Active visual search in non-stationary scenes: coping with temporal variability and uncertainty

Marija Ušćumlić and Benjamin Blankertz

Neurotechnology group, Technische Universität Berlin, Berlin, Germany

E-mail: marija.uscumlic@gmail.com

Abstract

Objective. State-of-the-art experiments for studying neural processes underlying visual cognition often constrain sensory inputs (e.g., static images) and our behavior (e.g., fixed eye-gaze, long eye fixations), isolating or simplifying the interaction of neural processes. Motivated by the non-stationarity of our natural visual environment, we investigated the electroencephalography (EEG) correlates of visual recognition while participants overtly performed visual search in non-stationary scenes. We hypothesized that visual effects (such as those typically used in human–computer interfaces) may increase temporal uncertainty (with reference to fixation onset) of cognition-related EEG activity in an active search task and therefore require novel techniques for single-trial detection. *Approach*. We addressed fixation-related EEG activity in an active search task with respect to stimulus-appearance styles and dynamics. Alongside *popping-up* stimuli, our experimental study embraces two composite appearance styles based on *fading-in, enlarging*, and *motion* effects. Additionally, we explored whether the knowledge obtained in the *pop-up* experimental setting can be exploited to boost the EEG-based intention-decoding performance when facing transitional changes of visual content. *Main results*. The results confirmed our initial hypothesis that the dynamic of visual content can increase temporal uncertainty of the cognition-related EEG activity in active search with respect to fixation onset. This temporal uncertainty challenges the pivotal aim to keep the decoding performance constant irrespective of visual effects. Importantly, the proposed approach for EEG decoding based on knowledge transfer between the different experimental settings gave a promising performance. *Significance*. Our study demonstrates that the non-stationarity of visual scenes is an important factor in the evolution of cognitive processes, as well as in the dynamic of ocular behavior (i.e., dwell time and fixation duration) in an active search task. In addition, our method to improve single-trial detection performance in this adverse scenario is an important step in making brain–computer interfacing technology available for human–computer interaction applications.

S Online supplementary data available from stacks.iop.org/jne/13/016015/mmedia

Keywords: EEG, fixation related potentials, active visual search, eye tracking, temporal uncertainty

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

1. Introduction

Information gathering in active visual search is guided by eye trajectory—a sequence of eye fixation positions that we make while performing a visual search task. Cognitive neural

processes in active viewing are typically studied in relation to the fixation onset, which requires simultaneous recording of electroencephalography (EEG) and eye tracking (see [1]). The emerging research on eye fixation-related potentials (EFRPs) [2–5] has shown that when the eye is fixating on the object of search, the event-related EEG potential (ERP) resembles the one in the classical visual oddball paradigm with the constrained gaze—e.g., the visual speller [6–8] and the search under the rapid serial visual presentation (RSVP) protocol [9–15]. Likewise, fixation-related EEG potentials are shown to be discriminative with respect to the attended object (target versus non-target depending on the search task).

Methods for the discriminative EEG analysis, inherited from studies on gaze-constrained paradigms, mainly operate in a temporal domain with the assumption that the potentials are evoked in an event-locked manner [16, 17]. In other words, it is assumed that a minimal trial-by-trial variation of ERP latency (i.e., minimal temporal uncertainty) occurs. Justification for this assumption is found in highly controlled experiments relating to stimulation timing, stimuli type, and subjects' behavior. Up-to-date EEG studies on active viewing therefore mainly use simplified stimuli (e.g., static scenes consisting of scattered artificial objects) and/or impose limited ocular behavior (e.g., longer fixations). While this controlled approach is successful in generating precise results concerning fundamental neurocognitive processes, it ignores some aspects of our natural viewing behavior that are of pivotal importance for a shift towards investigations in real-world conditions. For general human–computer interaction applications, in particular, the approach needs to be relaxed in several aspects—to include, for example, free viewing, diverse natural stimuli, and non-stationary scenes.

