Learning User Preferences for Trajectories from Brain Signals

Henrich Kolkhorst, Wolfram Burgard and Michael Tangermann

Abstract Robot motions in the presence of humans should not only be feasible and safe, but also conform to human preferences. This, however, requires user feedback on the robot's behavior. In this work, we propose a novel approach to leverage the user's brain signals as a feedback modality in order to decode the judgment of robot trajectories and rank them according to the user's preferences. We show that brain signals measured using electroencephalography during observation of a robotic arm's trajectory as well as in response to preference statements are informative regarding the user's target trajectory. Furthermore, we demonstrate that user feedback from brain signals can be used to reliably infer pairwise trajectory preferences as well as to retrieve the target trajectories of the user with a performance comparable to explicit behavioral feedback.

1 Introduction

In the vicinity of humans, it is not only important what a robot does, but also *how* these actions are performed. Especially in the context of robotic assistants, trajectories should not only be feasible and free of obstacles, but also comply with the user's preferences. However, preferences over trajectories may vary between users, environments, tasks and also time, which poses challenges to design general cost functions. Instead, it can be beneficial to learn preferences directly from the user.

To learn preferences, input from the user in the form of demonstrations or feedback on candidate trajectories is necessary. Particularly for robots with multiple degrees of freedom or for impaired users, providing trajectory demonstrations may be prohibitive. Giving feedback on the robot's behavior, however, is possible and—especially for relative instead of absolute ratings—does not require expert knowledge.

All authors are with the Department of Computer Science, University of Freiburg, Germany. H.K., W.B. and M.T. are with the Autonomous Intelligent Systems group and H.K. and M.T. are also with the Brain State Decoding Lab. Corresponding author's email: kolkhorst@informatik.uni-freiburg.de

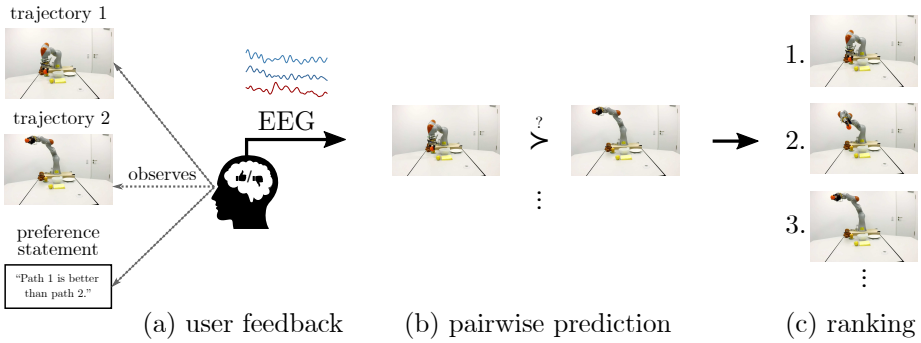

Fig. 1 Overview of our approach: (a) We measure the brain response of users during trajectory observation and in response to preference statements in a pairwise comparison setting using electroencephalography (EEG). (b) We predict pairwise preferences from the EEG data utilizing methods from Riemannian geometry. (c) We combine these predictions to rank the trajectories.

In order to obtain the human judgment of a trajectory, different modalities such as screen-based rating or speech are conceivable. Brain signals as a feedback modality can be desirable because their measurement does not interfere with the primary task and—especially in the context of robotic assistants for impaired users—can potentially be recorded from users who cannot reliably control robots through other modalities. However, signals measured using noninvasive electroencephalography (EEG) typically have an unfavorable signal-to-noise ratio that makes it challenging to utilize this information based on single brain responses.

In this work, we address the problem of learning trajectory preferences from EEG-based user feedback. For this, we let the user observe multiple trajectory pairs and additionally present a—potentially incorrect—preference statement that indicates which of the observed trajectories was supposedly better. This offers two informative sources for discriminative brain signals: As depicted in Fig. 1, we classify the user's brain signals during robot observation and after showing preference statements to predict the actual pairwise preferences of the user. Subsequently, we rank trajectories across pairs using the EEG-based predictions. To facilitate a quantitative evaluation that is comparable across users, we asked them to give feedback according to common target trajectories rather than personal preferences in this work. We compare our approach against explicit user feedback collected from button presses after the preference statements.

Our main contributions are: 1) We propose an approach to estimate user feedback to trajectories by classifying brain signals both during trajectory observation and in reaction to preference statements. 2) We show in experiments with 11 participants, who observed videos of real robot trajectories, that these brain signals are informative about trajectory judgment and that our approach enables reliable decoding of the preferred trajectory in pairwise comparisons in an offline manner. 3) We demonstrate that these pairwise preference predictions based on brain signals can be used to successfully identify the target trajectory from the observed ones with a performance that is comparable to explicit feedback in the form of button presses.

2 Related Work

In order for robots to operate successfully and adequately in human environments, motion planning should incorporate the user's demands. In the past, the modeling of human preferences has been addressed using cost functions [10, 7] and by integrating hand-crafted costs into planning [20]. The use of demonstrations allows to learn reward functions, e.g., from navigation behavior [17] or corrective actions for robot manipulators [2].

Rather than modeling costs manually or providing near-optimal demonstrations, behavioral feedback can also be used to improve robotic actions. For example, legibility can be optimized by minimizing the time until the user can press a button corresponding to the anticipated goal of the robot [6]. While absolute ratings of robotic actions would be desirable for ranking a set of options, it is easier for human users to give relative feedback by comparing a small set of items. Relative feedback has been used to model, e.g., human perception of "naturalness" of robot configurations [12] or preferences in simulated driving [23]. While many approaches assume label noise in the user feedback [1, 23], it is typically small and assumed to vary based on the reward differences. In contrast, strong measurement noise is typically encountered when using EEG signals. Closely related to our setting is the work by Jain et al. [11], which aims at learning preferences based on human feedback. However, they utilize screen-based reranking of trajectories and kinesthetic teaching as user feedback.

