence or additional (stimuli 2–6). The thin lines represent the class-wise average for each of subject and the thick lines the grand average over all six subjects. Note the differing y-axis limits between rows. In the top row, each response is averaged based on 108 target and 324 non-target epochs, respectively. In the bottom row, averages correspond to 540 and 1,620 epochs. Stimulation intervals are marked with green bars.

users rated the task's mental demand with 26 ± 26 %, the physical demand with 52 ± 28 % and the temporal demand with 21 ± 21 %. The required effort was rated with 56 ± 31 % on a scale from "low" to "high".

Electrophysiology of Responses:

In both conditions, we observed strong visual evoked potentials which differ between target and non-target stimuli, most prominently approximately 200 ms after stimulus onset (see Figure 4 and Figure 5). The first stimulus in each sequence on average evoked higher amplitudes than subsequent ones, with initial amplitudes surpassing 10 µV compared to the preceding baseline for some objects.

The evoked responses for condition C_{250} with heterogeneous objects are shown as thick lines in Figure 4. Different objects resulted in differing grand average ERP waveforms. Specifically object O4—having a black surface—resulted in a much smaller amplitude, with target responses for later stimuli having similar amplitudes as non-target responses for the other objects. Looking at individual subjects (shown as thin lines in Figure 4), ERP latencies are similar for objects O1 to O3 yet appear to vary substantially between subjects for O4.

In condition C_{500} with homogeneous object surfaces, initial stimuli similarly evoked higher potentials than subsequent ones, yet the difference is smaller than in condition C_{250} (see Figure 5). The grand average responses to the four different but homogeneous objects O5–O8 have a higher similarity than the ones in C_{250}.

Decoding results:

The classification of single epochs into targets and non-targets is possible with a mean AUC of 0.81 for C_{250} and 0.86 for C_{500}, as determined by cross-validation on the *pooled* data sets of the respective conditions. Analyzing the classification performance in this setting on all epochs corresponding to a single object, we found that it is similar for all (homogeneous) objects in C_{500}, while it differed substantially between (heterogeneous) objects in C_{250} (in decreasing order of performance: O1, O2, O3, O4).

As the ERP responses differed for object surface types (O1–O4) and sequence positions, we investigated partitioning the data into different subclasses. As this involved a reduction of training data per sub-classifier (c.f., Figure 3), we also examined the influence of data set sizes.

Partitioning based on sequence position: We found that treating stimulus position ("first" vs "additional") with specialized classifiers did not yield a clear performance improvement or decrease for any of the three data set sizes (100 %, 67 % and 33 %, see Figure 6).

Partitioning based on object: Training separate classifiers for each object in condition C_{250}, as shown in the first column of Figure 6, increased the AUC from 0.81 to 0.84, whereas we observed a performance decrease in the C_{500} setting with homogeneous objects (from 0.86 to 0.81). We found that the improvements obtained with object-specific classifiers for heterogeneous objects (C_{250}) depend on the availability of sufficient training data. As shown in the top row of Figure 6, a reduction of data size to 67 % reduced performance of the *obj* classifier to 0.83, and on 33 % the *obj* classifier performed worse than the *pooled* one (0.78 compared to 0.79). Looking at classification performance for individual epochs on full data, we observed gains for all objects, with greatest gains for O1. In condition C_{500} with homogeneous objects, the performance differences between classifiers did not change with the different data sizes evaluated.

 DOI: 10.3217/978-3-85125-682-6-42

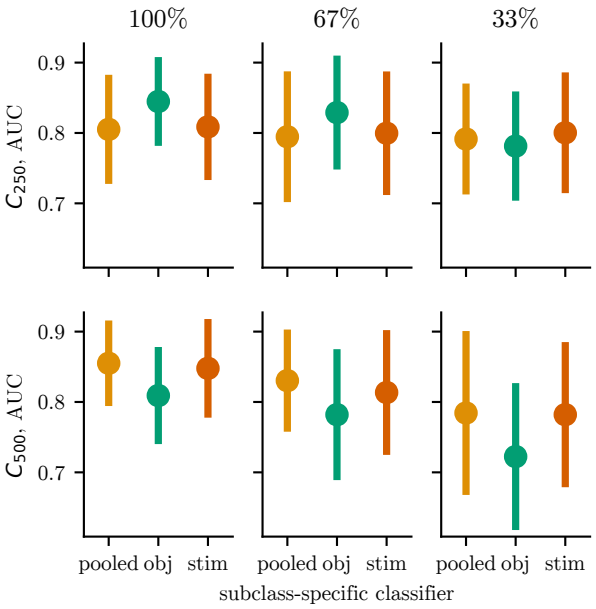

Figure 6: Classification performances of subclass-specific classifiers for different training data sizes (column) as well as SOA and object heterogeneity (rows). In the *pooled* case, data is not partitioned and a single classifier is trained. The means and standard deviations over the six subjects are reported.

DISCUSSION

Highlighting candidate objects directly reduces the levels of indirection of BCIs and can therefore entail greater usability. Yet it comes at the price of reduced control over stimulus presentation compared to traditional screen-based paradigms. This manifests itself in heterogeneous ERPs when object surfaces differ and similarly in distinct responses to initial stimuli (compared to subsequent ones). Whereas stimulus presentation is easily modifiable in screen-based solutions, it is neither desirable nor feasible to adapt properties of objects in the scene—leaving the question of how to handle the heterogeneity.

