Real-time inference of word relevance from electroencephalogram and eye gaze

M A Wenzel, M Bogojeski and B Blankertz

Technische Universität Berlin, Fachgebiet Neurotechnologie, Sekr. MAR 4-3, Marchstr. 23, 10587 Berlin, Germany

E-mail: markus.wenzel@tu-berlin.de

CrossMark

Abstract

Objective. Brain-computer interfaces can potentially map the subjective relevance of the visual surroundings, based on neural activity and eye movements, in order to infer the interest of a person in real-time. *Approach.* Readers looked for words belonging to one out of five semantic categories, while a stream of words passed at different locations on the screen. It was estimated in real-time which words and thus which semantic category interested each reader based on the electroencephalogram (EEG) and the eye gaze. *Main results.* Words that were subjectively relevant could be decoded online from the signals. The estimation resulted in an average rank of 1.62 for the category of interest among the five categories after a hundred words had been read. *Significance.* It was demonstrated that the interest of a reader can be inferred online from EEG and eye tracking signals, which can potentially be used in novel types of adaptive software, which enrich the interaction by adding implicit information about the interest of the user to the explicit interaction. The study is characterised by the following novelties. Interpretation with respect to the word meaning was necessary in contrast to the usual practice in brain-computer interfacing where stimulus recognition is sufficient. The typical counting task was avoided because it would not be sensible for implicit relevance detection. Several words were displayed at the same time, in contrast to the typical sequences of single stimuli. Neural activity was related with eye tracking to the words, which were scanned without restrictions on the eye movements.

Keywords: brain-computer interfacing, electroencephalography, eye movements, reading, relevance detection, semantics, unrestricted viewing

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

1. Introduction

Electromagnetic fields of the brain and eye movements may carry information about the subjective relevance of the single items present in the visual surrounding. This implicit information can potentially be decoded in real-time in order to infer the current interest of the individual person. Previous research on brain-computer interfacing (BCI) has shown that it can be estimated which stimuli aroused the interest, when a stimulus sequence is viewed—by detecting multivariate patterns in non-invasive recordings of the brain activity (e.g. [1, 2]).

However, familiar stimuli are typically presented again and again in BCI, and can therefore be easily recognised, regardless of whether they are letters, pictures of faces, geometric shapes or merely colours (e.g. [2–7]). In contrast, the regular visual environment contains items that have to be interpreted with respect to their meaning, most notably words in the case of written text. The interpretation of the semantics goes beyond the simple recognition of a previously known letter, picture, or shape that is repeatedly flashed (see [8] for a comparison).

Accordingly, the question was addressed if the relevance inference from the electroencephalogram (EEG) can be also
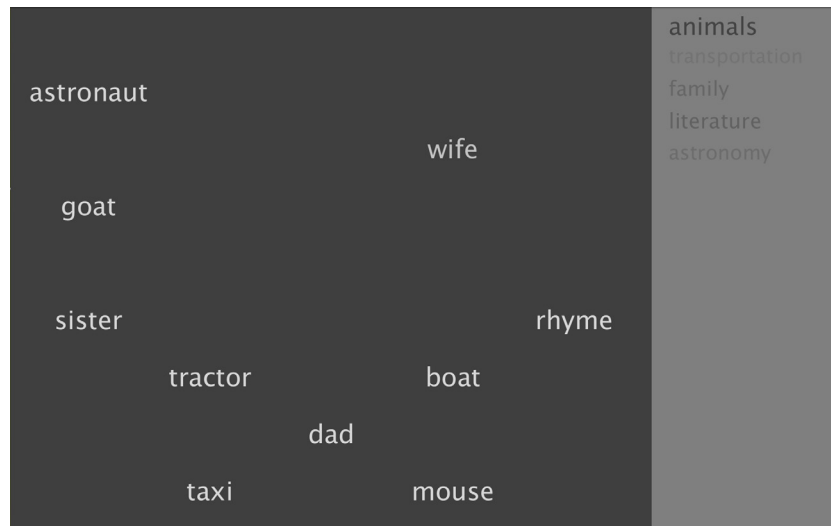
Figure 1. The participants looked for words (left) related to one out of five semantic categories (right). The semantic category of interest was estimated in real-time during the online phase of the experiment, based on implicit information present in the EEG and eye gaze. The current interest estimates were represented by the luminances of the five category names.

applied in settings where semantic content has to be interpreted. Readers looked for words belonging to one out of five semantic categories, while a stream of words passed at different locations on the screen (see figure 1). The words were dynamically replaced (when they had been fixated with the eye gaze) by new words fading in. It was estimated in real-time during the experiment which words and thus which semantic category interested the reader, based on information implicitly contained in the measured EEG and eye tracking signals. The estimates were visualised for demonstration purposes on the edge of the screen, and were updated as soon as a new word had been read. In this way, the reader could learn about the current estimates (for each of the five categories), and could observe how evidence was accumulated over time. Prior to the online inference (see section 2.2), a classifier had to be trained to estimate the word relevance based on the signals (see section 2.1).

In contrast to recent investigations with similar objectives [9–12], several words were displayed at the same time on the screen. The participants could scan the words without restrictions on the eye movements. Neural activity was related with eye tracking to the respective word looked at, like in studies on reading (e.g. [13–17]) and on visual search, that have shown that sought-for items evoke a detectable neural response when they are fixated with the eye gaze (see [18–25]).

The subjective relevance of the visual surrounding can be mapped with this approach by assigning relevance scores to the single items in view. The obtained information could be aggregated in order to characterise the current interest of the individual person. The resulting dynamic user interest profile would render possible novel types of adaptive software and personalised services, which enrich the interaction between human and computer by adding implicit information to the explicit interaction (see [11, 12, 21, 25–30]). Less obtrusive and more convenient EEG systems with sufficient signal quality are prerequisite for the application in practice (see [31–36]).