Brain–computer interfaces (BCIs) based on EEG signals are typically either driven by mental imagery of the user [5] or external stimuli [9, 25]. In reactive attention-based BCIs, discriminative information can be obtained from differing responses—mostly in the form of event-related potentials (ERPs)—to stimulus classes, such as identifying "surprising" outlier stimuli or stimulus changes [25]. Due to the low signal-to-noise ratio, prediction is typically limited to binary classification. A large body of different classification approaches specifically tailored to EEG signals exists [19]. State-of-the-art results increasingly utilize methods from Riemannian geometry [28, 3]: Assuming that relevant information about the mental state in a given time interval can be represented by the covariance matrix, the corresponding manifold structure suggests using non-Euclidean distance measures between data points. Generally, classification performance heavily depends on the experimental task—implying the mental states to be classified—and varies from user to user.

Many common BCIs are based solely on the appearance or the identity of an attended visual stimulus, whereas feedback on trajectories requires the decoding of the user's *judgment* of an action in a context. Error-related potentials [8], i.e., brain responses to committed or observed errors, form one common type of brain responses useful for classification. Error-related potentials are interesting since they are based on judgment of behavior, but typically also require its fast comprehension by the user. Brain activity from time windows that are not time-aligned to specific stimuli can also be informative—e.g., to predict workload [26] or upcoming task performance [21]. Although desirable for applications, predictive performance in such asynchronous settings is typically lower than in stimulus-aligned ones.

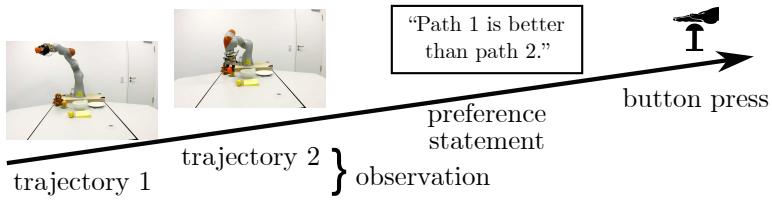

Fig. 2 Course of pairwise trajectory comparison: The user observes videos of the robot executing two trajectories. Subsequently, we present the user with a (potentially incorrect) *preference statement* of the form "Path 1 was better than path 2." (or vice versa). For evaluation purposes, after a delay we also ask the user to press a button if she agrees with the statement.

Recent work has incorporated brain responses to robotic behavior, e.g., for error recognition [24] or learning gesture mappings with reward signals from EEG [14]. In these contexts, the human judgment of the behavior could be performed near-instantaneously, greatly easing the temporal alignment of EEG signals for classification. However, in the case of observing more complex behavior (e.g., trajectories that have an identical start configuration but continuously deviate thereafter), the temporal alignment—i.e., the time point the human realizes one behavior is superior or inferior—cannot easily be inferred.

3 Decoding Trajectory Preferences from Brain Signals

We consider the problem of identifying a user's preferred trajectory based on feedback by the user: Given a set of trajectories $\Xi = \{\xi_i\}, i = 1, \ldots, N$, we want to find a *target trajectory* $\xi^* \in \arg\max_{\xi \in \Xi} R_h(\xi)$ that maximizes the (unknown) reward $R_h(\xi)$ of a human user in a given environment. In order to estimate ξ^*, we opt to query the user for relative preference feedback and use this feedback—from EEG or from explicit button presses for comparison—to rank trajectories.

After obtaining feedback, our approach consists of two prediction modules, which are depicted in Fig. 1: First, we propose to classify pairwise feedback to trajectories from brain signals during robot observation and in response to preference statements as described in Section 3.1. We utilize the covariance representations of signals, using methods from Riemannian geometry. Second, we use this feedback to rank trajectories both in an instance-based and a feature-based manner (see Section 3.2). Note that training of both modules is independent, i.e., classifiers for EEG decoding are not specific to different environments or preferences and only depend on the user.

In order to obtain pairwise preference feedback from a user, we perform a sequence of trajectory comparisons $C = \{c_1, \ldots, c_{N_c}\}$. Each comparison c_j consists of two trajectories $\xi_{i_{j,1}}, \xi_{i_{j,2}}$ that are presented to the user (i.e., $c_j = \{i_{j,1}, i_{j,2}\}$). To obtain comparative judgment of the two trajectories, we propose to use *preference statements*: After presenting a pair of trajectories, we show a textual statement to the user in the form of "Path 2 is better than path 1."

The correctness of the statement implies a preferred trajectory and should also induce differing decodable mental states. For control, we subsequently ask the user to press a button if she judges this statement to be correct (c.f., Fig. 2). Hence, for each comparison c_j, we observe a behavioral response by the user in response to the statement (i.e., a button press if the statement is correct), which implies the user's pairwise preference.

3.1 Decoding Pairwise Preferences from Brain Signals

We have two potentially informative sources of user feedback based on brain responses in each comparison: the brain activity during watching the trajectory execution by the robot—the *observation* setting—and the response to a subsequent *statement* (e.g., "Path 1 is better than path 2."). As discussed in Section 2, the latter synchronous case is better suited for classification using brain signals, whereas the former asynchronous one would be desirable to collect information passively during fluent human–robot interaction. We train separate classifiers for both types of signals as well as a combined one. As input, we extract features from fixed-sized windows of the continuous frequency-filtered EEG signal.