In this study, we specifically investigated the EEG correlates of visual recognition while participants overtly perform a search task in non-stationary scenes. To model scene non-stationarity, we rendered stimuli into the scene using several composite visual effects consisting of *fading-in*, *enlarging*, and *motion*. Visual effects made it so that no sudden-onset stimuli appear in subjects' visual fields. Instead, the appearance was smooth and gradual, mimicking our natural visual environment. We hypothesized that visual effects might enhance temporal uncertainty in recognition events with respect to fixation onset, affecting the latency and/or morphology of the corresponding EEG activity. To this end, we find particularly noteworthy the recent finding of a domain-independent decision signal in an EEG when performing a perceptual decision task [18–20]. Specifically, the evolution of a decision-signal was found to trace the dynamics of the evidence accumulation process, driven by sensory input.

Relatively little research has been conducted on EEG decoding in active viewing when facing more realistic visual stimuli. In an EEG study on active visual searches of the target face [21], the authors ensured original complexity of natural scenes by using images of crowds at stadiums as stimuli. Prior to the experiment, however, subjects were trained to carry out a certain behavior that promoted longer fixations. Visual recognition in a driving scenario was recently addressed in a joint EEG and eye tracking study for the efficient navigation of 3D naturalistic environments [22]. The results demonstrated that ocular data (i.e., dwell time, pupil dilation) may complement the EEG, improving target object detection in free-view tasks. Recently, a replica of previous statistical EEG studies on active search tasks was conducted using a more ecological paradigm [23]. Natural indoor and outdoor pictures appeared as stimuli, while no constraints on subjects visual behavior were imposed. The authors addressed the overlap issue of the EEG activity elicited by consecutive fixations with various duration, in particular, as concerns the interpretation of the EEG potentials. Neither of these studies, however, addressed the temporal uncertainty of the recognition event caused by content diversity or continuous scene changes.

On the other hand, a recent EEG study [24] addressed the recognition of time-evolving visual events. As stimuli, the authors used videos of an actor/avatar imitating several real-life human behaviors (e.g., leave/take a bag, wave). Videos were presented in the narrow field of view so that minimal changes in the gaze occurred. Regardless of the necessary integration of static and dynamic visual features for behavior recognition, the presence of robust, discriminative evoked EEG responses was observed. A decline in EEG decoding performance, however, was reported for one type of the events where a greater variation in the locking time of the event occurred.

The significance of modeling temporal variability for EEG decoding was previously demonstrated in an experiment on a short video RSVP search with a manual response upon target recognition [25, 26]. Designing a classification method that accounts for temporal variability in the neural response, the authors showed an improvement in the decoding performance compared to the state-of-the-art classification method.

In our study on active visual search in non-stationary scenes, we used modified *Landolt rings* as stimuli, which allowed the control of the semantic level of the search throughout the experiment. This is particularly important because a variability in the content semantic level is a potential source of temporal uncertainty in EEG responses [27, 28]. We applied discriminant analysis to evaluate the EEG-based decoding of recognition (i.e., target versus non-target). We further explored how in non-stationary scenes the performance could be preserved despite visual effects, knowing that state-of-the-art EEG decoding approaches are challenged by ERP's temporal uncertainty. This is of high significance for a potential symbiosis of brain–computer interface technology with real-world human–computer interaction that aims at constant performance irrespective of scene dynamics. As an illustration, intuitive human computer interfaces (HCIs) involve interactive entertainment and computer games in which visual effects are found strongly entangled.

2. Materials and methods

2.1. Experimental setup

In this study, EEG and eye movements were recorded simultaneously while subjects were performing a task. The eye movements were recorded with a remote eye tracker (RED 250, SMI, Teltow, Germany; sampling frequency of 250 Hz), attached to the protocol presentation screen. The screen resolution was 1680 by 1050 pixels (47.2 by 29.6 cm), corresponding to a visual angle of ~42° by 27°. A chin rest was used to reduce head movements while keeping the distance between eyes and screen constant (~61 cm). Physiological signals were recorded with two EEG amplifiers with 63 active EEG electrodes in total (BrainAmp, ActiCap, Brain-Products, Munich, Germany; sampling frequency of 1000 Hz). The EEG signals were re-referenced to the linked mastoid. We conducted the experimental study with sixteen healthy participants (13 male, 3 female, age 19–48). Ethical approval for the experiment was granted by the local ethic committee—Ethikkommission des Instituts für Psychologie und Arbeitswissenschaft (IPA) der TU Berlin. Following the ethical requirements, the participants gave informed written consent to take part in the study.

