Decoding Perceived Hazardousness from User's Brain States to Shape Human-Robot Interaction

Henrich Kolkhorst
University of Freiburg

Michael Tangermann
University of Freiburg

Wolfram Burgard
University of Freiburg

ABSTRACT

With growing availability of robots and rapid advances in robot autonomy, their proximity to humans and interaction with them continuously increases. In such interaction scenarios, it is often evident *what* a robot should do, yet unclear *how* the actions should be performed. Humans in the scene nevertheless have subjective preferences over the range of possible robot policies. Hence, robot policy optimization should incorporate the human's preferences. One option to gather online information is the decoding of the human's brain signals. We present ongoing work on decoding the perceived hazardousness of situations based on brain signals from electroencephalography (EEG). Based on experiments with participants watching potentially hazardous traffic situations, we show that such decoding is feasible and propose to extend the approach towards more complex environments such as robotic assistants. Ultimately, we aim to provide a closed-loop system for human-compliant adaptation of robot policies based on the decoding of EEG signals.

1. INTRODUCTION

With the advent of autonomous robots in a multitude of different environments as well as proximity and cooperation of humans and robots, multiple aspects of robotics, such as predictability of robotic behavior and compliance to the human's preferences are of increasing importance. In isolated environments, robot policies are typically optimized with regard to technical objectives (e.g., time for trajectory execution). However, a human might have different preferences (e.g., *perceived* complexity or safety).

While substantial work has been done to improve human-robot interaction by actively demonstrating desired behavior, such as socially-compliant navigation [7], it is often desirable to gather human preferences in a passive way. Several publications address inferring the user's goal from passive observations [6, 1], yet it poses challenging to infer *how* such a goal should be achieved.

This research was (partly) supported by the German Research Foundation (DFG, grant number EXC 1086).

HRI '17 Companion March 06-09, 2017, Vienna, Austria

ACM ISBN 978-1-4503-4885-0/17/03.

DOI: http://dx.doi.org/10.1145/3029798.3034826

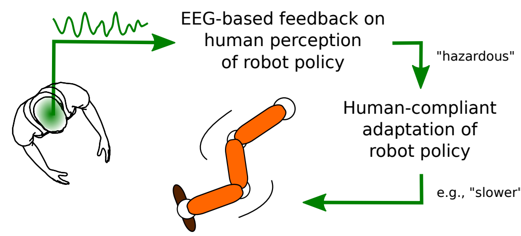

Figure 1: Schema of human-compliant robot adaptation based on brain signals with simple feedback.

Such desired properties of robot policies have been investigated, e.g., in the areas of human-robot handovers [9] or legibility of motion for observers [2]. Perception of properties such as hazardousness, however, differs between individuals and contexts. Hence, in many situations, robotic systems need to adapt their policies in an online fashion with the human in the loop. Gathering online feedback during execution is, however, very difficult, since asking for explicit user input can be distracting from the main task. Therefore, we propose to gather feedback in a passive way, using brain signals in the form of non-invasive electroencephalography (EEG) to extract information on the human scene perception, which can be utilized for improving robot policies (c.f., Figure 1).

Brain-computer interfaces (BCIs) are capable of estimating (noisy) user intentions (e.g., for device control [3, 10]) or decoding a user's brain state in single trial (e.g., for emergency braking [5]). BCI output is graded and can deliver a continuous stream of decoding results, which can be used for adapting human-computer interaction, e.g., modifying task difficulty based on detected workload [8].

In human-robot interaction, the application areas of autonomous driving and robotic assistants are particularly relevant due to the large number of non-expert robot users over the coming year. In autonomous driving, functional goals of reaching a target location leave a wide spectrum of (subject-dependent) choices regarding preferred driving style, e.g., speed or preferred following distance. Similarly, optimization of robotic assistants in workplace or assistive scenarios might face trade-offs between, e.g., minimizing execution time of trajectories and maximizing distance to the human user to reduce perceived hazardousness of the robot.

In the remainder of this paper, we focus on human perception of traffic scenes, e.g., as passengers in an autonomous car. We investigate whether it is possible to distinguish hazardous and non-hazardous events in traffic scenes based on brain signals.

2. METHODS AND RESULTS

In order to gather brain responses to events with varying hazardousness, we performed experiments with eleven subjects in a lab setting. Subjects watched realistic stimulus material in the form of videos (10 frames per second) of 48 traffic scenes from the KITTI dataset [4]. Subsections of scenes were modified with pictograms in order to add more potentially hazardous situations. Subjects were told they were passengers in an autonomous car and instructed to keep track of potentially dangerous situations and to press a button in these situations to relay this information on hazardousness to a vehicle. The traffic scenes contained 443 potentially hazardous candidate events such as a pedestrian appearing on the sidewalk (non-hazardous) or a child entering the street from an occlusion (hazardous, c.f. Figure 2a). EEG signals were recorded at 1 kHz using 63 passive Ag/AgCl electrodes.

For the analysis of hazardousness of events, intervals of 1000 ms aligned to the first visible frame of the potential hazard in the scene were extracted. After automatically rejecting intervals with artifacts (e.g., eye blinks), we trained a classifier (linear discriminant analysis) to classify single windows in single trial into hazardous and non-hazardous events based on electrode potentials in discriminative sub-intervals. Classifier training and evaluation have been performed in a 5-fold chronological cross-validation.