2. Materials and methods

2.1. Calibration

Labelled EEG and eye tracking data were recorded in order to train a classifier that could predict the relevance of the single words in the subsequent online phase (see section 2.2). The participants selected one out of five given semantic categories. Subsequently, twenty-two words were drawn randomly from the five categories, with a contribution of 20% per category on average. Words faded in on the screen at predefined positions in random order (see figure 1), and were faded out when they had been fixated with the eye gaze (with a delay of one second).

Examples of the categories and words are:

- Astronomy: orbit, galaxy, universe, meteorite.
- Time: future, seconds, hourglass, minute.
- Furniture: bathtub, closet, stool, bed.
- Transportation: taxi, canoe, tractor, helicopter.
- Visual art: palette, pencil, sculpture, crayon.

The participants were requested to remember the words that belonged to the chosen category. When the participants had looked at all words, they were asked to recall the relevant words from their memory. For this purpose, the words reappeared truncated (to about 40% of the original number of letters) at shuffled positions. Relevant words had to be selected with the mouse. Subsequently, the accuracy of the recall was checked and reported. This procedure helped to involve the participants in the task in order to mimic intrinsic interest in certain words but avoided interference of motor activity during the acquisition of the EEG data.

For the study, a corpus had been generated of seventeen semantic categories with twenty words each, both in English and German depending on the language skills of the participant (see section 3.1). The seventeen categories were: animals, furniture, transportation, body parts, family, food, literature, country names, astronomy, music, finance, buildings
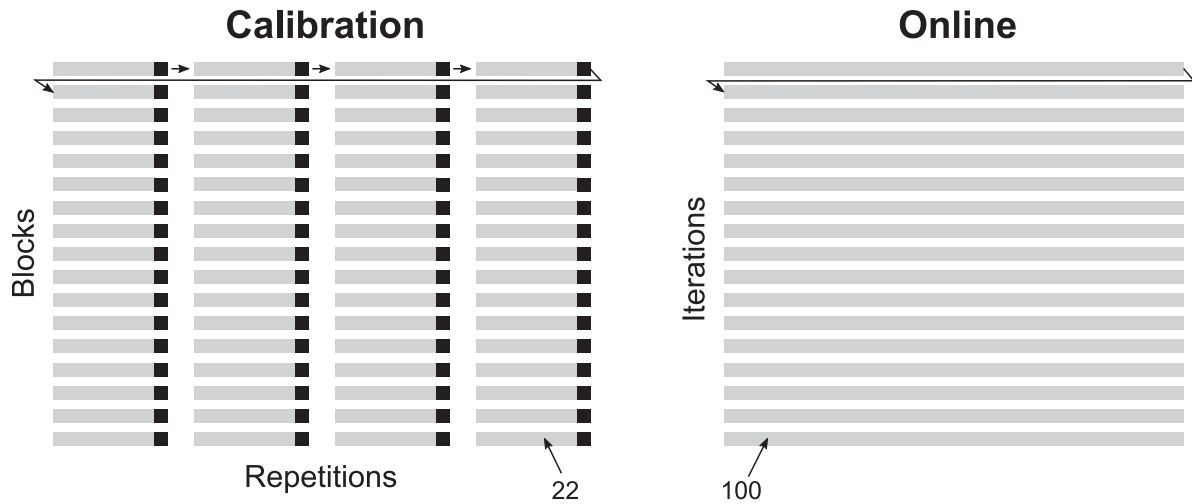
Calibration

Online

Figure 2. Overview of the procedure during the experiment. During the *calibration* phase, there were seventeen blocks with different semantic categories of interest. Each block was split into four repetitions. In each repetition, twenty-two words were viewed. After each repetition, words related to the respective category of interest had to be recalled (symbolised by black squares). During the *online* phase, there were seventeen iterations with different semantic categories of interest. In each iteration, hundred words were viewed, while feedback on the estimated interest was given in real-time.

and structures, healthcare, sports, time, clothes, and visual art. The calibration phase consisted in seventeen blocks with four repetitions each (see figure 2). At the beginning of each block, a semantic category (out of five options) could be chosen. The categories offered for selection changed during the course of the experiment. It was possible, but not necessarily the case, that each of the seventeen categories was chosen once, because the selection was not restricted. During the recording, it was tracked which category had been chosen by the participant and thus which single words were relevant.

Feature vectors were extracted from the recorded EEG and eye tracking data with the intention to capture processes related to word reading and categorisation (details below). The feature vectors were labelled depending on whether the word fixated at this moment was relevant or irrelevant to the chosen category of interest. Subsequently, a classification function was trained with regularized linear discriminant analysis [37] to discriminate the feature vectors of the 'relevant' and the 'irrelevant' class [1]. The shrinkage parameter was calculated with an analytic method [38, 39].

2.1.1. Feature extraction. The multi-channel EEG signal was re-referenced to the linked mastoids and low-pass filtered (with a second order Chebyshev filter; 42 Hz pass-band, 49 Hz stop-band). The continuous signal was segmented by extracting the interval from 100 ms to 800 ms after the onset of every eye fixation. Slow fluctuations in the signal were removed by baseline correction (i.e. by subtracting the mean of the signal within the first 50 ms after fixation onset from each epoch). The signal was downsampled from the original 1000 Hz to 20 Hz in order to decrease the dimensionality of the feature vectors to be obtained (14 values per channel). A low dimensionality in comparison to the number of available samples is beneficial for the classification performance, because the risk of overfitting to the training data is reduced [1]. The multi-channel signal

was vectorised by concatenating the values measured at the 62 scalp EEG channels at the 14 time points resulting in a $62 \times 14 = 868$ dimensional vector per epoch. The fixation duration was concatenated as additional feature to the EEG feature vector.