2.2. Experimental protocols

We used eight broken Landolt rings corresponding to the eight different directions of their openings as stimuli. While a traditional Landolt ring is uniformly colored (figure 1(a)), we modified the rings with a black/gray pattern (the intensity value of gray was set to 0.5) as illustrated in figure 1(b) (for more details, see the supplementary material). A target Landolt ring (with respect to the direction of its opening) was shown to the subject at the beginning of each stimuli sequence and he/she was instructed to (silently) count how many times it occurred. At the end of the sequence, the subject reported the corresponding number. The chance of a target appearing in the stimuli sequence was 25%. Target-stimulus frequency was higher than the frequency of individual non-target stimuli which was about 10.7%. We raised the probability of targets to 25%, since 12.5% was considered too little and using only four directions too small variability in stimuli.

Three different conditions of stimuli presentation were considered: 'pop-up' (PU), 'smooth appearance' (SA), and 'motion appearance' (MA). Stimuli presentation in these conditions is illustrated in figure 1(c).

PU. Stimuli pop up one by one with a random time interval between them (1–1.5 s).

SA. Stimuli appear smoothly, fading in one-by-one, linearly increasing their size and opacity. Stimulus disclosure in the full size and opacity therefore takes 1 s and 2 s respectively. Stimuli appear at fixed positions on the screen with a random interval between them (1–1.5 s).

In both PU and SA conditions, each presented stimulus starts to slowly fade out 5 s after appearing. The next stimulus appears at a random position, but in a limited screen area with respect to the position of its predecessor. The distance between two successive stimuli was in the range of 170–480 pixels, corresponding to a visual angle from 4.5° to 12°. Following a stimuli sequence of random length (50–80 stimuli), the screen was erased and the next task was set. Variable-length sequences were used to discourage anticipation of the end of sequences.

MA. In regard to motion appearance, stimuli move in either top-down or bottom-up direction. All stimuli start in minimized size (dots) at the top or bottom central location and move slowly downwards or upwards. Movement directions were counterbalanced across sequences. The velocity of the stimuli ranges from 1.85 to 2.25°/s. While its vertical component was constant across the stimuli (1.85°/s), its horizontal component depends on the initial horizontal position of the stimuli from the central top or bottom position (within +/−10% of the screen width), resulting in a velocity range of 0 to 1.4°/s. Stimuli were continuously enlarged and their transparency continuously decreases until they reach the center of the screen height (i.e., the complete stimulus enclosure requires 6 to 7.5 s depending on its initial horizontal position in the screen) when their transparency starts being modified in the opposite manner while the size was preserved. Hence, compared to the SA condition, the manipulation of the size and intensity of the stimuli was slower.

Finally, the level of the background gray color of the screen was set to 0.58 for PU and SA, and 0.7 for the MA condition. Consequently, in the MA condition the contrast between the gray pattern (the level of 0.5) and the background is higher, facilitating perception. The length of the sequences ranged from 50 to 80 stimuli for the PU and SA conditions, while it was set to 100 stimuli for the MA condition.

The experiment was organized in blocks of six sequences (i.e., two search tasks for each condition). On average, subjects performed twelve search tasks per condition.