3.1.1 Segmentation and Labeling of Brain Signals

For the *observation* setting, we use a 2 s window temporally centered in the trajectory execution, leading to feature vectors $X^o_{j,k} \in \mathbb{R}^{N_{\text{ch}} \times N^o_s}$ for trajectory $\xi_{i_{j,k}}$ in comparison c_j ($k \in \{1, 2\}$). Here, N_{ch} denotes the number of channels in the recording and N^o_s the duration of the window in samples. The temporal centering encodes our hypothesis that the user judgment evolves, with increasing confidence, during the observation and—with a high uncertainty of the user at the start and a low one at the end—intermediate signals might capture discriminative mental states. We assume supervision in the form of a training dataset for which the user's preference—the target trajectory ξ^*—is known. While we aim to infer the pairwise preference in each comparison, such a comparative judgment is only possible after having observed both trajectories. Hence, comparative labels are not applicable for observation windows. Instead, we opt to use the similarity between observed candidate trajectories and the target trajectory. As a trajectory similarity measure, we use a geometric distance: For trajectory ξ, we temporally normalize trajectory durations to $[0, 1]$ and interpolate trajectory waypoints using cubic splines to get a trajectory representation $f_\xi(t)$, which we use to calculate the distance

$$d(\xi_1, \xi_2) = \int_0^1 \left\| f_{\xi_1}(t) - f_{\xi_2}(t) \right\|_2 \, \mathrm{d}t. \tag{1}$$

Equipped with this distance, we can calculate $d_{\text{target}}(\xi_i) = d(\xi_i, \xi^*)$ for each trajectory in our training data. We obtain binary labels $y^o_{j,k}$ for windows $X^o_{j,k}$ by thresholding $d_{\text{target}}(\xi_{i_{j,k}})$ on the median of all distances in the training data. Other absolute trajectory metrics or ratings could be used alternatively. Given the predicted distance class $\hat{y}^o_{j,k}$ for each of the two trajectories in comparison j ($k \in \{1, 2\}$), we derive the pairwise-preferred trajectory $\hat{i}_j$ based on the smaller of the two.

For the *statement* response, we extract a 1 s window $X^s_j \in \mathbb{R}^{N_{\text{ch}} \times N^s_s}$ for each comparison c_j, starting with the onset of the statement. As a label, we use the correctness of each statement $y^s_j \in \{correct, erroneous\}$ based on d_{target} of the trajectories. Since the statements are comparative, the predicted correctness $\hat{y}^s_j$ implies the pairwise-preferred trajectories $\hat{i}_j$. In the rare cases in which the behavioral response of the user did not correspond to our label (i.e., the button press implied a preference for the trajectory further away from the target), we assumed that the user's brain state also reflected the behavioral judgment and consequently corrected these labels.

We also *combine* both feedback types in addition to classifying observation and statement windows separately. For this, we concatenate the predictions for the statement and the observation windows ($\hat{y}^s_j, \hat{y}^o_{j,k}, \hat{y}^o_{j,k'}$) on the comparison level ($k, k' \in \{1, 2\}$), where each prediction is the output of the separately trained classifiers for each window type. Note that we order the observation-based predictions such that the first coincides with the supposedly better trajectory in the statement. For the combined setting, we use the labels y^s_j from the corresponding statements to train a logistic regression classifier.

3.1.2 Feature Extraction and Classification of Brain Signals

We use covariance-based features for the classification of both observation and statement windows. Since we expect time-locked event-related potentials only after the statement (c.f., Section 2), we perform baselining and augmentation for these windows: We baseline each of the statement windows by subtracting the channelwise average activity in the 200 ms preceding the window and augment them with prototype responses [4]. As prototypes P^+, P^-, we use the mean response for each class in the training data. Additionally, we reduce the channel count by projecting the data using $W^+, W^- \in \mathbb{R}^{3 \times N_{\text{ch}}}$ obtained by selecting three components per class based on the largest eigenvalues from an xDAWN decomposition [22]. This leads to augmented statement windows $\tilde{X^s_j} = \left(W^+P^+,\ \ W^-P^-,\ \ W^+X^s_j,\ \ W^-X^s_j\right)^T \in \mathbb{R}^{12 \times N_{\text{samples}}}$.

For both feedback types ($X^o_{j,k}$ or $\tilde{X^s_j}$), we calculate window-wise covariances C using a Ledoit-Wolf regularization. To account for the symmetric positive-definiteness and the corresponding undesirable properties of the Euclidean distance [28], we project each of the covariance representations into the corresponding tangent space at the Frechét mean C_{ref} of the training samples of the corresponding window type (c.f., [16, 3]):

$$S = \text{logm}\left(C_{\text{ref}}^{-1/2} C C_{\text{ref}}^{-1/2}\right) \tag{2}$$

Here, logm denotes the logarithm of a diagonalizable matrix (i.e., the logarithm of each element of the diagonal after the corresponding decomposition). The upper triangular entries of these projected covariances S are used as input to the classifier. Due to the typically small number of training samples that are available per user, we use L_2-regularized logistic regression classifiers.

3.2 Trajectory Ranking based on Noisy Pairwise Preferences

In order to rank a set of trajectories based on pairwise preferences, we follow two conceptual approaches: In the *instance-based* setting, ranking is performed solely based on the identity of the compared trajectories (i.e., the index i), disregarding any additional (geometric) information about the trajectory. Alternatively, learning to rank can be based on (e.g., geometric) trajectory features $\phi : \Xi \to \mathbb{R}^{N_{\text{feat}}}$. While only the latter *feature-based* approach allows to rank trajectories for which no user feedback has been observed, it depends on the expressiveness of the feature representations ϕ. Hence, we propose both an instance-based and a feature-based ranking approach for combining the pairwise predictions to identify preferred trajectories.