Note that other eye tracking features, e.g. the gaze velocity, could not be exploited, because they are not provided in real-time by the application programming interface of the device, and that two additional EEG electrodes, which were not situated on the scalp and served for re-referencing and electrooculography, were excluded from the set of 64 electrodes in total. The distance between the words and the font size were chosen such that the words had to be fixated for reading, which made it possible to relate the continuous EEG signal to the respective word looked at. However, it can not be excluded that some words could be recognised also in peripheral vision (see [23]). Eye-movement-related signal components were not removed from the EEG, which makes online operation simpler. Moreover, the employed multivariate methods can project out artefacts of various kinds.

2.2. Online prediction

The subjective relevance of words to a semantic category was inferred online with the previously trained classifier. Again, the participants read words and were asked to look for words related to one out of five semantic categories. The words faded in and out similar to the calibration phase but vacant positions were replaced by new words fading in. In this way, all hundred words of the five involved categories were shown per iteration (see figure 2). Usually, several words were present on the screen at the same time. The classifier predicted online for each fixated word if it was relevant to the category of interest or not, based on the incoming EEG and eye tracking data.

The class membership probability estimates for the single words were assigned to the corresponding semantic category

and all estimates obtained so far were averaged per category. The resulting five-dimensional vector indicated how likely each category was of interest. The vector was normalised to unit length, determined the font size and luminance of the visualisation of the five category names on the right side of the screen (see figure 1), and was updated when a new word had been fixated with the eye gaze. It was initialised with neutral values for the initial period when only few words had been read and not every category was captured. The participants were informed about the predictive mechanism underlying the adaptive visualisation in order to foster task engagement. The feedback may have driven strategies in the participants that would have not occurred otherwise. However, if there was no feedback, the participants would have had hardly any intrinsic motivation to look for words 'of interest'. Besides, relevance feedback would be also part of the envisioned novel types of adaptive software (see section 1).

A memorisation task like in the calibration phase (see section 2.1) was not included in view of the objective to exploit only *implicit* information. Otherwise, the detected information may be related to the memorisation and not to the subjective experience of finding a relevant word. The procedure was iterated seventeen times with new combinations of five categories (see figure 2). At the beginning of each iteration, the participants indicated the selected category of interest for later validation, and the previously collected relevance estimates were cleared.

The participants became more familiar with the corpus of words during the course of the experiment. Nevertheless, the participants had to read each word, interpret the word meaning, and decide if the word belonged to the chosen category of interest. In contrast, only a small set of few shapes/colours is repeatedly flashed in brain-computer interfacing and stimulus recognition is sufficient (see figure 1 in [4]).

The difference in the stimulus presentation between the calibration and the online phase has the following two reasons. (a) The words were not replaced during the calibration phase in order to limit the number of words to remember and thus the difficulty of the memory task. (b) The replacement during the online phase allowed for accumulating evidence over more data during one task iteration. Note that the spatial resolution of the eye tracker limits the words that can be displayed at the same time. For maximal similarity with the online phase, words faded in and out in the calibration phase as well.

Remark for the sake of completeness: the classifier output was dichotomised to zero or one in the actual visualisation during the experiment. In contrast, class membership probability estimates ranging between zero and one were employed for the figures presented in this paper.

2.3. Experimental setup

An apparatus was developed that allowed for making inferences from combined EEG and eye tracking data in real-time and displaying this information in an adaptive graphic visualisation.
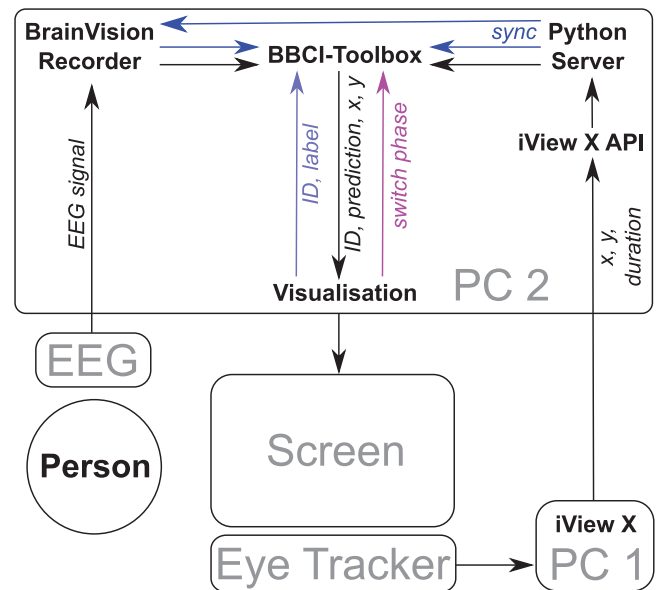

Figure 3. The apparatus allows for making inferences from EEG and eye tracking signals in real-time and displaying the obtained information in an adaptive graphic visualisation.