2.3. Data preprocessing and EFRP extraction

2.3.1. Ocular data. We used the SMI high speed event detection algorithm of BeGaze2 (peak velocity threshold: 40 °/s, min fixation duration: 50 ms) to detect eye-movement events, i.e., fixations, saccades, and blinks (IDF Event Detector, SensoMotoric Instruments, Teltow, Germany, http://www.smivision.com, 2014). In addition, we analyzed ocular behavior by comparing *gaze response* and *dwell time* across different experimental conditions and stimulus types (target versus non-target). To this end, the region of interest (ROI) was introduced because stimuli recognition requires foveal vision. We defined the ROI as the area enclosed by a circle with a diameter of 100 pixels centered on a stimulus, corresponding to the visual angle of ~2.5° at viewing distances of ~61 cm. First, we look into duration histograms of the fixations within ROI. We compared the distributions between the first fixations on target and non-target stimuli for each condition separately. We measured the *gaze response* as the time between the stimulus appearance in the scene until the first fixation to the corresponding ROI. Additionally, we measured the *dwell time* as the time duration of attending to
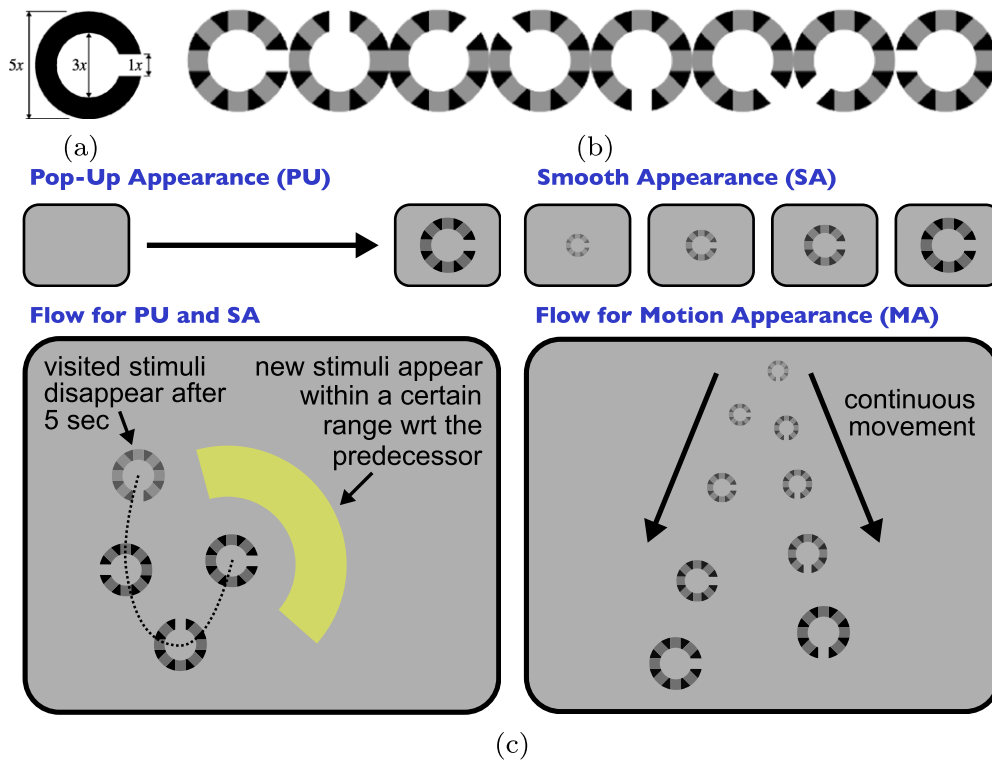
Figure 1. (a) A standard Landolt broken ring. (b) Eight modified Landolt rings that we used in our study. (c) Illustration of the stimuli presentation flow for different conditions. (top-left) Three intermediate steps of stimulus evolution in time are presented for the SA condition, followed by a completely revealed stimulus. (bottom-right) The dashed line indicates the order of stimuli appearance in the PU and SA conditions. (bottom-left) Several intermediate steps of the evolution of two successive stimuli are illustrated for the MA condition. The arrows indicate the direction of their continuous motion. This illustration is simplified, since multiple objects were present on the screen in the MA condition (see section 2.2). Note: stimuli are enlarged, compared to the real screen dimension.

the ROI between the first fixation to it and the gaze shift to the ROI of the next stimulus that is attended. Finally, we inspected to see whether there was a significant difference in these measures between target and non-target stimuli. To this end, we applied Wilcoxon rank-sum and Bonferroni correction for testing across conditions, for each subject independently.

2.3.2. Fixation-related EEG potentials.
First, the EEG signals were filtered between 1 and 30 Hz (Butterworth filter) and downsampled to 250 Hz. Then, single trial EEG activity was extracted with respect to the fixation. Specifically, we considered only the first fixation on the ROI after the appearance of a stimulus on the screen.

