OPEN ACCESS
IOP Publishing
Journal of Neural Engineering
J. Neural Eng. **15** (2018) 026002 (10pp)
https://doi.org/10.1088/1741-2552/aa9999

Implicit relevance feedback from electroencephalography and eye tracking in image search

Jan-Eike Golenia[1,2], Markus A Wenzel[1,2], Mihail Bogojeski[1] and Benjamin Blankertz[1]

[1] Fachgebiet Neurotechnologie, Technische Universität Berlin, Marchstr. 23, 10587 Berlin, Germany
[2] Equal contributions.

E-mail: jagoleni@uos.de, markus.wenzel@hhi.fraunhofer.de, mihail.bogojeski@campus.tu-berlin.de and benjamin.blankertz@tu-berlin.de

Received 14 April 2016, revised 25 October 2017
Accepted for publication 10 November 2017
Published 24 January 2018

Abstract

Objective. Methods from brain–computer interfacing (BCI) open a direct access to the mental processes of computer users, which offers particular benefits in comparison to standard methods for inferring user-related information. The signals can be recorded unobtrusively in the background, which circumvents the time-consuming and distracting need for the users to give explicit feedback to questions concerning the individual interest. The obtained implicit information makes it possible to create dynamic user interest profiles in real-time, that can be taken into account by novel types of adaptive, personalised software. In the present study, the potential of implicit relevance feedback from electroencephalography (EEG) and eye tracking was explored with a demonstrator application that simulated an image search engine. *Approach.* The participants of the study queried for ambiguous search terms, having in mind one of the two possible interpretations of the respective term. Subsequently, they viewed different images arranged in a grid that were related to the query. The ambiguity of the underspecified search term was resolved with implicit information present in the recorded signals. For this purpose, feature vectors were extracted from the signals and used by multivariate classifiers that estimated the intended interpretation of the ambiguous query. *Main result.* The intended interpretation was inferred correctly from a combination of EEG and eye tracking signals in 86% of the cases on average. Information provided by the two measurement modalities turned out to be complementary. *Significance.* It was demonstrated that BCI methods can extract implicit user-related information in a setting of human-computer interaction. Novelties of the study are the implicit online feedback from EEG and eye tracking, the approximation to a realistic use case in a simulation, and the presentation of a large set of photographies that had to be interpreted with respect to the content.

Keywords: eye fixation related potentials, implicit relevance feedback, eye tracking, brain-computer interfacing, electroencephalography

(Some figures may appear in colour only in the online journal)

1. Introduction

Signals from the brain may contain implicit information about the users of computers, which can potentially be decoded with methods from brain–computer interfacing (BCI) [1–4]. Such a direct access to the mental processes of the users offers particular benefits in comparison to standard methods for the inference of user-related information, e.g. asking the user for explicit feedback, or observing the user's interaction with the device. Physiological signals can be recorded unobtrusively in the background, and their analysis would circumvent the time-consuming and distracting need for the user to give explicit feedback to questions concerning the individual interest, as well as a possible response bias. The obtained implicit information could augment standard input devices (e.g. computer mouse and keyboard) for the interaction between human and machine.

Research on BCI has shown that humans can volitionally generate 'neural signatures' that can be detected in the electroencephalogram (EEG) with pattern recognition methods in real-time. The extracted information can be translated into a signal serving for control or communication [5–9]. Some BCI methods exploit the phenomenon that stimuli of interest, which are flashed in a stimulus sequence, elicit a detectable attention-related neural response [10–13]. Combining this BCI technique with eye tracking makes it possible to infer the subjective relevance of the single elements of the visual surrounding [14–21].

The present study demonstrates that it is possible to decode from EEG and eye tracking signals which images were subjectively relevant for the user of a simulated web image search engine (see 'Flickr' or 'Google Images'). The resulting relevance map of the computer screen, where numerous images were displayed at the same time in a grid, made it possible to characterise the current interest of the individual user. Implicit relevance information can be aggregated in dynamic user interest profiles, that could be taken into account by novel types of adaptive, personalised software. This potential is explored here with a demonstrator application that infers the user interest online from implicit information hidden in the signals. Novelties of the study are the implicit online feedback from a combination of EEG and eye tracking signals, the approximation to a realistic use case in a simulation, and the presentation of a large set of photographies that had to be interpreted with respect to the content (which goes beyond the mere recognition of previously known simple stimuli that are typical for BCI paradigms based on event-related potentials). The demonstrator is not considered to be a final application of its own right, but may be an important step towards future applications that are informed by the insights gained.

The presented novel approach may show promise in light of the increasing interest of customers and large technology companies in wearable physiological sensors [22] and recently developed, deployable eye tracking and EEG systems, which will make the signal acquisition during daily life more and more feasible—in contrast to the bulky, expensive, inconvenient, and stationary equipment of the past. Examples of the technological innovations are affordable eye trackers [23] and mobile EEG systems [24–26] with gel-free [27–30], miniaturised [31] electrodes that can be placed hardly visible in/on/around the ear [32–36]. Moreover, in-ear headphones with different physiological sensors including EEG, which connect with a smartphone, are under development (e.g. 'The Aware' from 'United Sciences', Atlanta, USA).

2. Methods

2.1. Experimental design

The participants of the study queried for ambiguous terms in a simulated image search engine, and viewed different images that were related to the respective search term. During image viewing, the EEG was recorded and the eye movements were tracked. Feature vectors were extracted from the signals in order to train a classifier that estimated the intended interpretation of the ambiguous search term. First, the participants were asked to choose one of two possible interpretations (like 'animal-nature-wildlife' versus 'baseball-ball-sports') of an ambiguous search term (here 'bat'). Then, they viewed 24 square images arranged in a four-times-six grid on the screen that were related to either one or the other meaning of the query (see figure 1; non-square images were cropped). Finally, they were asked to report the number of the pictures belonging the chosen category and got feedback on whether their response was correct. This procedure was repeated 154 times with different ambiguous search terms. Further examples of the queries are 'jam' with the possible interpretations 'cream-tea-scone' versus 'music-guitar-band', 'deck' ('ship-sea-boat' versus 'skateboard-skate-board'), and 'tick' ('macro-insect-bug' versus 'time-clock-tock'). The participants were instructed to quickly skim the images instead of prioritizing the correct accomplishment of the counting task, assuming that this behaviour is typical when browsing image search results. Before the appearance of the image mosaic, a fixation cross directed the gaze to the upper left corner of the screen. Each picture shown in the image mosaic was picked randomly from one of the two given categories with a probability of $p = 11/24$. In addition, few 'odd' pictures, which were not related to the query, were displayed with a probability of $p = 2/24$. The odd pictures were randomly selected from the remainder of the image collection.