2.3.1. Key constituents of the system. The system comprised an EEG device, an eye tracker, two computers and a screen that the test person was looking at (see figure 3). EEG was recorded with 64 active electrodes arranged according to the international 10–20 system (*ActiCap*, *BrainAmp*, BrainProducts, Munich, Germany; sampling frequency of 1000 Hz). The ground electrode was placed on the forehead and electrodes at the linked-mastoids served as references. An eye tracker, connected to a computer (PC 1), detected eye fixations in real-time (*RED 250*, *iView X*, SensoMotoric Instruments, Teltow, Germany; sampling frequency of 250 Hz; details of the online fixation detection algorithm were not disclosed by the manufacturer). A second computer (PC 2) acquired raw signals from the EEG device (with the software *BrainVision Recorder*, BrainProducts, Munich, Germany), and obtained preprocessed eye tracking data from PC 1 over network using the *iView X API* and a custom server written in *Python 2.7* (https://python.org). EEG and eye gaze data were then streamed to in-house software written within the framework of the *BBCI-Toolbox* (https://github.com/bbci/bbci_public) running in *Matlab 2014b* (MathWorks, Natick, USA). The graphic visualisation was computed with custom software written in *Processing 3* (https://processing.org) and displayed on the screen (60 Hz, 1680×1050 pixel, $47.2 \,\text{cm} \times 29.6 \,\text{cm}$).

2.3.2. Synchronisation of EEG and eye tracking signals. When the data acquisition started, the Python server sent a sync-trigger into the EEG signal and transmitted the current time stamp of the eye tracker to the BBCI-Toolbox. These simultaneous markers allowed for synchronising the two measurement modalities.

2.3.3. Workflow of the system. The experiment included several phases, which could be switched by the visualisation software with messages sent over a TCP connection. During the
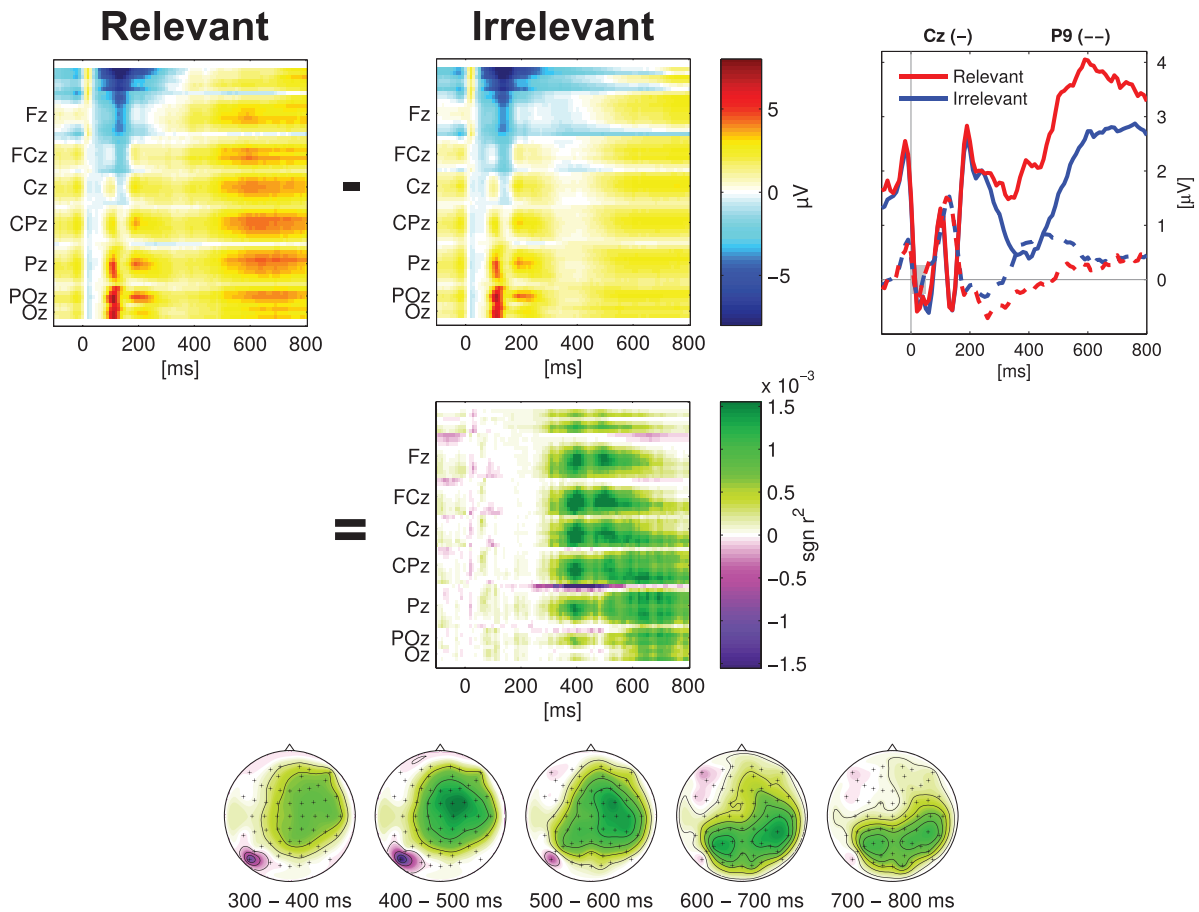
Figure 4. Patterns in the EEG differed when the word read was relevant to the category of interest or irrelevant (calibration phase). *Top*: EEG time series for relevant and irrelevant words (for all channels sorted from front to back and from left to right, and for two selected channels). *Centre*: Difference. *Bottom*: topographies of the difference.
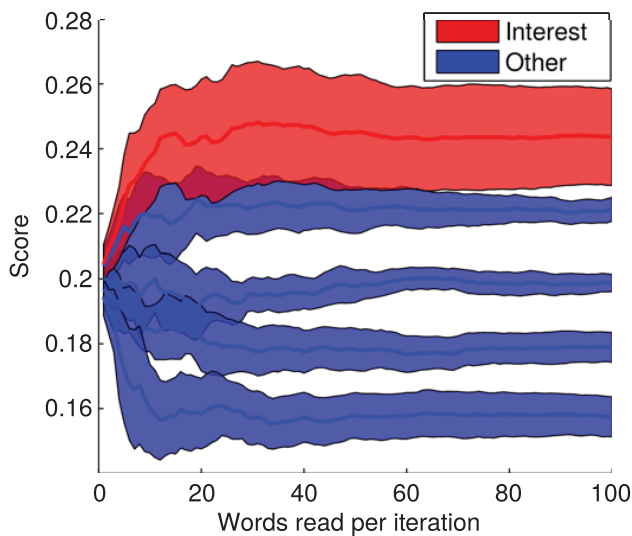