2.3.3. Eye movement artifacts.
Eye movements, as an inseparable aspect of free-viewing tasks, represent a significant source of artifacts in the corresponding EEG data. Therefore, we applied the 'interference subtraction' method to reduce eye-related artifacts in the EEG data [16]. Contaminated EEG data prior to the start of the experiment were used instead of the calibration session. In particular, we considered saccadic intervals. Eye-tracker data were used for detecting the time intervals of the contamination. Following the linear model, i.e., the linear relation between the scalp potentials and the sources, the activity generated by eye

movements was subtracted from the recorded EEG activity. To this end, the forward matrix corresponding to the source of artifacts was estimated as a component that captures the maximum difference in the signal during the contaminated period (left versus right horizontal, up versus down vertical movements figure 9).

2.4. Discriminative analysis

2.4.1. LDA-based classification methods.
Numerous successful classification approaches in brain–computer interfacing (BCI) are based on linear discriminant analysis (LDA), despite its remarkable simplicity. Here, we will consider two of them (i) shrinkage LDA, and (ii) hierarchical discriminant component analysis (HDCA), where both demonstrated good performance in a high-dimensional-data low-sample-size setting.

In contrast with classical LDA, shrinkage LDA implies the regularization of the empirical covariance matrix by shrinkage, while retaining LDA's simplicity [17]. On the other side, HDCA reduces the complexity of the classification problem by disregarding correlations of single-channel potential values across different time intervals. Additionally, it uses a more complex (nonlinear) classification step for combining classifier outputs from individual time intervals. HDCA classifiers therefore consist of two steps. First, one individual LDA classifier was trained per each of the non-

overlapping time intervals within a trial. Then, the LDA classifiers' outputs were combined by a logistic regression classifier to produce a final interest score for the given trial. The HDCA classifier has been successfully applied for single trial EEG analysis in visual BCI paradigms [9, 22, 29].

Concerning the temporal uncertainty (i.e., trial-to-trial temporal variability) of the EEG activity, the sliding HDCA (sHDCA) has been proposed [30]. It can be interpreted as an extension of the HDCA classifier with an additional level. First, the score space was obtained by applying an HDCA classifier, trained for a specific time interval, in a sliding manner i.e., with different time shifts with respect to the event onset. Then, in the decision space, the logistic regression that was trained to combine the obtained scores (i.e., the outputs of the HDCA for different shifts) was applied for the final decision.

2.4.2. Spatio-temporal localization of the discriminant activity.
A shrinkage LDA was applied for each EEG channel independently, using a time window between 100 and 800 ms of the fixation onset. As a result, a spatial distribution of classification performance in the ten-fold cross-validation setting indicates the most informative locations of the EEG channels for the discrimination between the target and non-target fixation. On the other hand, the most informative time intervals for discrimination were detected by estimating the peak classification performance on a time window of 50 ms in a sliding manner over a trial (with a temporal shift of 20 ms) by exploiting all EEG channels.

We also visualized a separability measure (target versus non-target) for each pair of channels and time points over a trial. As a separability measure we considered a signed squared value ($\text{sign} r^2$) of point-biserial correlation coefficients between the EEG activity and the stimuli type labels (target versus non-target).

2.4.3. Intra- and inter-protocol EEG decoding.
We based our discriminative analysis on the shrinkage LDA and the classical HDCA as explained further in the text. Additionally, we suggest a novel approach for dealing with the temporal uncertainty motivated by sHDCA classification method and transfer learning principle.

Intra-protocol decoding approach—We estimated classification performance of an HDCA classifier for each of the protocols separately, in a ten-fold crossvalidation setting. To this end, the classification interval spanned from 100 to 800 ms of the fixation onset (non-overlapping windows of 100 ms).

Inter-protocols decoding approach—Assuming that sufficient degree of phase locking of the EEG activity appears in the 'PU' condition, we trained an HDCA classifier on the most discriminant time interval (see section 2.4.2) using only the data from that condition. Then, we applied the classifier in a sliding manner with a step size of 50 ms, to the data from the other two conditions independently. The starting position was at 200 ms. For each position of a sliding window (seven positions in total) the posterior probability that target recognition happens was estimated. The final decision was made considering only the maximum of the obtained posterior probability estimates per trial.