For *instance-based* ranking solely based on the pairwise comparison outcome, we use a modified *Borda counting* method [27]: For each trajectory ξ_i (which is a candidate in the comparisons $J(i) = \{j = 1, \ldots, N_c,\ i \in c_j\}$), we count the number of comparisons in which ξ_i is the predicted pairwise preference, yielding $\hat{R}_{\text{instance}}(\xi_i) = \sum_{j \in J(i)} \mathbf{1}\{\hat{i}_j = i\}$. Here $\mathbf{1}\{\cdot\}$ denotes the indicator function that returns 1 iff the argument is true. While conceptually simple, the method is asymptotically optimal for retrieving the most highly ranked items from noisy user observations [27]. However, in the case of EEG-based pairwise preferences, we also have to account for substantial noise in the predictions, which typically differs between comparisons yet should be correlated to the classifier's predicted probability $\hat{y}$. We propose a heuristic extension to the Borda counting that weighs comparisons based on the confidence $conf(\hat{y}_j) = |\hat{y}_j - 0.5|$:

$$\hat{R}_{\text{instance(conf)}}(\xi_i) = \sum_{j \in J(i)} conf(\hat{y}_j)\mathbf{1}\{\hat{i}_j = i\} \tag{3}$$

For the *feature-based* ranking approach, we follow the common assumption that the reward is linear in the (geometric) feature representation ϕ of a trajectory [23, 11]:

$$\hat{R}_{\text{feat}}(\xi_i) = \boldsymbol{\theta}^T \phi(\xi_i) \tag{4}$$

Hence, the goal is to find a $\boldsymbol{\theta} \in \mathbb{R}^{N_{\text{feat}}}$ such that a pairwise preference of ξ_m over ξ_n in comparison j implies $\boldsymbol{\theta}^T \phi(\xi_m) > \boldsymbol{\theta}^T \phi(\xi_n)$. Consequently, we want the projected feature differences $\boldsymbol{\theta}^T\ (\phi(\xi_m) - \phi(\xi_n))$ to be positive and can use them to train a binary classifier [13]. As labels, we use the predicted pairwise preference relations $\mathbf{1}\{\hat{i}_j = m\}$ based on the brain signals of the user (c.f., Section 3.1.1).

For the feature representation ϕ we include geometric information on the robot movement and on the environment interaction: For the movement, we include the end effector's mean squared velocity, mean squared acceleration and mean and maximal squared jerk as well as mean and maximal joint velocities over the trajectory. For the environment interactions, we use the features proposed in [11] (e.g., minimal distances to scene objects and distance from the goal), yielding a total of $N_{\text{feat}} = 120$ features. Note that also other—potentially more discriminative—feature representations could easily be used instead. We train a L_2-regularized logistic regression classifier on this data and use the classifier predictions to rank trajectories.

4 Experiments

In order to evaluate our approach for learning trajectory preferences from brain signals, we assessed the performance of different feedback types for predicting pairwise preferences as well the inference of target trajectories based on combining these predictions into a ranking. We performed the evaluation using data from experiments with 11 participants. Specifically, we wanted to answer the following questions: (1) Are brain signals as a feedback modality—passively during trajectory observation or reactively after a preference statement—informative about user preferences? (2) Is it possible to classify responses to single comparisons in order to predict the pairwise-preferred trajectory? (3) Can we use these predictions to select a trajectory that is close to the user's target trajectory?

The evaluation design requires a balancing of the diversity of (preference) trajectories against reproducibility, reduction of confounders and comparability across participants. To have identical trajectory execution for all participants, we opted to record videos of a Kuka iiwa robotic arm executing trajectories and showed these videos to the participants. While there is a mismatch between watching videos and observing a robot in the scene, we believe that directly observing the robot would likely be more immersive, and thus might lead to even stronger brain responses. Preference statements could also easily be presented by the robot similar to our setting (e.g., using audio). The recording quality of EEG in the proximity of a robot still allows successful decoding [16].

While it would be desirable to let participants give feedback according to their personal preference trajectories, this can lead to infeasible desired target trajectories (especially for novice users) and differing class distributions in the pairwise preference prediction task, hindering comparability of performance. Crucially, a geometric representation is not available for personal preference trajectories, which precludes a quantitative evaluation of the predicted preference trajectory based on geometric proximity to the ground truth. Therefore, we presented *reference* trajectories to the participants and instructed them to give preference feedback according to these references. To validate the adequacy of the references, we also asked participants to rate the personal agreement with the reference trajectories.

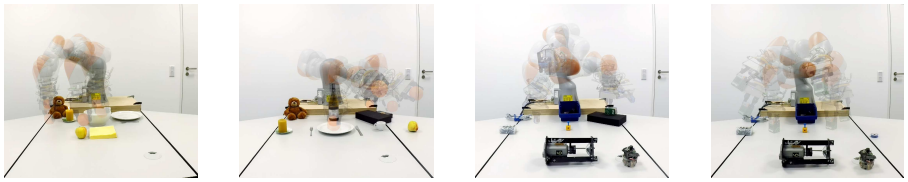

Fig. 3 Examples for the different tabletop environments used in our experiments. To give an intuition about the robot's path, we overlaid screenshots of intermediate configurations.