Figure 5. Evolution of the scores corresponding to the category of interest (red) and to the four other categories (blue, sorted according to the respective final score) during the online phase (combined EEG and gaze features). Tubes indicate the standard error of the mean.
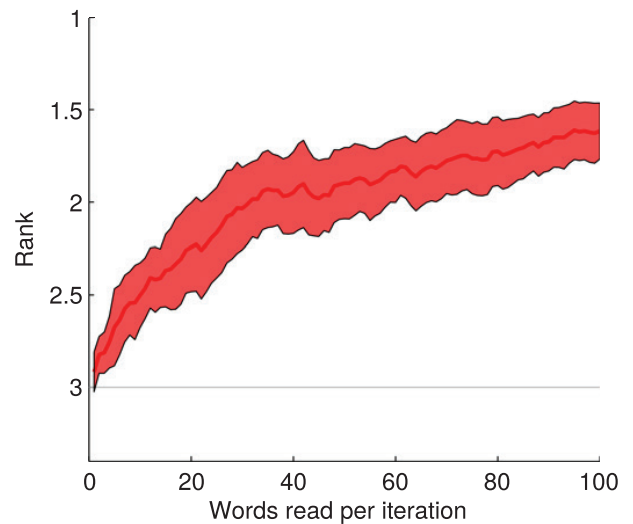
Figure 6. Evolution of the rank of the category of interest among the five categories during the online phase (combined EEG and gaze features; note the direction of the *y*-axis with the top rank of 1 on top; the shaded area indicates the standard error of the mean).

calibration phase, EEG and eye tracking data were recorded to train a model (see section 2.1) that was supposed to predict the relevance of each word read by the subject in the subsequent online phase (see section 2.2). Feature vectors were extracted from the ongoing EEG and eye tracking signals, every time the eye tracker had detected a new eye fixation (see section 2.1). The visualisation software checked if the eye fixation was

situated on a word displayed on the screen, according to the received *x*-*y*-coordinates. During the calibration phase, the feature vectors were labelled, depending on whether the word belonged to the category of interest or not. Labels and feature vectors were matched according to a unique identifier (ID) of each eye fixation. During the online phase, the graphic visualisation adapted according to the incoming predictions. The architecture of the system is modular and the visualisation module can easily be replaced by other software for novel applications that depend on making real-time inferences from EEG and eye tracking signals. The communication protocol that enables the visualisation module to interact with the other parts of the system offers three types of interactions. The visualisation module can (a) switch between calibration and online phase and an initial adjustment of the eye tracker, (b) can receive relevance estimates from the BBCI-Toolbox, and (c) can mark events and stop data acquisition by sending markers into the EEG.

2.4. Data acquisition

Experiments with three female and twelve male participants with normal or corrected to normal vision, no report of eye or neurological diseases and ages ranging from 21 to 40 yr (median of 28 yr) were conducted while EEG, eye tracking and behavioural data were recorded. Ten people performed the experiment in their mother tongue of German and five people with other first languages accomplished the task in English, which was not their mother tongue. The subjects gave their informed written consent (a) to participate in the experiment and (b) to the publication of the recorded data in anonymous form without personal information. The study was approved by the ethics committee of the Department of Psychology and Ergonomics of the Technische Universität Berlin (reference BL_03_20150109).

3. Results

3.1. Calibration

The participants recalled the words that were relevant to the category of interest with an average accuracy of 80%, ranging from 72% to 84% in the individuals. Classifiers were trained individually for each participant to detect relevant words with EEG and eye tracking data recorded during the calibration phase (see section 2.1). In the subsequent online phase, the classifiers were applied to the data incoming in real-time (see section 2.2).

Additionally, the performance of the classifiers was assessed in ten-fold cross-validations using only the data recorded during the calibration phase. The area under the curve (AUC) of the receiver operating characteristic served as performance metric [40]. An AUC of 0.63 ± 0.01 (mean ± standard error of the mean) was measured for the single-trial classifications with EEG feature vectors from the calibration phase, which was significantly better than the

Table 1. Final rank of the category of interest in the online phase when hundred words per iteration had been read (averages over the seventeen iterations per participant, as well as over all participants). The combined and the single modalities are listed separately.

Participant	EEG & Gaze	EEG	Gaze
1	1.29	1.35	2.59
2	1.12	1.12	1.18
3	1.53	1.65	4.59
4	1.53	1.47	2.00
5	2.12	2.47	2.24
6	1.76	1.76	2.53
7	1.06	1.12	1.76
8	1.53	1.53	2.41
9	1.47	1.47	2.76
10	1.65	1.65	2.00
11	1.88	1.88	3.29
12	1.76	2.06	1.47
13	2.00	2.00	1.88
14	1.65	1.71	2.82
15	1.88	1.94	1.53
Mean ± SEM	1.62 ± 0.08	1.68 ± 0.09	2.34 ± 0.22

chance level of 0.5 ($Z = 3.37$, $p < 0.05$). Adding the fixation duration as extra feature did not improve the results, the AUC remained at the same level (significantly better than chance; $Z = 3.37$, $p < 0.05$). When only the fixation duration served as feature, an AUC of 0.51 ± 0.01 was obtained, which was not significantly better than chance ($Z = 1.05$, $p > 0.05$, Bonferroni corrected for the three Wilcoxon signed rank tests on the population level).