2.2. Experimental stimuli

All pictures were obtained from Flickr [37], a service for sharing pictures aimed at amateur and professional photographers. Flickr provides access to a large collection of user annotated pictures via an application programming interface (*'API'*; [38]). Flickr clusters the images into categories that contain images with similar content according to the user annotations (tags). These clusters can be accessed via the API with the 'cluster search' function. Called with a single search term, the function returns up to four clusters. Each cluster is described by a list of tags and named after the first three tags. Several lists of homonyms (e.g. [39]) served as query terms for the cluster search function, and a collection of 63 110

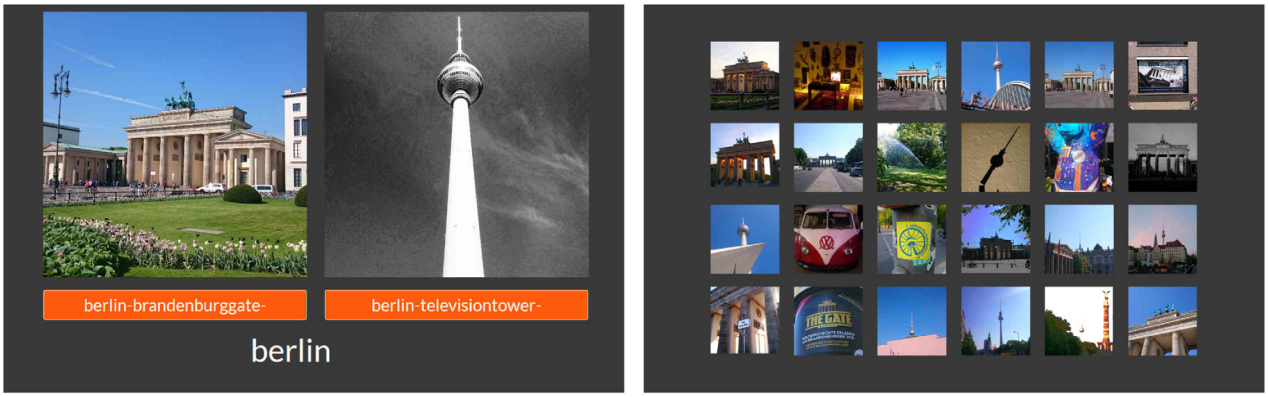

Figure 1. Exemplary stimulus presentation. *Left:* selecting one category of the underspecified search term 'Berlin'. *Right:* the result page contains pictures from both categories (either 'Berlin-Brandenburg Gate' or 'Berlin-Television Tower') and few 'odd' pictures (room, park, car) that are not related to the search term. The original photographies were replaced by similar own pictures in this illustration due to copyright restrictions.

images related to 936 ambiguous terms was downloaded. Search terms were picked that generated two clusters with more than 18 pictures each that could be clearly associated with the name of the respective cluster. A manual review was necessary, because many pictures were hardly in any relation to the cluster name or query term.

Ambiguity was rarely the result of lexical homonymy, but more often due to underspecified search queries. The search term 'filter' resulted, for instance, in images of coffee filters, in pictures of filter lenses made of glass and in photographies processed by different digital filters. The two categories were illustrated for the participant by the first three tags and one example picture per cluster (see section 2.1 and figure 1). Some categories could be easily distinguished, others not. For instance, the categories 'hyacinth-flower-blue' and 'fruit-green-macro' of the search term 'grape' could be easily discerned. The former consisted of close-up photographies of blue hyacinth flowers in a grape shaped form, the latter contained grapes and other fruits that were never blue. In contrast, it was difficult to distinguish the categories 'paint-art-painting' and 'makeup-eyeshadow-cosmetics' of the search term 'palette', because the images of both categories depicted colour palettes, that contained either make-up or paint for drawing.

2.3. Data acquisition

Fourteen persons with normal vision and no report of eye or neurological diseases participated in the experiments. The age of the five female and nine male subjects ranged from 22 to 33 yr with a mean age of 27.7 yr (standard deviation: 2.96). The first subject viewed 123 result pages and all others 154 result pages. One recording session included giving an informed written consent to take part in the study, vision tests for eye dominance, preparation of the sensors, eye tracker calibration and validation, introduction to the task and the main experiment (with a duration of about 1.5 h). The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (application number BL_03_20150109).

The participant sat at a distance of 60 cm in front of a computer screen and entered the number of the counted target pictures with a keyboard. Physiological signals were recorded with two amplifiers with 62 active EEG electrodes (BrainAmp, ActiCap, BrainProducts, Munich, Germany; sampling frequency of 1000 Hz) and one active electrode for electrooculography (EOG). An eye tracker (RED 250, SensoMotoric Instruments, Teltow, Germany; sampling frequency of 250 Hz) was attached to the screen. A chin rest gave orientation for a stable position of the head. The screen had a resolution of 1680 pixels × 1050 pixels, a size of 47.2 cm × 29.6 cm and subtended a visual angle of 38.2° in horizontal and 26.3° in vertical direction.

EEG was acquired and analysed with Wyrm and Mushu [40, 41]. The synchronously recorded EEG and eye tracking signals were aligned with the help of sync-triggers. Client-side JavaScript Ajax (asynchronous JavaScript and XML) calls sent HTTP requests every 500 ms that in turn called a function on the backend (Flask web server) that elicited the subsequent recording of EEG and eye tracking time-stamps. These time-stamps were used to estimate the parameters of a linear regression function for the mapping of eye-tracker-time to EEG-time. The EEG data were low-pass filtered with a second order Chebyshev filter (42 Hz passband, 49 Hz stop band), down-sampled to 100 Hz, re-referenced to the digitally linked-mastoids and high-pass filtered with a Butterworth filter at 0.5 Hz. The last 500 ms of each stimulus presentation were not considered for the analysis in order to avoid confounds from the terminating button press. The first three result pages were only used for practice and not for analysis.