Compared to the sliding HDCA, in our *inter-protocol decoding approach*, we preserved the same principle for creating the score space—a single HDCA was applied in a sliding manner. In contrast to sHDCA, however, we use the classifier trained on data of different origin (another experimental paradigm) than the actual data to be classified. Furthermore, in our approach, no additional training was required in the decision space because the final decision is based on the maximum of the estimated scores.

Intra-protocol decoding with a sliding time window—The inter-protocol decoding approach was motivated by transfer learning. Yet, it additionally relaxes the requirement of temporal locking of the ERP by introducing a sliding window. We therefore also performed the intra-protocol decoding in a sliding manner to support our hypothesis—the major advantage of the proposed inter-protocol approach originates in transferring the knowledge from the PU paradigm to the other paradigms.

Electrical activity caused by eye movements interferes with EEG, but predominantly in the frontal scalp regions. Hence, as eye-movement data in active viewing are linked to visual cognition, we also compared the intra-protocol classification results when exploiting all the EEG channels with the case when [Fp1, Fp2, F9, F10 and AF7, AF8] channels are excluded.

The area under the receiver operating characteristic curve (AUC) was used as a measure of classification performance. The chance level of the classifier was estimated using a random permutation procedure, where the data labels were randomly permuted in the training step in ten-fold cross-validation setting.

3. Results

3.1. Eye movements and behavioral data analysis

Subjects were engaged in the search task by counting the appearance of the target Landolt symbol and reporting the final score. Subjects correctly responded (i.e., the percentage of runs in which the correct number of targets was reported) to 89.55% (PU), 88.18% (SA), and 46.89% (MA) stimuli sequences. Considering only the incorrect responses, the differences between the subjects' responses and the actual number of targets were on average $-0.61 +/- 1.08$ (PU), $-0.077 +/- 2.24$ (SA) and $-1.25 +/- 1.06$ (MA). Note that the average error in responses was low in all three conditions.

In all the three visual presentation protocols, the appearance of a new stimulus (i.e., the change in the content of non-stationary scene) drives our gaze allocation. When the gaze is shifted to the ROI, as a response to the last presented stimulus, multiple fixations may be detected within it. We present, first, the fixation duration histograms distinguishing the fixation orders (figures 2(a)–(c)). Ocular-related artifacts
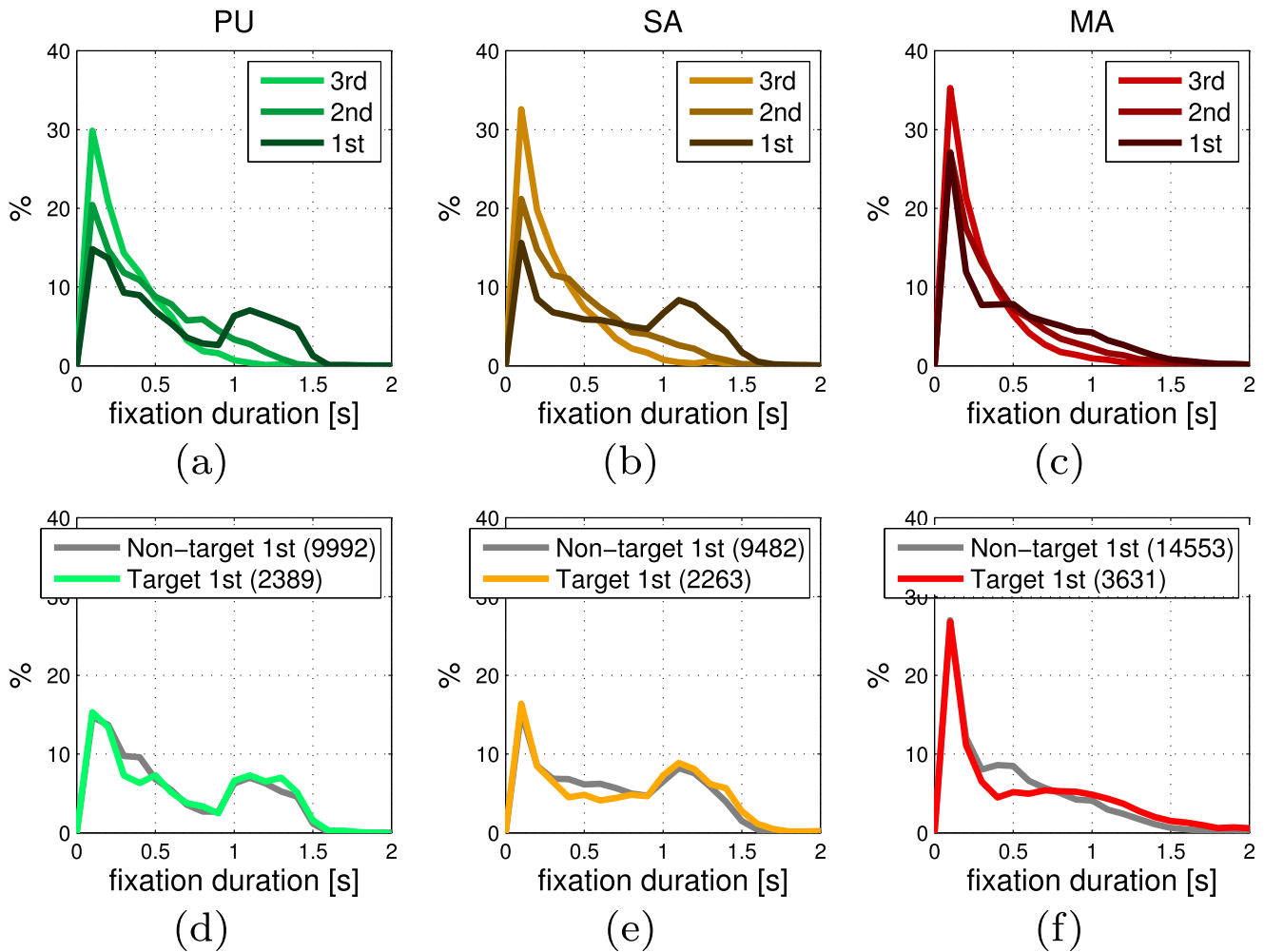
Figure 2. Histograms of the fixation duration: (a)–(c) distinguishing the order of the fixations within the ROI starting not later than 1 s from the stimulus onset, (d)–(f) the first fixations within the ROI distinguishing target and non-target stimuli (the total number of corresponding fixations is specified).