4.1 Experimental Setup

Each experiment session consisted of 16 *preference tasks* with differing target trajectories. With nine comparisons per target trajectory, each participant observed a total of 288 videos in 144 comparisons. We recorded videos in four different environments with differing objects (resembling assembly or dinner settings as depicted in Fig. 3). In each environment, there were two different pairs of initial and final end-effector poses. For each of these pairs, two dissimilar reference trajectories were selected (e.g., passing on different sides of an obstacle). Each was the target of a single preference task. Hence, all trajectories in a task had identical initial and final end-effector poses and only differed in the intermediate configurations (please see the supplemental video for examples). The trajectory videos had a mean duration of 5 ± 2 s. In order to resemble a realistic robotic setup, a minority of videos also contained "failures" (e.g., the arm touches scene objects or an item is dropped from the gripper). We sampled candidate trajectories using RRT-Connect [18] and selected a subset to assure variance within tasks.

For each preference task (i.e., an experimental block with a single target trajectory), we first showed the reference trajectory. Subsequently, we performed nine comparisons, each consisting of two videos, followed by a preference statement as depicted in Fig. 2. We only prompted for the button response 2 s after the appearance of the statement to reduce a possible influence of motor activity on the brain response. The order of trajectories in the statement was randomized and statements were balanced with respect to the expected behavioral response of the participants. Including pauses between videos and comparisons, one preference task took approximately 7 minutes.

We acquired the brain signals using a cap holding Ag/AgCl gel-based passive electrodes positioned according to the extended 10–20 system with a nose reference. Channel impedances were kept below 20 kΩ. The amplifier sampled the EEG signals at 1 kHz. We used $N_{\text{ch}} = 32$ channels whose signals were frequency filtered to a band of 0.50 Hz to 40 Hz. For the statement responses, we resampled the data to 100 Hz.

We conducted experiment sessions with 14 participants. Following the declaration of Helsinki, we received approval by the local ethics committee and obtained written informed consent from participants. Participants familiarized themselves with the setting in five comparisons that were not analyzed. In data from three of the sessions, we observed a large fraction of artifacts after the statements (more than 30% of windows X^s exceeded a max–min difference of 100 μV in any channel), likely caused by eye blinks.

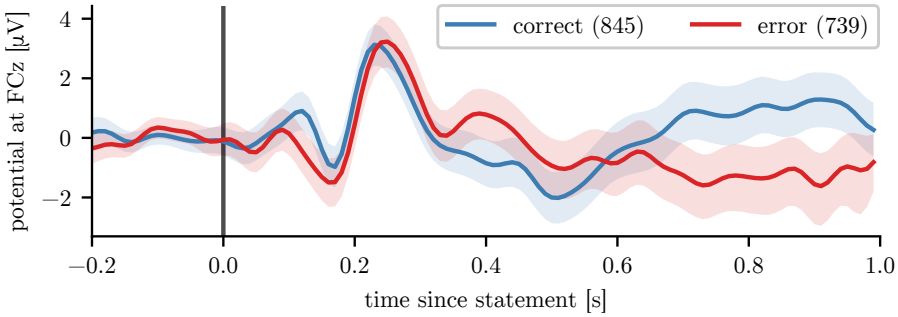

Fig. 4 Grand average response to preference statements for all 11 participants at frontal electrode FCz. Shaded areas correspond to bootstrapped 95% confidence intervals of the mean.

While the EEG data of these sessions still contains discriminative information, classification would likely primarily be based on muscular artifacts rather than brain signals. Hence, we only kept data from the other 11 participants. We did not reject any data of the remaining participants, enabling identical sample counts and ranking tasks. We trained separate classifiers for each participant and evaluated them in a chronological 5-fold cross-validation.

4.2 Results for Pairwise Preference Prediction

Before analyzing ranking performance, we discuss the intermediate pairwise preference prediction. We evaluated it based on the *comparison accuracy* using the trajectory with a smaller d_{target} as the ground truth, which assures identical labels for all participants. Hence, also the behavioral response of the participants did not achieve a perfect score (which is in line with the assumption of noisily rational behavior). Nevertheless, the mean accuracy based on button presses (0.92, see Fig. 5) indicated that the participants followed the task. We asked the last nine participants after the experiment to indicate whether the reference paths agreed with their personal preference. On a visual analog scale from “little” (0) to “much” (1), participants marked an average of 0.69, with seven of nine indicating a tendency to agree (upper half of scale). When asked after each task, participants stated that reference trajectory matched their preference in 78 % of cases (42 % “agree” and 35 % “strong agree”).

4.2.1 Electrophysiology of Statement Responses

As an introspection into the signals used for classification, Fig. 4 shows the classwise average potential in a frontocentral channel of all participants in response to the preference statements (which is better suited for visualization due to the stimulus alignment).

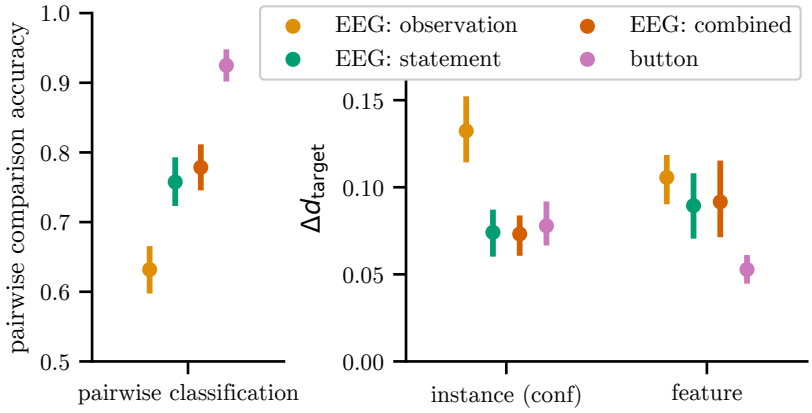

Fig. 5 Mean accuracies for all 11 participants of predictions on a comparison level (left, higher is better) and distance differences Δd_{target} from the target for the top-ranked trajectory for each preference (right, lower is better). Error bars correspond to bootstrapped 95% confidence intervals.