Furthermore, the EEG patterns corresponding to relevant and irrelevant words were characterised in order to understand on which processes the classification success was based on (see figure 4). The EEG signal was inspected that followed the landing of the eye gaze on the words. The onset of the eye fixation was situated at $t = 0$ ms. Early components (until about 150 ms) were related to the saccade offset (respectively the fixation onset) and occurred equally in both conditions. Later components differed depending on whether the word was relevant or irrelevant. Relevant words evoked a left lateralised posterior negativity in comparison to irrelevant words and a positivity that shifted from fronto-central to parietal sites on both hemispheres. For this analysis, all EEG epochs of all participants were averaged separately for relevant and irrelevant words (see figure 4, top) and the difference between the two classes was assessed with signed squared biserial correlation coefficients (see figure 4, centre and bottom). Each time point measured at each EEG electrode was treated separately in order to characterise the spatio-temporal evolution. A significance threshold was not applied in order to show also subtle differences that can potentially be detected by a multivariate classifier.

Relevant words were fixated for about 227.4 ms ± 8.7 ms and irrelevant words for about 216.8 ms ± 7.8 ms during calibration (mean ± standard error of the mean). A paired t-test detected a significant difference between the two classes on the population level; $t(14) = 4.3$, $p < 0.05$.
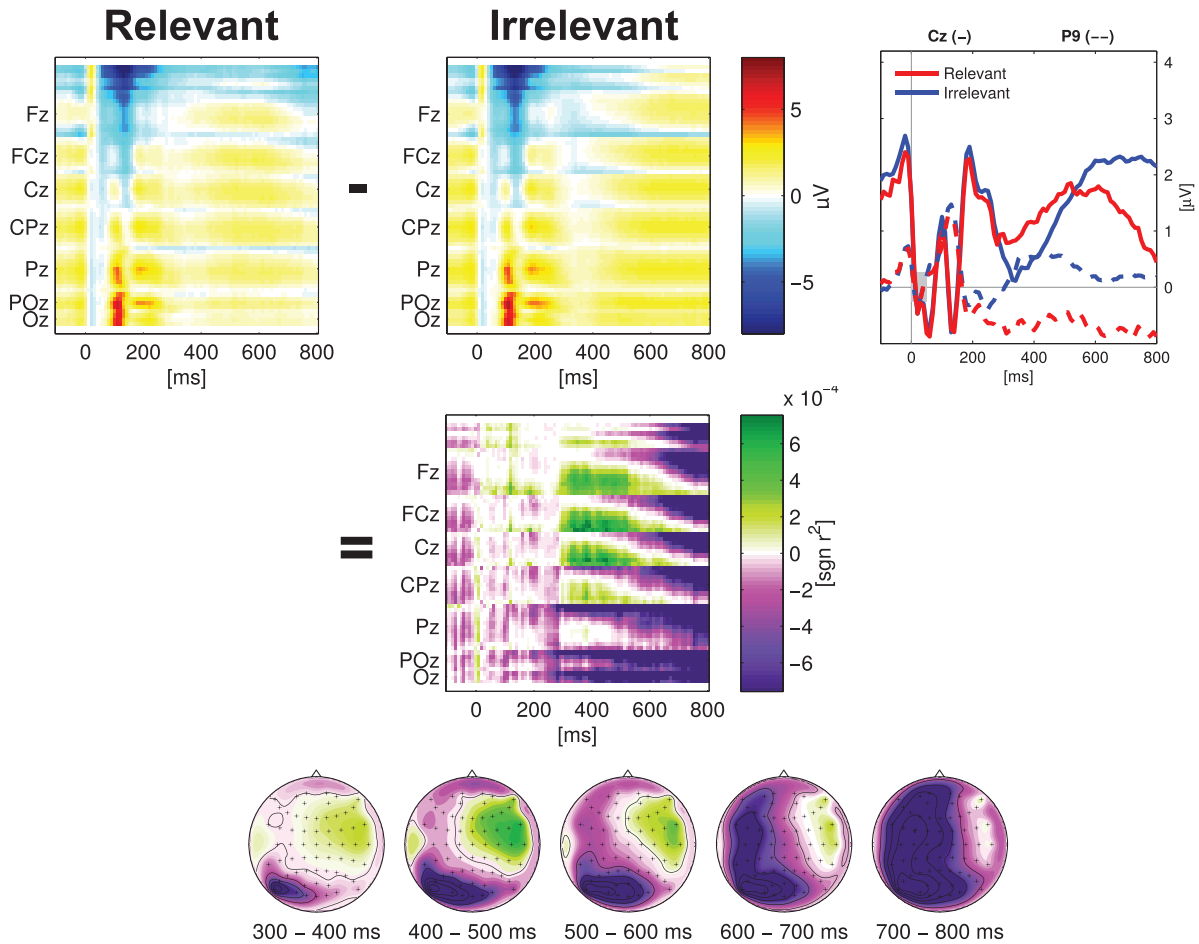
Figure 7. EEG patterns during the online phase. *Top*: EEG time series for relevant and irrelevant words (for all and for two selected channels). *Centre*: difference. *Bottom*: topographies of the difference.