The proper calibration of the eye tracker was re-validated at least four times during the experiment and more often if the subject was unsteady and moved a lot. A picture was considered as fixated if the location detected by the online algorithm of the eye tracker was situated within the borders of the picture plus 20 pixels (0.52°). The pictures had a side length of 186 pixels and subtended a visual angle of 5.0°. The size was picked to fit approximately into the area with high foveal resolution. The distance between the pictures was 35 pixels (0.9°) in horizontal direction, 40 pixels (1.1°) in vertical direction,

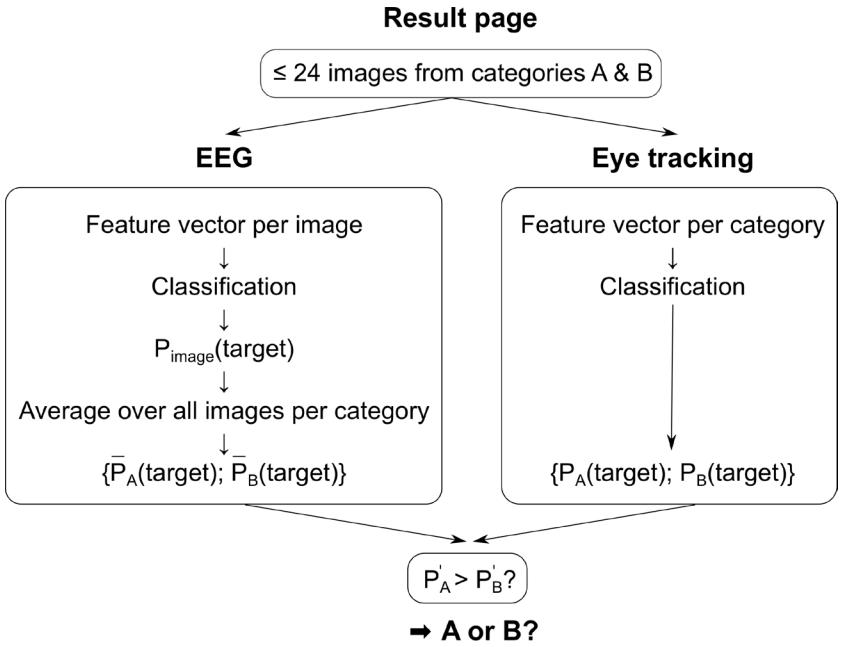

Figure 2. Prediction of the category of interest (flow chart of the data analysis described in section 2.4.1).

181 pixels (4.8°) to the horizontal screen borders and 100 pixels (2.7°) to the vertical screen borders.

The stimuli were presented with web technologies in order to explore the compatibility of the BCI-based relevance detector with common software applications, which are not optimised for the presentation of experimental stimuli (frontend: HTML5, CSS, JavaScript, jQuery, Ajax, Bootstrap, backend: Flask). The experiment was interactive and not a static prearranged sequence of stimuli. The user could navigate between different menu pages (e.g. a page for trial selection) and could calibrate and validate the eye tracker inside the browser under the supervision of the experimenter. For demonstration purposes, it was additionally possible to train a classification model with the data recorded so far using a preliminary version of the classification procedure presented in section 2.4.1. This option was given to the participants after the end of the main recording session. Then, a 'feedback mode' could be launched, that allowed for an online prediction of the respective category of interest (not described further in this paper).

2.4. Data analysis

2.4.1. Prediction of the category of interest. Every result page contained pictures of the two possible interpretations of the ambiguous search term, which will be referred to as *categories*. In addition, few odd pictures were mixed in, which did not belong to any of the two categories. The subjects selected one category of interest before the display of each result page and labelled it as *target category* by pressing a button. The respective other category was labelled as *non-target category*. The selected target category of every result page was inferred from feature vectors extracted from the EEG and eye tracking signals in two steps (see figure 2). First, EEG- and eye-tracking-based feature vectors were classified separately (details are set out below). Then, information from EEG and eye tracking was combined by averaging the classifier estimates of the two measurement modalities. The category with the larger average target estimate was considered to be the target category of the respective result page. Binary classifications were performed, because the odd images were not considered. Linear discriminant analysis (LDA) with shrinkage served as classifier, which regularises the estimated covariance matrix and, thereby, reduces the likelihood of overfitting in the case of high-dimensional data and a limited number of samples [42, 43]. The optimal shrinkage parameter was calculated analytically using the closed form equation derived in [44], which is computationally less expensive than choosing the optimal parameter by cross-validation. Posterior probabilities were computed from the classifier scores because probabilities are well suited for combining different classifier estimates due to the clear upper and lower limit and the same scale. The predictive performance was assessed in ten-fold cross-validations using the classification accuracy as metric.

For the EEG-based prediction, feature vectors corresponding to each fixated image were classified as being either members of the target or the non-target category. The target probabilities of all feature vectors per category were averaged. The category with the larger average target probability can be assumed to be the selected category of interest of the respective result page. Feature vectors were extracted from the continuous multi-channel EEG signals as follows. One second long epochs aligned to the onsets of the longest eye fixations of each image were cut out (fixation-related potentials; 'FRPs') and downsampled to 20 Hz (which reduced the dimensionality of the feature vectors and thereby the risk of overfitting to the training data). The data of all 62 channels were concatenated in one feature vector with 1240 dimensions. The number of samples (longest fixations on either target or non-target images) ranged from 2821 to 5165 per

single subject, with slightly unbalanced classes, because target images were fixated more often than non-target images. Note, that only fixated images could contribute to the inference. For a performance comparison, the first and the last fixation were also tested as time markers of reference—in addition to the default usage of the longest fixation. Methods for artefact rejection were not applied in order to let the classifier learn to deal with potential artefacts in the signals. From experience, this approach is superior to artefact rejection/correction in laboratory experiments with artefacts that are not too severe. A robust classifier can deal with artefacts during online operation, while artefact rejection would lead to missing data, which is critical in many online applications.

For the eye-tracking-based prediction, feature vectors were extracted separately per category and result page, and were classified with shrinkage LDA. These *screen*-based eye tracking features comprised the mean dwell time, the median and maximum fixation duration and the average fixation number. The category with the larger target probability was considered to be the selected category of interest of the respective result page. In addition, an alternative classification strategy was examined, which resembled the procedure of the EEG-based prediction: each image was first classified as member of the target or non-target category based on the dwell time on each image (*single*-image eye tracking features). Then, the single probabilities were averaged per category (*aggregated* eye tracking probabilities). Shrinkage was not necessary in this case because covariances can not be considered for this univariate feature.

In addition, feature vectors extracted from the EOG were classified in order to assess a possible contribution of eye movements to the EEG-based prediction (horizontal eye movements were captured by subtracting channel F10 from channel F9 and vertical eye movements by subtracting channel Fp1 from the signal of the electrode below the eye).