The visually evoked responses after the statement's appearance are similar for both classes until approximately 300 ms. However, we observed a second positive deflection only for erroneous statements approximately 400 ms after the statement. Such a response is plausible since the early response will mainly depend on the visual appearance of the stimulus and not upon the content of the text. The discriminative response could be observed relatively early considering the need for language processing. However, due to the repetitive nature of the comparisons, participants likely did not have to parse the language of the statement, but rather look for which trajectory number was mentioned first. The signal differences in the later part of the time window (approximately 600 ms to 1,000 ms) might not solely be explained by the statement stimulus, but could also be influenced by the preparation of motor activity for the button press. However, the earliest allowed occurrence of button presses was 2 s after the statement (mean 2.61 ± 0.29 s across all participants), and results from a different experimental paradigm indicate that the decoding performance in judgment tasks does not rely solely on button press activity and that it can be improved when reducing motor-related signal components [15].

4.2.2 EEG-based Classification of Pairwise Preferences

Inspecting the classification results for individual comparisons (as depicted in the left half of Fig. 5), we achieved a mean accuracy of 0.76 when using the windows time-aligned to the preference statements. In the more difficult asynchronous *observation* setting—where we had no temporal alignment to specific stimuli—we still achieved an accuracy of 0.63. This is especially interesting since such information could potentially be recorded during regular interaction with the robot, with increasing confidence after repeated observations.

To check our expectations about the suitability of windows centered in the trajectory (c.f., Section 3.1.1), we also evaluated two additional window alignments during observation: Extracting time windows relative to the start or end of trajectories rather than the center performs worse, with accuracies of 0.53 and 0.59, respectively.

Combining classifier output from both window types (accuracy of 0.78) yielded improvements over only using the statement response. Note that in our evaluation setting (using identical target trajectories across all participants), "nonconforming" perception by the participant (indicated by a button accuracy of less than 1) is possible. This likely also affected the decodable brain states and therefore the corresponding accuracies for these trajectories.

Answering our first two questions, (1) brain signals both during trajectory observation and after preference statements were informative about user preferences and (2) this translated into an accuracy of 0.78 on a single comparison level using combined observation and statement responses.

4.3 Results for Trajectory Ranking

Evaluating the identification of target trajectories from pairwise preferences, we examined 176 preference learning tasks, each consisting of nine pairwise comparisons. The pairwise preferences were estimated from brain signals or—for comparison with behavioral feedback—from button presses. Our experiment design allowed us to evaluate the ranking performance using geometric distances since we had the ground-truth target trajectories available. For this, we used $d_{\text{target}}(\xi)$ as defined in Equation 1 in order to calculate the difference Δd_{target} between the obtained and the best possible distance to the preference. In addition to this absolute score, we also report the normalized discounted cumulative gain (nDCG), which has a range from 0 to 1 and shows the relative performance based on the best possible obtainable ranking. As the *relevance* for nDCG calculation, we used the negative distance from the preference, shifted by the maximum. We denote the nDCG of the top-ranked item with nDCG@1 and use nDCG@3 for the metric calculated based on the three highest-ranked trajectories. As a baseline method, we compared our ranking approaches to the trajectory preference perceptron proposed by Jain et al. [11], using identical features for all methods (c.f., Section 3.2).

4.3.1 Mean ranking performance

Inspecting the results of instance-based ranking with incorporated confidences (i.e., using $\hat{R}_{\text{instance(conf)}}$) for different feedback types (see Fig. 5 and Table 1), we observed that it is indeed possible to successfully learn trajectory preferences from EEG-based user feedback. For this, the use of the statement or combined EEG features is needed. The latter achieved a mean Δd_{target} of 0.07 for the top-ranked trajectory.

	Δd_{target}				nDCG@1				nDCG@3			
feedback type	obs	stmt	comb	bttn	obs	stmt	comb	bttn	obs	stmt	comb	bttn
ranking												
traj. perceptron [11]	0.15	0.11	0.11	0.08	0.64	0.72	0.74	0.81	0.68	0.72	0.73	0.77
instance	0.13	0.11	0.10	0.08	0.68	0.74	0.75	0.81	0.68	0.73	0.75	0.82
instance (conf)	0.13	**0.07**	**0.07**	0.08	0.68	**0.82**	**0.82**	0.81	0.68	**0.82**	**0.84**	0.82
feature	**0.11**	0.09	0.09	**0.05**	**0.74**	0.78	0.79	**0.87**	**0.73**	0.78	0.78	**0.85**

Table 1 Ranking performances of our methods and the baseline from [11] on the different feedback types: EEG response to observations (obs), statements (stmt), and both combined (comb) as well as button presses (bttn). As performance measures we use the difference to the best obtainable d_{target} (lower is better) and the normalized discounted cumulative gain (nDCG, higher is better) for the top-ranked (nDCG@1) and the three highest-ranked trajectories (nDCG@3).

Considering the three highest-ranked trajectories, the use of statement windows yielded an nDCG@3 of 0.82, compared to 0.84 for the combined features. Limited by the lower performance on the comparison level, the ranking based on the observation setting performed substantially worse (Δd_{target} of 0.13 and nDCG@3 of 0.68).

Interestingly, the ranking based on combined EEG signals achieved a similar performance as the ranking obtained by explicit button presses (nDCG@1 of 0.82 for combined EEG vs. 0.81 for button and nDCG@3 of 0.84 vs. 0.82, respectively). Note that the improvements (which might appear counterintuitive at first) were due to taking into account confidence values from the EEG classifiers, which were not available for the button press (c.f., instance-based results without confidences in Table 1). One possible explanation for this is that for comparisons without a clear preference, this ambivalence results in nondiscriminative brain signals whereas the button press forces an arbitrary choice.