3.2. Online prediction

The previously trained classifiers were applied during the online phase to the incoming data and it was predicted for each word if it was relevant to the category of interest or not. The class membership probability estimates were averaged per semantic category and the obtained five-dimensional vector was normalised to unit length (see section 2.2). Figure 5 displays the evolution of the resulting scores corresponding to the category of interest and to the four other categories, which were sorted according to the respective final score (combined EEG and gaze features; average over all participants). With more words being read by the participant, the score of the category of interest grew in comparison to the other categories. Note that the (blue) score curves of the four 'other' categories in figure 5 diverge due to a selection effect: for each iteration, the 'other' categories were ranked according to their final score (other #1, ⋯, other #4) and the statistics were calculated separately for each of those ranks, across iterations. The ranking allows for comparing the score curve of the category of interest (red) with the best competitor per iteration (top blue curve). Without the ranking, the blue curves would look alike.

Figure 6 shows the evolution of the rank of the category of interest among the five semantic categories (combined EEG and gaze features; average over all participants). The category of interest started with an average rank of three and moved towards the top of the ranking with more words being read (note the direction of the *y*-axis).

Table 1 lists the average final rank of the category of interest for each single participant (i.e. when all hundred words per iteration had been read; see section 2.2). The predictions were based on feature vectors including either the EEG data or the fixation duration, or a combination of the two measurement modalities (columns in the table). The final rank was below three in every single participant when only EEG features were used and even smaller when the fixation duration was added as extra feature. Deploying the fixation duration as single feature resulted in a comparably large final rank. On the population level, the final rank was significantly below three for all feature types ($Z_{\text{EEG\&Gaze}} = -3.38, Z_{\text{EEG}} = -3.38, Z_{\text{Gaze}} = -2.41, p < 0.05$, Bonferroni corrected for the three Wilcoxon signed rank tests).

Figure 7 displays the EEG patterns during the online phase for relevant and irrelevant words. Relevant words evoked a posterior negativity and a central positivity in comparison to irrelevant words, which is similar to the calibration phase (see figure 4). Additionally, a negativity arose on the left hemisphere in the online phase, in contrast to the calibration phase.

Relevant words were fixated for about $239.5\,ms \pm 12.4\,ms$ and irrelevant words for about $208.2\,ms \pm 7.0\,ms$ during the online phase (mean $\pm$ standard error of the mean). The two classes differed significantly on the population level according to a paired *t*-test; $t(14) = 4.7, p < 0.05$.

4. Discussion

4.1. Calibration

All participants complied with the task instructions because they recalled the words that were relevant to the selected semantic category with an accuracy of at least 72% (giving random answers would result in an expected accuracy of about 20% due to the five possible categories). EEG and eye tracking signals recorded during the calibration phase were used to train classifiers (individually for each participant) to discriminate relevant words from irrelevant words.

The trained EEG-based classifiers were able to generalise to unseen data, because the cross-validation results with calibration data were significantly better than it can be expected from random guessing (see section 3.1; note that the AUC served as straightforward metric here, in contrast to the online phase where the ranking of the categories provided a more descriptive metric). Classification was apparently possible because relevant words evoked a different neural response than irrelevant words (see section 3.1 and figure 4). In previous research on brain-computer interfacing, the stimuli of interest evoked a similar neural response with a left lateralised negativity and a central positivity (see figure 2, right panel, in [4]), even though the stimuli used in the cited study were not words but geometric shapes flashed on the screen while the eyes did not move. Hence, it was shown with the present investigation that the methods developed for brain-computer interfacing can be employed for inferring the relevance of words under unrestricted viewing conditions.

Concatenating the fixation duration to the feature vectors did not improve the predictive performance, and single-trial classification based on the fixation duration alone was not possible better than random (when data from the calibration phase were used). Nevertheless, a small but significant difference of the fixation duration between the two classes was found on average (see section 3.1).

4.2. Online prediction

It was predicted in real-time which words were relevant for the reader, who was looking for words related to a semantic category of interest. The five categories were ranked according to the normalised five-dimensional average score vector. Perfect prediction of the category of interest would have resulted in a score of 1 and a rank of 1 for the category of interest. If each word was classified randomly as relevant or irrelevant, an average score of 0.2 and an average rank of 3 can be expected. The score and the rank of the category of interest started at this chance level, as it can be assumed. With more words being read, the score grew and the rank decreased (see figures 5 and 6).

Apparently, evidence could be accumulated by integrating information over the incoming single predictions.

The combination of EEG and fixation duration resulted in the best predictive performance (see table 1). The gaze did not contribute much to the relevance estimate because features from the EEG alone were more informative than when the fixation duration was used as single feature (while it has to be considered that information about the eye gaze is required for the EEG feature extraction, because the EEG signals had to be related to the corresponding words looked at; see section 2.1).

The successful transfer of the classifiers from the calibration phase to the online phase is reflected in the underlying data. The EEG patterns, that made it possible to distinguish relevant and irrelevant words, evolved similarly in the calibration and in the online phase in the first period after fixation onset (see figures 4 and 7). The later discrepancy is presumably a result of the different tasks, because the relevant words had to be memorised only in the calibration phase. Moreover, during the online phase, fixated words were replaced by new words fading in, the words were already familiar, and relevance feedback was displayed (see sections 2.1 and 2.2). Despite of these differences, generalisation from the calibration to the online phase was possible. The discriminative EEG patterns may correspond to two components of the event related potential: the 'P300', which is associated with attention mechanisms (and subsequent memory processing) [41, 42], and the 'N400', which is related to language processing [43]. The found fixation durations might be comparable to the average numbers reported in the literature, e.g. of $224\,ms \pm 25$ ms in table 1 in [14].

4.3. Conclusion

The study demonstrates that the subjective relevance of words for a reader can be inferred from EEG and eye gaze in real-time. The methods employed are rooted in research on brain-computer interfacing based on event-related potentials, where stimulus recognition is usually sufficient, and where sequences of single stimuli are typically flashed. In contrast, the investigation presented here is characterised by the requirement to interpret words with respect to their semantics. Furthermore, several words were presented at the same time and neural activity was related with eye tracking to the respective word read. The typically employed counting task was avoided because it would not be sensible for *implicit* relevance detection (see [44]). The task instruction during the online phase was merely to look for (and not to count) words relevant to the category of interest. In this way, the subjective experience of encountering a relevant word should be approximated, which can be vague in comparison to the well-defined counting task.