2.4.2. Characteristics of the EEG and eye tracking features. The characteristics of the EEG epochs, which served as features for the classifications, were assessed separately for the three groups of the corresponding images (targets, non-targets, odds). Discriminative information between target versus non-target EEG epochs, between target versus odd EEG epochs, and between non-target versus odd EEG epochs was inspected for each time point and each EEG channel with the point biserial correlation coefficient, which was squared while retaining the sign (r^2). The eye movements were characterised with fixation maps of the result pages, and by computing the statistics of the dwell time, of the number of fixations and of the median and maximum fixation duration of target, non-target and odd images.

2.4.3. Task performance. The behavioural performance and compliance of each participant with the task instructions was assessed by computing the percentage of correct answers, the deviation of the number entered by the subject from the true number of images belonging to the selected category, and the trial durations.

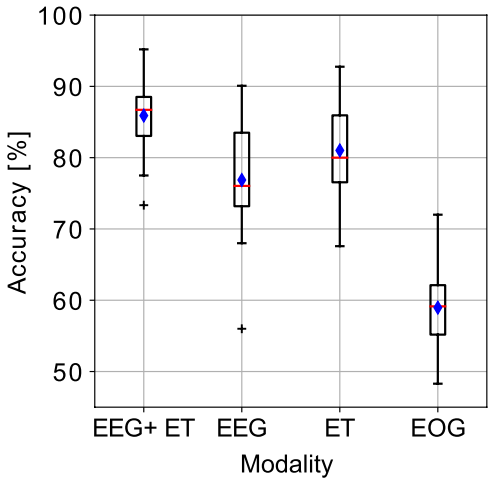

Figure 3. Identifying the selected target category of the ambiguous search term with information extracted from the different signals ('ET' stands for eye tracking). The classification accuracy served as metric for the predictive performance. The chance level of a random classifier would be situated at 50%. Every boxplot represents the average cross-validation results of the participants of the study. Red lines indicate the median values, blue diamonds the mean, black boxes the 25th and 75th percentiles, whiskers the range, and crosses the outliers. EEG and EOG epochs used for the classifications were aligned to the longest fixation.

3. Results

3.1. Prediction of the category of interest

The chosen category of interest of the ambiguous search term could be inferred with an accuracy of 85.9% ± 5.8%, when information from EEG and eye tracking was combined (mean ± standard deviation; the results of the single subjects ranged from 73% to 95%; see figure 3). This outcome is significantly better than the chance level of 50% that can be expected from random guessing ($p < 0.05$, Wilcoxon signed rank test on the population level). When only EEG features were used, the estimates were correct in 76.9% ± 8.7% ($p < 0.05$; ranging from 56.0% to 90.1%), and in 81.0% ± 6.7% for predictions with screen-based eye tracking features only ($p < 0.05$; ranging from 67.6% to 92.8%). The complementarity of information provided by the single modalities was evaluated separately for EEG and eye tracking. A subset of the samples was selected where the prediction based on the respective alternative modality was wrong (i.e. the full set of samples was reduced by about 81.0% and 76.9% respectively). The predictive performance on the subset decreased merely for about five percentage points in comparison to the full set, and was still significantly better than random (EEG if eye tracking wrong: 71.5%, eye tracking if EEG wrong: 76.8%), which indicates complementarity (Wilcoxon signed rank tests, $p <= 0.05$). EOG features resulted in a predictive performance closer to the chance level of 50% in comparison to the other modalities (see figure 3).

The predictive performance based on EEG features only is shown in figure 4. The category of interest was estimated by aggregating the category membership probability estimates of the single images (see black and grey boxplots in figure 4).

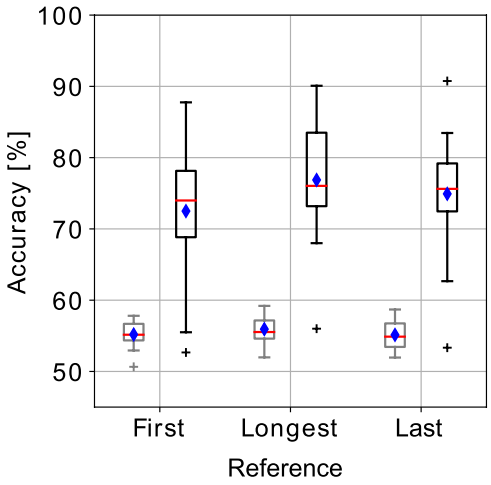

Figure 4. Predictive performance with EEG features only. The category of interest was predicted (*black*) by aggregating the category membership estimates of the single images (*grey*; class-wise weighted accuracies). Either the first, the longest, or the last fixation on an image served as time marker of reference for the feature extraction from the continuous EEG.

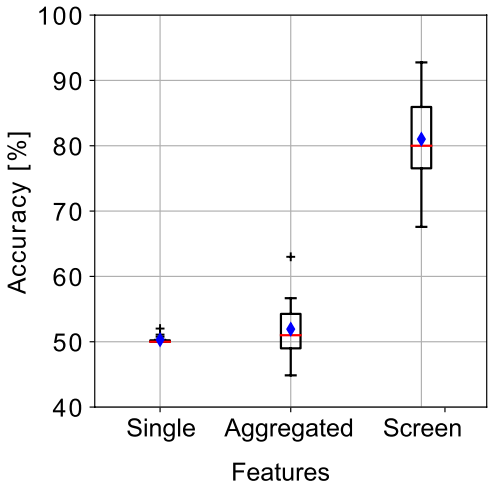

Figure 5. Predictive performance using either the *single*-image eye tracking features, the *aggregated* eye tracking probabilities, or the *screen*-based eye tracking features. The chance level of a random classifier would be situated at 50%.

The class-wise normalised accuracy, which is insensitive to class imbalances, served as performance metric in the case of the single image classification, because target images were fixated more often than non-target images. Using the longest fixation as time marker of reference (for the feature extraction from the continuously recorded EEG) resulted in a slightly better accuracy in comparison to the usage of the first or the last fixation on an image.

The category of interest could be predicted better than random with screen-based eye tracking features, but not with single-image eye tracking features (also when the resulting probabilities were aggregated per category; see figure 5).

3.2. Characteristics of the EEG and eye tracking features

Characteristic neural responses were elicited when either target, non-target or odd images were fixated. An EEG component occurred at about 500 ms to 700 ms after the onset of the longest fixation, and allowed for discriminating targets from non-targets and odds (see figures 6 and 7). Differences between the corresponding EEG epochs were most prominent at central and parietal electrodes. For conciseness, we only display the results of the longest fixation, because the spatial distributions and time courses of the different fixations were very similar (with a small time lag).