If maximal performance is not required, the non-i.i.d. characteristics of the data can be ignored, since using a single classifier based on the pooled data (i.e., ignoring the subclass information) still allows an effective decoding on average (however, performance on difficult subclasses suffers). The higher performance for C_{500} compared to C_{250} in the pooled setting can be attributed partly to the homogeneous object surfaces and partly to the larger SOA, which implies a larger fraction of (stronger) responses to initial stimuli in this condition.

Addressing the differing surfaces in the condition C_{250}, object-specific classifiers yield an improvement in AUC. This improvement by specialization is particularly interesting considering the reduced amount of training samples available for each classifier (c.f., Figure 3). The lack of heterogeneity in surfaces is the likely reason for the subpar performance of object-specific classifiers in condition C_{500}, since in condition C_{250} even reduced data allows for improvement compared to the pooled one.

Contrasting the improvement using object-wise classifiers to the lack thereof for sequence-specific ones, two observations can be made: First, the aforementioned reduced data size and therefore higher variance of subclassifiers can offer an explanation for the lack of improvement since the minority class (first stimulus) makes up only 17 % (C_{500}) or 8 % (C_{250}) of the data. Second, the (pooled) covariance-based features (as opposed to utilizing mean potentials) might be more robust to variations in amplitude (as observed between sequence positions) compared to variations in waveforms (for heterogeneous objects), avoiding the need for specialization in the former case.

Hence, subclass-specific classification of heterogeneous responses is feasible and can offer improved performance in the presence of highly differing ERPs and sufficient data. They are, however, not a “silver bullet” solution to this problem, and—without automatic relevance selection—inspecting the subclass data is necessary to identify suitable subclass partitions.

CONCLUSION

Utilizing in-the-scene stimuli for a brain-computer interface implies an influence of scene properties on observed brain responses. In this paper, we find that heterogeneous object surfaces as well as stimulus position within a sequence result in differing ERPs. Addressing this with subclass-specific classifiers can offer improved decoding performance, with the caveat of increased data requirements as well as unfavorable performance in the absence of subclass differences. For future work it would be interesting to automatically infer relevant subclasses and to investigate approaches capable of handling low-data scenarios [12].

ACKNOWLEDGMENTS

This work was (partly) supported by BrainLinks- BrainTools, Cluster of Excellence funded by the German Research Foundation (DFG, grant number EXC 1086). Additional support was received from the German Research Foundation through grant INST 39/963-1 FUGG and from the Ministry of Science, Research and the Arts of Baden-Württemberg for bwHPC.

REFERENCES

[1] Millan J., Galan F., Vanhooydonck D., Lew E., Philips J., Nuttin M. Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009. Sep. 2009, 3361–3364.

[2] Burget F., Fiederer L. D. J., Kuhner D., *et al.* Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. In: 2017 Euro-

 DOI: 10.3217/978-3-85125-682-6-42

pean Conference on Mobile Robots (ECMR). Sep. 2017, 1–6.
[3] Höhne J., Schreuder M., Blankertz B., Tangermann M. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Front. Neurosci.. 2011;5.
[4] Schreuder M., Blankertz B., Tangermann M. A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue. PLoS ONE. 2010;5(4):e9813.
[5] Waal M. van der, Severens M., Geuze J., Desain P. Introducing the tactile speller: An ERP-based brain–computer interface for communication. J. Neural Eng.. 2012;9(4):045002.
[6] Bin G., Gao X., Wang Y., Hong B., Gao S. VEP-based brain-computer interfaces: Time, frequency, and code modulations [Research Frontier]. IEEE Computational Intelligence Magazine. 2009;4(4):22–26.
[7] Chen X., Wang Y., Nakanishi M., Gao X., Jung T.-P., Gao S. High-speed spelling with a noninvasive brain–computer interface. PNAS. 2015;112(44):E6058–E6067.
[8] Guo F., Hong B., Gao X., Gao S. A brain–computer interface using motion-onset visual evoked potential. J. Neural Eng.. 2008;5(4):477–485.
[9] Thielen J., Broek P. van den, Farquhar J., Desain P. Broad-Band Visually Evoked Potentials: Re(con)volution in Brain-Computer Interfacing. PLOS ONE. 2015;10(7):e0133797.
[10] Kolkhorst H., Tangermann M., Burgard W. Guess What I Attend: Interface-Free Object Selection Using Brain Signals. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Oct. 2018, 7111–7116.
[11] Hübner D., Tangermann M. Challenging the assumption that auditory event-related potentials are independent and identically distributed. In: Proceedings of the 7th Graz Brain-Computer Interface Conference 2017. Verlag der Technischen Universität Graz, 2017, 192–197.
[12] Höhne J., Bartz D., Hebart M. N., Müller K.-R., Blankertz B. Analyzing neuroimaging data with subclasses: A shrinkage approach. NeuroImage. 2016;124:740–751.
[13] Barachant A., Bonnet S., Congedo M., Jutten C. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. IEEE Transactions on Biomedical Engineering. 2012;59(4):920–928.
[14] Barachant A., Congedo M. A Plug&Play P300 BCI Using Information Geometry. ArXiv:1409.0107 [cs, stat]. 2014.
[15] Rivet B., Souloumiac A., Attina V., Gibert G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.