Using our feature-based trajectory ranking approach ($\hat{R}_{\text{feat}}$, where training labels are based on user feedback), we could observe that for low label noise (i.e., button presses), it outperformed all other ranking settings (mean d_{target} of 0.05 and a mean nDCG@1 of 0.87). Since label uncertainty in the form of comparison confidences is, however, not incorporated in this approach, it performed worse than the instance-based ranking when utilizing the less accurate EEG-based feedback. In the adequate comparison with the instance-based approach without confidences, the feature-based ranking performed better. As shown in Table 1, our feature-based approach outperformed the trajectory preference perceptron proposed in [11] in all feedback settings.

4.3.2 Analysis of Results for Individual Participants and Preferences

Performing identical experiments with all participants allowed us to gain insight into the influence of different tasks and users on the ranking performance. The heat maps in Fig. 6 indicate the performance of participants for the different preference tasks. Looking at the button results with low feedback noise in the top two matrices, the similarities within each individual column show that performance variations were systematic and depended on the set of trajectories and targets.

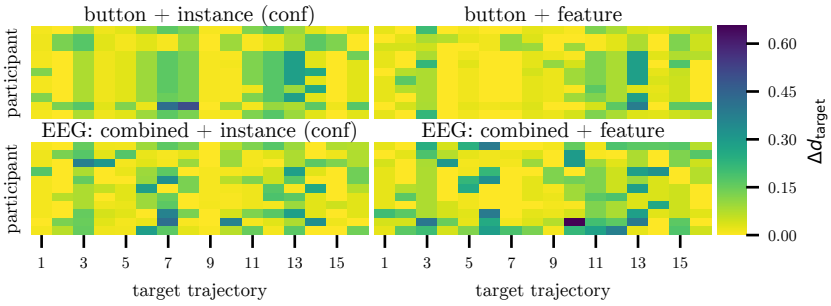

Fig. 6 Ranking performance for individual participants and preferences: The top matrices are based on button press and the bottom matrices are based on the combined EEG responses. The left matrices use instance-based ranking without additional trajectory information and the right ones use the feature-based ranking using trajectory and scene features. Entries in each matrix correspond to a single participant (row) for a single target trajectory (column). Performances are measured using the difference from the target Δd_{target} for the top-ranked trajectory, lower is better)

Specifically, we observed comparatively low performance for target trajectories 3 and 13. Here, also the button feedback accuracy (not shown) is lower for most participants, indicating "harder" comparisons that also translated into worse EEG-based decoding and ranking performance. While behavioral performance is similar across most participants, we observed a higher variance in the EEG-based setting due to the inherent interperson variability in EEG measurements. However, a recovery of preferences was possible in most of the preference tasks for all participants.

Answering question (3), ranking based on EEG-based predictions was possible both in the instance-based and the feature-based setting. Moreover, we found that utilizing EEG-based confidence values in addition to the pairwise preference prediction allowed a ranking performance based on brain signals that is comparable to results obtained by button presses.

5 Conclusion

In this work, we presented a novel approach to learn user preferences for trajectories from brain signals based on pairwise comparisons. Our approach predicts pairwise preferences from EEG data both during passive observation of trajectories and in response to explicit statements. We utilize these predictions to rank observed trajectories.

In extensive experiments, we demonstrated that brain signals as a feedback modality were informative about a user's target trajectory and the latter could be reliably predicted in a pairwise setting. Furthermore, we showed that—despite the low signal-to-noise ratio of EEG signals—ranking trajectories using the EEG-based pairwise preference predictions allowed us to identify the target trajectories with a performance comparable to explicit button presses.

Our results open up paths for future work both on EEG-based active learning of preferences [23, 1]—where feedback could also be repeated on demand to reduce measurement noise—and on utilizing brain signals in a passive way during human–robot interaction to improve robotic behavior without explicitly querying the human.

Acknowledgements The authors would like to thank Robin Burchard for help recording the robot trajectories and Joseline Veit for help conducting the EEG experiments. 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 Federal Ministry of Education and Research under grant OML, the DFG through INST 39/963-1 FUGG as well as the Ministry of Science, Research and the Arts of Baden-Württemberg for bwHPC.