Analyzing the brain signals, we observe event-related reactions for both classes of events yet find them to differ substantially between classes. For hazardous compared to non-hazardous events, we observe for most subjects an early negative deflection primarily at occipital electrodes and a later positive deflection at central electrodes. Class-wise averages over all events of two EEG channels for a subject with average classification performance are depicted in Figure 2b. Visualizing all electrode locations for two sub-intervals of the 1000 ms window (Figure 2c) shows that the class-discriminative information can be observed in a large number of channels.

Single-trial classification of events into hazardous and non-hazardous based on the 1000 ms windows was evaluated using the area under the curve of the receiver operating characteristic (AUC). The classification yielded a mean AUC over all subjects of 0.73 with a standard deviation of 0.05.

3. DISCUSSION AND FUTURE WORK

The performed experiments suggest the feasibility to decode perceived hazardousness from the brain signals of human observers. Discriminative information can be observed early, providing the opportunity to utilize this information in online systems. The experiments were, however, performed in a controlled environments and under the assumption that the timing of candidate events is known.

Building on these results, we plan to next extend the work in two directions. First, we plan to decode the perceived scene complexity as a continuous feedback metric by the human. Such a score can be utilized to adapt behavior (such as reducing execution speed or modifying planning cost functions to minimize the perceived complexity). Second, we aim to investigate the perceived hazardousness in the more complex domain of robotic assistants in collaborative environments. Here, we plan to directly associate brain response with varying robot policies or cost functions.

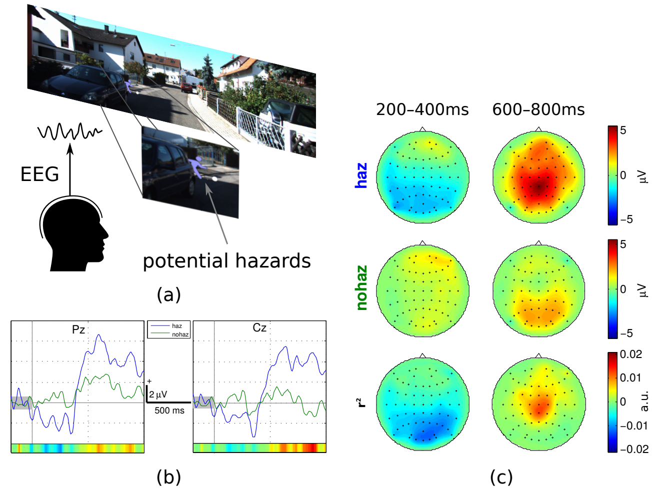

Figure 2: (a) Experimental setup. (b) Exemplary averaged event-related response of one subject (electrodes Pz, Cz), aligned to the first frame of the event (N=70+305). Hazardous events (blue) cause a distinct response. (c) view of all electrodes of the same subject in intervals 200ms and 600ms after the event, showing (top to bottom) hazardous and non-hazardous events and the discriminability (r^2) .

Utilizing the user's brain signals, we expect to reduce uncertainty over the user's preferences regarding robot policies. Having a better estimate of the preferred behavior, we plan to incorporate this into the policy optimization for adaptive human-compliant policies. Ultimately, we envision brain state decoding as a non-intrusive online feedback channel which can provide a teaching signal as well as an evaluation measure for the quality of human-robot interaction.

4. REFERENCES

[1] A. Bestick, R. Bajcsy, and A. D. Dragan. Implicitly Assisting Humans to Choose Good Grasps in Robot to Human Handovers. In *ISER*, Tokyo, 2016.

[2] A. Dragan and S. Srinivasa. Integrating human observer inferences into robot motion planning. *Auton Robot*, 37(4):351–368, 2014.

[3] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán. A brain-actuated wheelchair: Asynchronous and non-invasive Brain–computer interfaces for continuous control of robots. *Clinical Neurophysiology*, 119(9):2159–2169, Sept. 2008.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets Robotics: The KITTI Dataset. *The International Journal of Robotics Research*, page 0278364913491297, Aug. 2013.

[5] S. Haufe, J.-W. Kim, I.-H. Kim, A. Sonnleitner, M. Schrauf, G. Curio, and B. Blankertz. Electrophysiology-based detection of emergency braking intention in real-world driving. *J. Neural Eng.*, 11(5):056011, Oct. 2014.

[6] S. Javdani, S. Srinivasa, and A. Bagnell. Shared Autonomy via Hindsight Optimization. In *RSS*, Rome, 2015.

[7] M. Kuderer, H. Kretzschmar, and W. Burgard. Teaching mobile robots to cooperatively navigate in populated environments. In *IROS*, pages 3138–3143, 2013.

[8] M. Schultze-Kraft, S. Dähne, M. Gugler, G. Curio, and B. Blankertz. Unsupervised classification of operator workload from brain signals. *J. Neural Eng.*, 13(3):036008, 2016.

[9] K. W. Strabala, M. K. Lee, A. D. Dragan, J. L. Forlizzi, S. Srinivasa, M. Cakmak, and V. Micelli. Towards Seamless Human-Robot Handovers. *J of Human-Robot Interaction*, 2(1):112–132, 2013.

[10] M. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller. Playing Pinball with non-invasive BCI. In *NIPS 21*, pages 1641–1648. 2009.