Result pages were scanned in a systematic order, starting in the upper left corner (at the position of the fixation cross) and then continuing row by row to the bottom right (see figure 8 for a typical fixation map). Few subjects examined column by column, almost all subjects applied the same search strategy to most of the search screens.

Different fixation patterns were observed for target, non-target and odd pictures (see figure 9). Dwell time, number of fixations and median and maximum fixation duration were significantly larger for targets than for non-targets ($p < 0.05$, Wilcoxon signed rank test across all subjects; medians: 523 ms versus 339 ms, 2.15 versus 1.6, 218 ms versus 196 ms, 551 ms versus 428 ms) and distributed more broadly as indicated by the standard deviations (dwell time: 248 ms versus 210 ms, number of fixations: 0.82 versus 0.73, fixation duration: 44 ms versus 40 ms, maximum fixation duration: 221 ms versus 207 ms). The odd distributions have non-empty bins at zero, because sometimes all odd images of a result page were skipped.

3.3. Task performance

Correct answers were given in $45.7 \pm 14.2\%$ of the cases (mean $\pm$ standard deviation), ranging from 20% to 63%. Participants tended to miss a target rather than counting too many (see figure 10, bottom). The participants spent a median time of about 15 s and rarely more than 20 s on each result page with 24 images. Accordingly, single images were typically viewed less than one second.

4. Discussion

4.1. Prediction of the category of interest

Ambiguity in image search was resolved by inferring the intended meaning of the underspecified query term from information present in EEG and/or eye tracking signals. Predicting the category of interest was possible with both measurement modalities. Combining the modalities improved the predictive performance, which suggests that EEG and eye tracking provide complementary information (see section 3.1 and figure 3). The following findings give further evidence for this claim: testing only samples that were misclassified by the respective other modality resulted in an accuracy that was still significantly better than random (see section 3.1). Thus, the classifiers made different mistakes and exploited different information. Moreover, discriminative information present in the fixation-related EEG

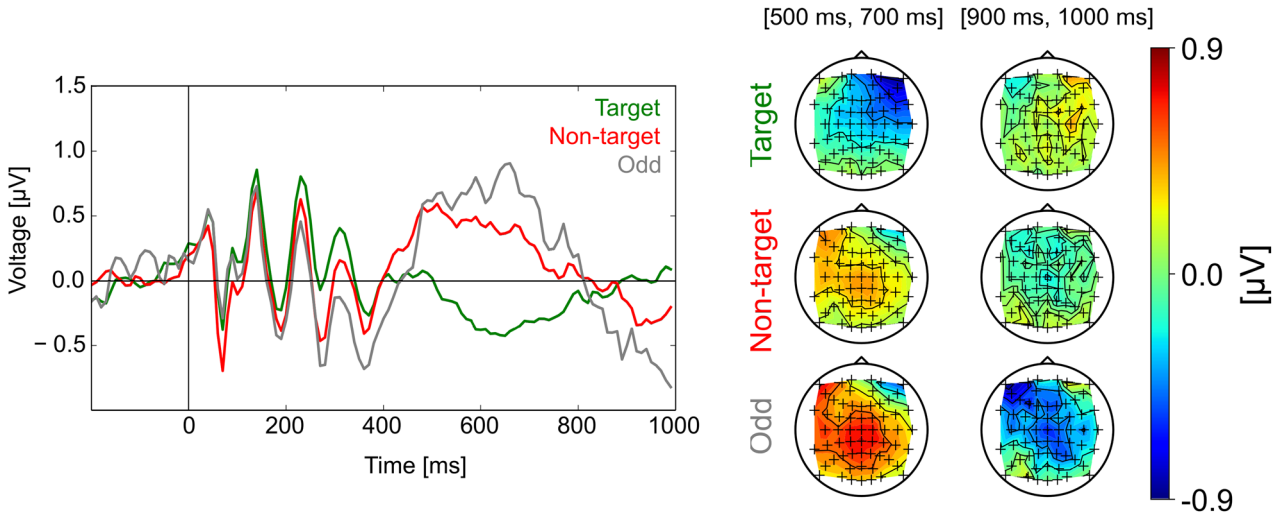

Figure 6. Average EEG responses to the longest fixation of target, non-target and odd pictures (*left:* time courses at electrode Cz; *right:* scalp maps with all electrodes in two selected temporal intervals; averages over all EEG epochs of all participants).