Task engagement was additionally fostered by explaining the predictive mechanism underlying the adaptive visualisation. The experiment exploits a situation that allows for integrating implicit information across several single words. In a next step, the methods could be applied to a situation where sentences or entire texts are being read, which will entail a number of new challenges for the data analysis, because the

single words are syntactically and semantically interdependent in this case. While this study serves as a proof-of-principle, the methods can potentially be used in the future for mapping the subjective relevance of the field of view in novel applications (see section 1). In summary, this study represents a further step towards inferring the interest of a person from information implicitly contained in neurophysiological signals.

Acknowledgments

References

[1] 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

[2] 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

[3] Kaufmann T, Schulz S M, Grünzinger C and Kübler A 2011 Flashing characters with famous faces improves ERP-based brain–computer interface performance *J. Neural Eng.* **8** 056016

[4] 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

[5] Acqualagna L, Treder M S and Blankertz B 2013 Chroma speller: isotropic visual stimuli for truly gaze-independent spelling *6th Int. IEEE/EMBS Conf. on Neural Engineering* (IEEE) pp 1041–4

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

[7] Seoane L F, Gabler S and Blankertz B 2015 Images from the mind: BCI image evolution based on rapid serial visual presentation of polygon primitives *Brain Comput. Interfaces* **2** 40–56

[8] Wenzel M A, Moreira C, Lungu I-A, Bogojeski M and Blankertz B 2015 Neural responses to abstract, linguistic stimuli with variable recognition latency *Symbiotic Interaction* (*Lecture Notes in Computer Science* vol 9359) ed B Blankertz *et al* (Berlin: Springer) pp 172–8

[9] Geuze J, van Gerven M A J, Farquhar J and Desain P 2013 Detecting semantic priming at the single-trial level *PLoS One* **8** e60377

[10] Geuze J, Farquhar J and Desain P 2014 Towards a communication brain computer interface based on semantic relations *PLoS One* **9** e87511

[11] Eugster M J A, Ruotsalo T, Spapé M M, Kosunen I, Barral O, Ravaja N, Jacucci G and Kaski S 2014 Predicting term-relevance from brain signals *Proc. of the 37th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval* pp 425–434

[12] Eugster M J A, Ruotsalo T, Spapé M M, Barral O, Ravaja N, Jacucci G and Kaski S 2016 Natural brain-information interfaces: recommending information by relevance inferred from human brain signals *Sci. Rep.* **6** 38580

[13] Baccino T and Manunta Y 2005 Eye-fixation-related potentials: insight into parafoveal processing *J. Psychophysiol.* **19** 204–15

[14] Dimigen O, Sommer W, Hohlfeld A, Jacobs A M and Kliegl R 2011 Coregistration of eye movements and EEG in natural reading: analyses and review. *J. Exp. Psychol. Gen.* **140** 552

[15] Dimigen O, Kliegl R and Sommer W 2012 Trans-saccadic parafoveal preview benefits in fluent reading: a study with fixation-related brain potentials *NeuroImage* **62** 381–93

[16] Kliegl R, Dambacher M, Dimigen O, Jacobs A M and Sommer W 2012 Eye movements and brain electric potentials during reading *Psychol. Res.* **76** 145–58

[17] Kornrumpf B, Niefind F, Sommer W and Dimigen O 2016 Neural correlates of word recognition: a systematic comparison of natural reading and rapid serial visual presentation *J. Cogn. Neurosci.* **28** 1374–91

[18] 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

[19] 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

[20] 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

[21] 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

[22] Golenia J-E, Wenzel M A and Blankertz B 2015 Live demonstrator of EEG and eye-tracking input for disambiguation of image search results *Symbiotic Interaction* (Berlin: Springer) pp 81–6

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

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

[25] 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

[26] Pohlmeyer E A, Wang J, Jangraw D C, Lou B, Chang S-F and Sajda P 2011 Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases *J. Neural Eng.* **8** 036025

[27] 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

[28] Ušćumlić M, Chavarriaga R and Millán J d R 2013 An iterative framework for EEG-based image search: robust retrieval with weak classifiers *PLoS One* **8** e72018

[29] Jangraw D C, Wang J, Lance B J, Chang S-F and Sajda P 2014 Neurally and ocularly informed graph-based models for searching 3D environments *J. Neural Eng.* **11** 046003

[30] Blankertz B, Acqualagna L, Dähne S, Haufe S, Schultze-Kraft M, Sturm I, Ušćumlić 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

[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 and wearable BCI *Brain-Computer Interface Research* (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 Sens. 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 2016 Hearables: multimodal physiological in-ear sensing (arxiv:1609.03330)

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

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

[39] Schäfer J and Strimmer K 2005 A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics *Stat. Appl. Genet. Mol. Biol.* **4** 16646851

[40] Fawcett T 2006 An introduction to ROC analysis *Pattern Recognit. Lett.* **27** 861–74

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

[42] Polich J 2007 Updating P300: an integrative theory of P3a and P3b *Clin. Neurophysiol.* **118** 2128–48

[43] Kutas M and Federmeier K D 2011 Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP) *Annu. Rev. Psychol.* **62** 621–47

[44] Wenzel M A, Almeida I and Blankertz B 2016 Is neural activity detected by ERP-based brain-computer interfaces task specific? *PLoS One* **11** e0165556