References

1. Akrour, R., Schoenauer, M., Sebag, M., Souplet, J.C.: Programming by Feedback. In: E.P. Xing, T. Jebara (eds.) Proceedings of the 31st International Conference on Machine Learning, *Proceedings of Machine Learning Research*, vol. 32, pp. 1503–1511. PMLR, Beijing, China (2014)
2. Bajcsy, A., Losey, D.P., O'Malley, M.K., Dragan, A.D.: Learning Robot Objectives from Physical Human Interaction. In: S. Levine, V. Vanhoucke, K. Goldberg (eds.) Proceedings of the 1st Annual Conference on Robot Learning, *Proceedings of Machine Learning Research*, vol. 78, pp. 217–226. PMLR (2017)
3. Barachant, A., Bonnet, S., Congedo, M., Jutten, C.: Classification of covariance matrices using a Riemannian-based kernel for BCI applications. Neurocomputing **112**, 172–178 (2013). DOI 10.1016/j.neucom.2012.12.039
4. Barachant, A., Congedo, M.: A Plug&Play P300 BCI Using Information Geometry. arXiv:1409.0107 [cs, stat] (2014)
5. Burget, F., Fiederer, L.D.J., Kuhner, D., Völker, M., Aldinger, J., Schirrmeister, R.T., Do, C., Boedecker, J., Nebel, B., Ball, T., Burgard, W.: Acting thoughts: Towards a mobile robotic service assistant for users with limited communication skills. In: 2017 European Conference on Mobile Robots (ECMR), pp. 1–6 (2017). DOI 10.1109/ECMR.2017.8098658
6. Busch, B., Grizou, J., Lopes, M., Stulp, F.: Learning Legible Motion from Human–Robot Interactions. Int J of Soc Robotics pp. 1–15 (2017). DOI 10.1007/s12369-017-0400-4
7. Busch, B., Maeda, G., Mollard, Y., Demangeat, M., Lopes, M.: Postural optimization for an ergonomic human-robot interaction. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2778–2785 (2017). DOI 10.1109/IROS.2017.8206107
8. Chavarriaga, R., Sobolewski, A., Millán, J.d.R.: Errare machinale est: The use of error-related potentials in brain-machine interfaces. Front. Neurosci **8**, 208 (2014). DOI 10.3389/fnins.2014.00208
9. Chen, X., Wang, Y., Nakanishi, M., Gao, X., Jung, T.P., Gao, S.: High-speed spelling with a noninvasive brain–computer interface. PNAS **112**(44), E6058–E6067 (2015). DOI 10.1073/pnas.1508080112
10. Dragan, A., Srinivasa, S.: Integrating human observer inferences into robot motion planning. Auton Robot **37**(4), 351–368 (2014). DOI 10.1007/s10514-014-9408-x
11. Jain, A., Sharma, S., Joachims, T., Saxena, A.: Learning preferences for manipulation tasks from online coactive feedback. The International Journal of Robotics Research **34**(10), 1296–1313 (2015). DOI 10.1177/0278364915581193
12. Jeon, H.J., Dragan, A.D.: Configuration Space Metrics. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5101–5108 (2018). DOI 10.1109/IROS.2018.8593564
13. Joachims, T.: Optimizing Search Engines Using Clickthrough Data. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02, pp. 133–142. ACM, New York, NY, USA (2002). DOI 10.1145/775047.775067

14. Kim, S.K., Kirchner, E.A., Stefes, A., Kirchner, F.: Intrinsic interactive reinforcement learning – Using error-related potentials for real world human-robot interaction. Scientific Reports **7**(1), 17562 (2017). DOI 10.1038/s41598-017-17682-7
15. Kolkhorst, H., Kärkkäinen, S., Raheim, A.F., Burgard, W., Tangermann, M.: Influence of User Tasks on EEG-based Classification Performance in a Hazard Detection Paradigm. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6758–6761 (2019). DOI 10.1109/EMBC.2019.8857812
16. 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), pp. 7111–7116 (2018). DOI 10.1109/IROS.2018.8593992
17. Kretzschmar, H., Spies, M., Sprunk, C., Burgard, W.: Socially compliant mobile robot navigation via inverse reinforcement learning. The International Journal of Robotics Research p. 0278364915619772 (2016). DOI 10.1177/0278364915619772
18. Kuffner, J.J., LaValle, S.M.: RRT-connect: An efficient approach to single-query path planning. In: Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), vol. 2, pp. 995–1001 vol.2 (2000). DOI 10.1109/ROBOT.2000.844730
19. Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., Yger, F.: A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. J. Neural Eng. (2018). DOI 10.1088/1741-2552/aab2f2
20. Mainprice, J., Sisbot, E., Jaillet, L., Cortes, J., Alami, R., Simeon, T.: Planning human-aware motions using a sampling-based costmap planner. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 5012–5017 (2011). DOI 10.1109/ICRA.2011.5980048
21. Meinel, A., Castaño-Candamil, S., Reis, J., Tangermann, M.: Pre-Trial EEG-Based Single-Trial Motor Performance Prediction to Enhance Neuroergonomics for a Hand Force Task. Front. Hum. Neurosci p. 170 (2016). DOI 10.3389/fnhum.2016.00170
22. Rivet, B., Souloumiac, A., Attina, V., Gibert, G.: xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface. IEEE Transactions on Biomedical Engineering **56**(8), 2035–2043 (2009). DOI 10.1109/TBME.2009.2012869
23. Sadigh, D., Dragan, A., Sastry, S., Seshia, S.: Active Preference-Based Learning of Reward Functions. In: Proceedings of Robotics: Science and Systems. Cambridge, Massachusetts (2017). DOI 10.15607/RSS.2017.XIII.053
24. Salazar-Gomez, A.F., DelPreto, J., Gil, S., Guenther, F.H., Rus, D.: Correcting robot mistakes in real time using EEG signals. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 6570–6577 (2017). DOI 10.1109/ICRA.2017.7989777
25. Schreuder, M., Rost, T., Tangermann, M.: Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Front. Neurosci. **5**, 112 (2011). DOI 10.3389/fnins.2011.00112
26. Schultze-Kraft, M., Dähne, S., Gugler, M., Curio, G., Blankertz, B.: Unsupervised classification of operator workload from brain signals. J. Neural Eng. **13**(3), 036008 (2016). DOI 10.1088/1741-2560/13/3/036008
27. Shah, N.B., Wainwright, M.J.: Simple, Robust and Optimal Ranking from Pairwise Comparisons. Journal of Machine Learning Research **18**(199), 1–38 (2018)
28. Yger, F., Berar, M., Lotte, F.: Riemannian Approaches in Brain-Computer Interfaces: A Review. IEEE Transactions on Neural Systems and Rehabilitation Engineering **25**(10), 1753–1762 (2017). DOI 10.1109/TNSRE.2016.2627016