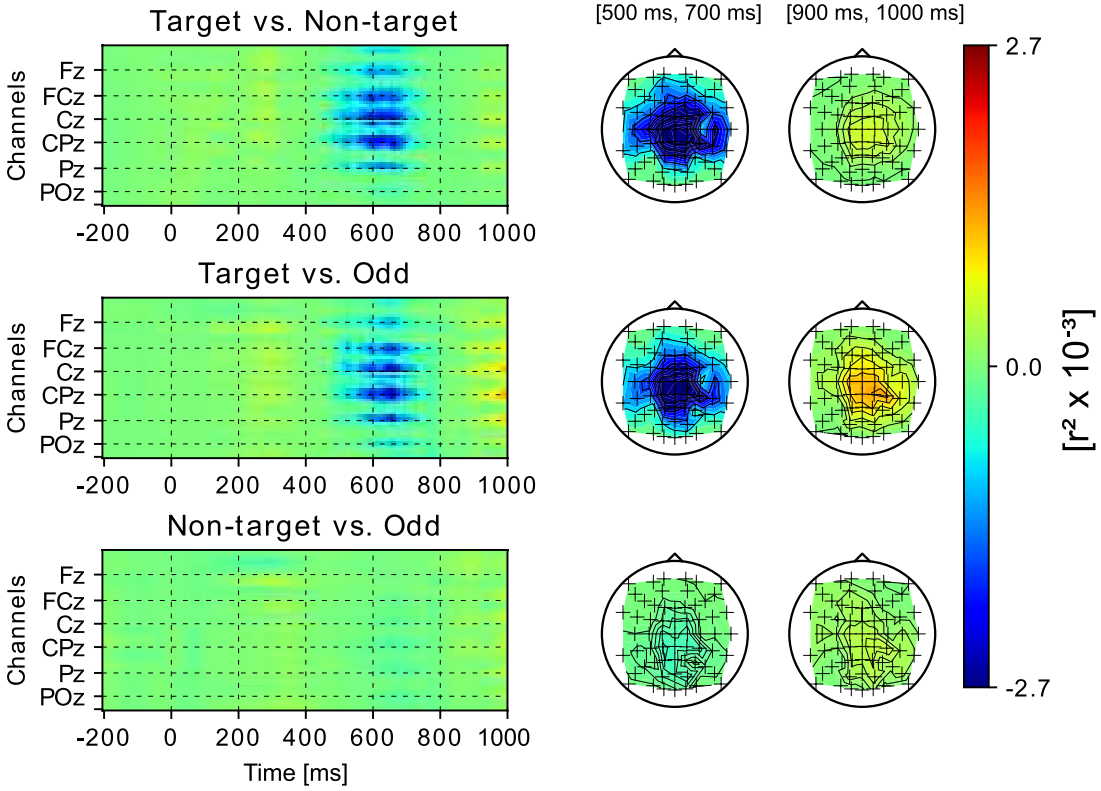

Figure 7. Statistical differences (signed r^2 values) between target versus non-target EEG epochs (top), between target versus odd EEG epochs (centre), and between non-target versus odd EEG epochs (bottom). The epochs were aligned to the longest fixations of the images. The channels are ordered from the front to the back and from the left to the right side of the head. Averages over all subjects of the study are shown for all time points (*left*) and for two selected intervals as scalp maps (*right*). A significance threshold was not applied in order to keep also subtle differences that can potentially be exploited by the multivariate classifier (see section 2.4.1).

epochs was found mainly at central electrodes, which are presumably less confounded by eye movements than electrodes at outer positions (see figure 7; eye movements may have influenced the EEG responses of the single classes; see the topographies in figure 6). Besides, differences in the EEG started at about 500 ms after fixation onset (see figure 7), and, therefore, mainly after the onset of the following eye movement (see figure 9).

Accumulating evidence (classifier probabilities) over several feature vectors considerably improved the EEG-based predictive performance (see section 3.1 and 4). Thus, the findings demonstrate that the inherent uncertainty of the single relevance estimates (here: for single images) can be overcome by including information about the membership to a more general category (here: possible interpretations of an ambiguous term). This insight can be taken into account also by

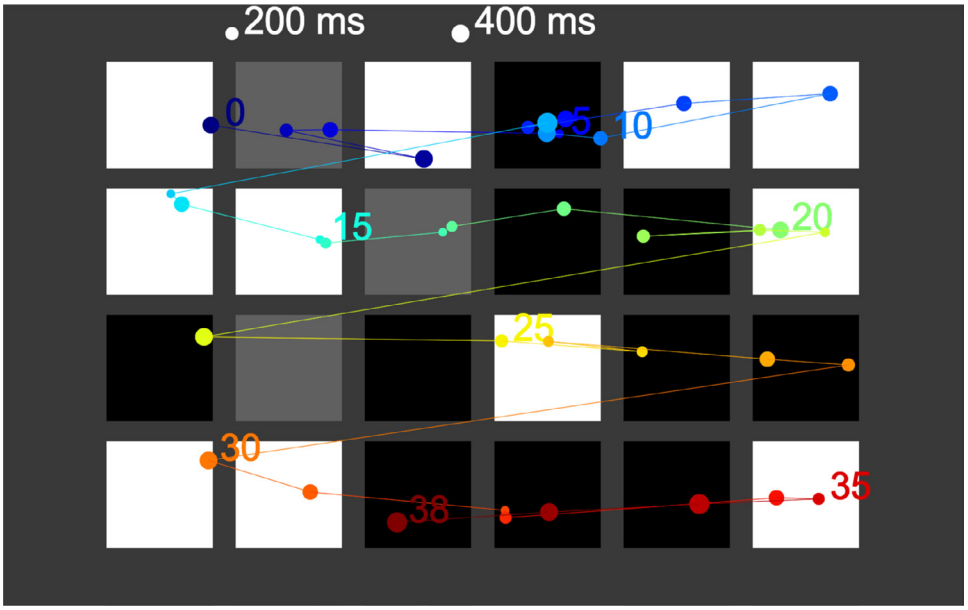

Figure 8. Exemplary fixation map of a participant inspecting a result page. The participant searched pictures of one category of the ambiguous search term that are represented here by white tiles (due to copyright restrictions) and was less interested in pictures of the second category (black tiles). Three 'odd' pictures (grey tiles) were not related to the search term. Eye fixations are indicated by blobs with surfaces proportional to the respective fixation duration. The first, the last and every fifth fixation are labelled. Colours indicate the order of the fixations (from blue to red).

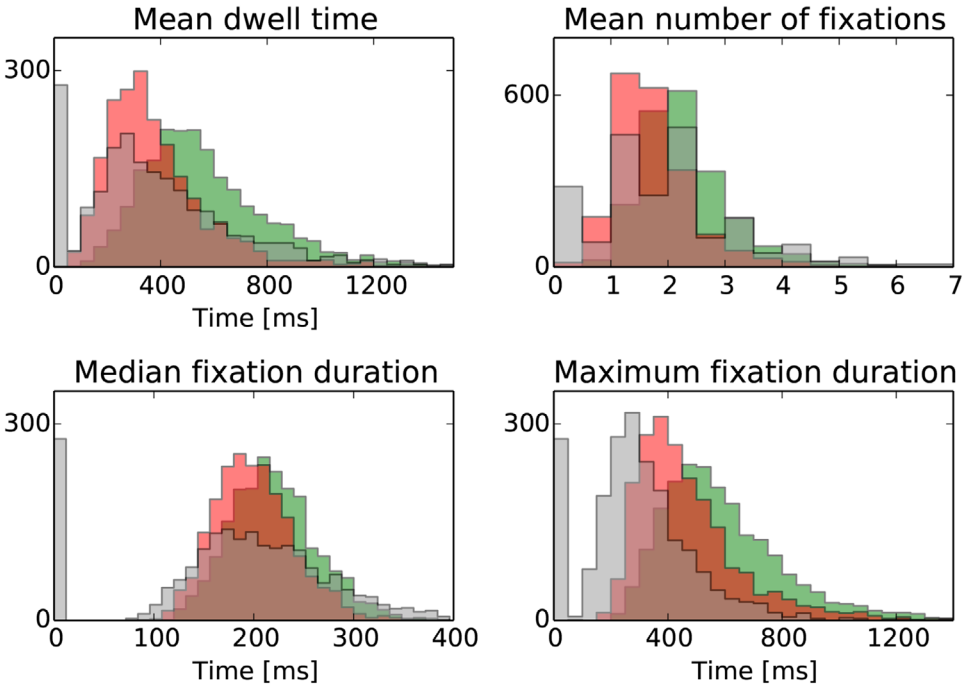

Figure 9. Distributions of the four eye tracking features, averaged over all subjects, for the three categories 'targets' (green), 'non-targets' (red) and 'odds' (grey).

future efforts that apply brain-computer interfacing to human-computer interaction.

The effect of the number of *test* samples used for evidence accumulation on the certitude of the final prediction is inspected in more detail in [21]. In addition, the expectable generalization performance of a predictive model typically grows with more *training* samples available, but has to be weighed up against the effort and the duration to acquire more training samples. This trade-off depends on the specifics of the future application. We therefore decided not to investigate this dependency in more detail for the current study, which investigates merely a demonstrator.

The longest fixations may have resulted in the best EEG-based predictive performance (see section 3.1 and figure 4) because they presumably served for a closer inspection of informative spots of the picture (and were not only intermediate stops on negligible spots).

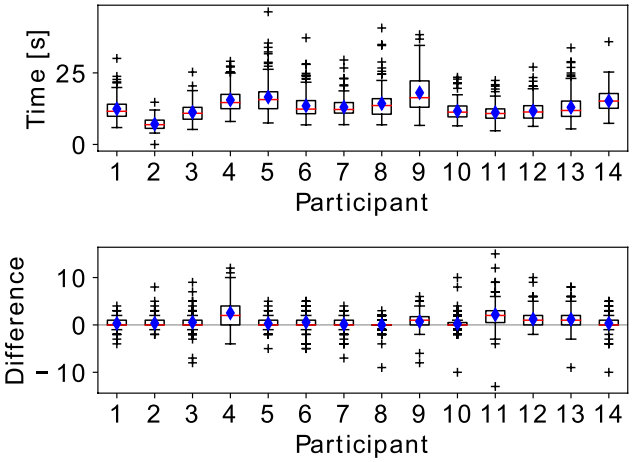

Figure 10. Top: viewing durations of the result pages for every participant of the study. Bottom: mean absolute differences between the true target number and the targets counted by the subject.

4.2. Characteristics of the EEG and eye tracking features

The fixation of non-target and odd images evoked a late positive complex, in contrast to target images (see section 3.2 and figures 6 and 7). The effect occurred later than it can be expected from the EEG component 'P300', which is evoked by the oddball paradigm [45]. The stimuli were photographies that differed not only in low-level features, which could be quickly recognised (e.g. texture, contrast, colour), but also in high-level features, which had to be interpreted (e.g. scene or object depicted). Note that the experimental design does not exactly match the classic oddball paradigm, because the probabilities of target and non-target stimuli were equal. *Non*-target and odd images did not fit the expectations of the participant, stood out in the 'regular train of standard stimuli' [45], and might be compared to the so called *target* stimuli of the classic oddball paradigm. For this reason, the late positive complex may appear to be inverted at the first glance (see an alternative explanation below).

Images were often fixated only once (see section 3.2). Thus, the longest fixation was in many cases the first and the last fixation at the same time. The distributions of the eye tracking features corresponding to the three image categories (target, non-target, odd) overlap, but are clearly not the same (see figure 9). Target images were, in general, fixated longer and more frequently than non-target images. Thus, an image was more likely followed by a target image than by a non-target (or odd) image, even though the presentation probability was the same for target and non-target images (see section 2.1). Imbalanced dwell times and transition probabilities may have systematically distorted the event-related potentials at later time points, when the next image was already fixated, and could have resulted in the found late positive complex.

4.3. Task performance

The participants complied with the task instructions, because the images were skimmed quickly and not inspected thoroughly, as suggested by the comparably short time spent on each result page and the rather low counting accuracy (see section 3.3 and figure 10).

5. Conclusion

The study shows that EEG and eye tracking signals can be used to infer the subjective relevance of screen content. This implicit information can be extracted from the signals in the background and makes it possible to create dynamic user interest profiles in real-time without an explicit relevance feedback from the user. A whole new range of applications can be conceived on the basis of the introduced technologies, even though the purpose of use presented in this paper is rather specific (ambiguities in image search were resolved). Computer users could navigate rapidly through large data sets with little effort using novel interfaces tailored to the implicit relevance feedback from the sensors. Eye tracking is especially promising considering the progress made with regard to technology and cost [23]. Nevertheless, recently developed miniaturised EEG systems with dry electrodes can be set-up quickly and hassle-free (see section 1), and a small set of electrodes may be sufficient, because central areas of the scalp were particularly informative (see section 3.2 and figures 6 and 7). While both measurement modalities turned out to be complementary (see sections 3.1 and 4.1), information provided by eye tracking might vanish in a more realistic setting (but is nevertheless required for the feature extraction from the EEG). Discriminative information present in fixation duration and dwell time could be corrupted when the user starts pondering and interrupts the flow of the eye movements. In contrast, spatio-temporal patterns in short fixation-related EEG epochs may remain unaffected. Besides, EEG contained information about the relevance of the single images, which could be used for more fine-grained user interest profiles (see figure 4), in contrast to eye tracking, which allowed only for estimating the relevance of the entire page (see figure 5).

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 611570. The work of Benjamin Blankertz was additionally funded by the Bundesministerium für Bildung und Forschung under contract 01GQ0850.

ORCID iDs

Markus A Wenzel https://orcid.org/0000-0002-6540-1476

References

[1] Blankertz B *et al* 2010 The Berlin brain-computer interface: non-medical uses of BCI technology *Frontiers Neurosci.* **4** 198

[2] Müller K-R, Tangermann M, Dornhege G, Krauledat M, Curio G and Blankertz B 2008 Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring *J. Neurosci. Methods* **167** 82–90

[3] Zander T O and Kothe C 2011 Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general *J. Neural Eng.* **8** 025005

[4] Blankertz B, Acqualagna L, Dähne S, Haufe S, Schultze-Kraft M, Sturm I, Uščumlić M, Wenzel M A, Curio G and Müller K-R 2016 The Berlin brain-computer interface: progress beyond communication and control *Frontiers Neurosci.* **10** 530

[5] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

[6] Lebedev M A and Nicolelis M A L 2006 Brain-machine interfaces: past, present and future *Trends Neurosci.* **29** 536–46

[7] Dornhege G R, del R J, Hinterberger T, McFarland D and Müller K-R (ed) 2007 *Toward Brain-Computer Interfacing* (Cambridge, MA: MIT Press)

[8] Mak J N and Wolpaw J R 2009 Clinical applications of brain-computer interfaces: current state and future prospects *IEEE Rev. Biomed. Eng.* **2** 187

[9] Wolpaw J and Wolpaw E W (ed) 2012 *Brain-Computer Interfaces: Principles and Practice* 1st edn (Oxford: Oxford University Press)

[10] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23

[11] Treder M S, Schmidt N M and Blankertz B 2011 Gaze-independent brain-computer interfaces based on covert attention and feature attention *J. Neural Eng.* **8** 066003

[12] Acqualagna L and Blankertz B 2013 Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP) *Clin. Neurophysiol.* **124** 901–8

[13] Wenzel M A, Almeida I and Blankertz B 2016 Is neural activity detected by ERP-based brain-computer interfaces task specific? *PLoS One* **11** 1–16

[14] Kamienkowski J E, Ison M J, Quiroga R Q and Sigman M 2012 Fixation-related potentials in visual search: a combined EEG and eye tracking study *J. Vis.* **12** 4

[15] Brouwer A-M, Reuderink B, Vincent J, van Gerven M A J and van Erp J B F 2013 Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials *J. Vis.* **13** 17

[16] Kaunitz L N, Kamienkowski J E, Varatharajah A, Sigman M, Quiroga R Q and Ison M J 2014 Looking for a face in the crowd: fixation-related potentials in an eye-movement visual search task *NeuroImage* **89** 297–305

[17] Kauppi J-P, Kandemir M, Saarinen V-M, Hirvenkari L, Parkkonen L, Klami A, Hari R and Kaski S 2015 Towards brain-activity-controlled information retrieval: decoding image relevance from MEG signals *NeuroImage* **112** 288–98

[18] Uščumlić M and Blankertz B 2016 Active visual search in non-stationary scenes: coping with temporal variability and uncertainty *J. Neural Eng.* **13** 016015

[19] Wenzel M A, Golenia J-E and Blankertz B 2016 Classification of eye fixation related potentials for variable stimulus saliency *Frontiers Neuroprosthetics* **10** 23

[20] Finke A, Essig K, Marchioro G and Ritter H 2016 Toward FRP-based brain-machine interfaces—single-trial classification of fixation-related potentials *PLoS One* **11** e0146848

[21] Wenzel M A, Bogojeski M and Blankertz B 2017 Real-time inference of word relevance from electroencephalogram and eye gaze *J. Neural Eng.* **14** 056007

[22] Piwek L, Ellis D A, Andrews S and Joinson A 2016 The rise of consumer health wearables: promises and barriers *PLoS Med.* **13** e1001953

[23] Dalmaijer E 2014 Is the low-cost eyetribe eye tracker any good for research? *PeerJ PrePrints* **2** e585v1

[24] Stopczynski A, Stahlhut C, Larsen J E, Petersen M K and Hansen L K 2014 The smartphone brain scanner: a portable real-time neuroimaging system *PloS One* **9** e86733

[25] De Vos M, Gandras K and Debener S 2014 Towards a truly mobile auditory brain-computer interface: exploring the P300 to take away *Int. J. Psychophysiol.* **91** 46–53

[26] Mullen T R, Kothe C A E, Chi Y M, Ojeda A, Kerth T, Makeig S, Jung T-P and Cauwenberghs G 2015 Real-time neuroimaging and cognitive monitoring using wearable dry EEG *IEEE Trans. Biomed. Eng.* **62** 2553–67

[27] Popescu F, Fazli S, Badower Y, Blankertz B and Müller K-R 2007 Single trial classification of motor imagination using 6 dry EEG electrodes *PLoS One* **2** e637

[28] Grozea C, Voinescu C D and Fazli S 2011 Bristle-sensors-low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications *J. Neural Eng.* **8** 025008

[29] Zander T O, Lehne M, Ihme K, Jatzev S, Correia J, Kothe C, Picht B and Nijboer F 2011 A dry EEG-system for scientific research and brain-computer interfaces *Frontiers Neurosci.* **5** 1–10

[30] Guger C, Krausz G, Allison B Z and Edlinger G 2012 Comparison of dry and gel based electrodes for P300 brain-computer interfaces *Frontiers Neurosci.* **6** 60

[31] Nikulin V V, Kegeles J and Curio G 2010 Miniaturized electroencephalographic scalp electrode for optimal wearing comfort *Clin. Neurophysiol.* **121** 1007–14

[32] Looney D, Kidmose P and Mandic D P 2014 Ear-EEG: user-centered and wearable BCI *Brain-Computer Interface Research (Biosystems & Biorobotics* vol 6) ed C Guger *et al* (Berlin: Springer) pp 41–50

[33] Debener S, Emkes R, De Vos M and Bleichner M 2015 Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear *Sci. Rep.* **5** 16743

[34] Norton J J S *et al* 2015 Soft, curved electrode systems capable of integration on the auricle as a persistent brain-computer interface *Proc. Natl Acad. Sci.* **112** 3920–5

[35] Goverdovsky V, Looney D, Kidmose P and Mandic D P 2016 In-ear EEG from viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring *IEEE Sensors J.* **16** 271–7

[36] Goverdovsky V, von Rosenberg W, Nakamura T, Looney D, Sharp D J, Papavassiliou C, Morrell M J and Mandic D P 2017 Hearables: multimodal physiological in-ear sensing *Scientific Reports* **7** 6948

[37] Flickr www.flickr.com/

[38] Flickr API documentation www.flickr.com/services/api/

[39] Wikipedia list of true homonyms http://en.wikipedia.org/wiki/List_of_true_homonyms

[40] Venthur B, Dähne S, Höhne J, Heller H and Blankertz B 2015 Wyrm: a brain-computer interface toolbox in Python *J. Neuroinformatics* **13** 471–86

[41] Venthur B and Blankertz B 2012 Mushu, a free-and open source BCI signal acquisition, written in python *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* (IEEE) pp 1786–8

[42] Blankertz B, Lemm S, Treder M, Haufe S and Müller K-R 2011 Single-trial analysis and classification of ERP components—a tutorial *NeuroImage* **56** 814–25

[43] Friedman J H 1989 Regularized discriminant analysis *J. Am. Stat. Assoc.* **84** 165

[44] Ledoit O and Wolf M 2004 A well-conditioned estimator for large-dimensional covariance matrices *J. Multivariate Anal.* **88** 365–411

[45] Picton T W 1992 The P300 wave of the human event-related potential *Clin. Neurophysiol.* **9** 456–79