were induced in the EEG prior the onset of succeeding fixations. Therefore, a general characterization of this behavioral aspect is of interest for the analysis of the later EFRP's components. The first peak in all of the distributions across the protocols appears earlier than 100 ms. In the PU and SA conditions, however, the distribution of the first fixation duration was characterized by the second peak appearing after 1 s. The histograms of the first fixation durations, distinguishing fixations at target and non-target, are given in figures 2(d)–(f). A statistically significant difference between the fixation durations was found for 4 (PU), 5 (SA), and 9 (MA) subjects (out of 16 subjects) by applying the non-parametric Wilcoxon rank-sum test at 5% significance level (with the Bonferroni correction for testing across the three conditions).

Gaze response distributions across experimental conditions and stimuli types are presented in figure 3(a). The gaze shift happens most rapidly in the PU condition (the median gaze response time of 240 ms) with the smallest inter-trial variability. This is in line with the experimental design since a stimulus was entirely disclosed at the moment of appearance in the PU condition. In the SA condition, the visual effects

delay the gaze response (the median gaze response time being 590 ms). The most prolonged gaze response (the median gaze response time, 1.32 s) with the largest inter-trial variability was observed in the MA condition. No difference in gaze response between the stimuli types was observed in any condition.

Dwell time distributions across experimental conditions and stimuli types are presented in figure 3(b). The biggest difference in dwell time between the stimuli types was observed in the MA condition, where longer mean dwell time corresponds to target stimuli. The difference was found to be significant for 5 (PU), 6 (SA), and 15 (MA) subjects (out of 16 subjects). To this end, we applied the non-parametric Wilcoxon rank-sum test at a significance level of 5% (with the Bonferroni correction for testing across the three conditions).

3.2. Discriminative analysis

Grand average fixation-related EEG potentials are presented in figure 6, distinguishing target and non-target fixations. Scalp maps show the spatial distribution of the EEG components in the selected time intervals, indicated as two-tone
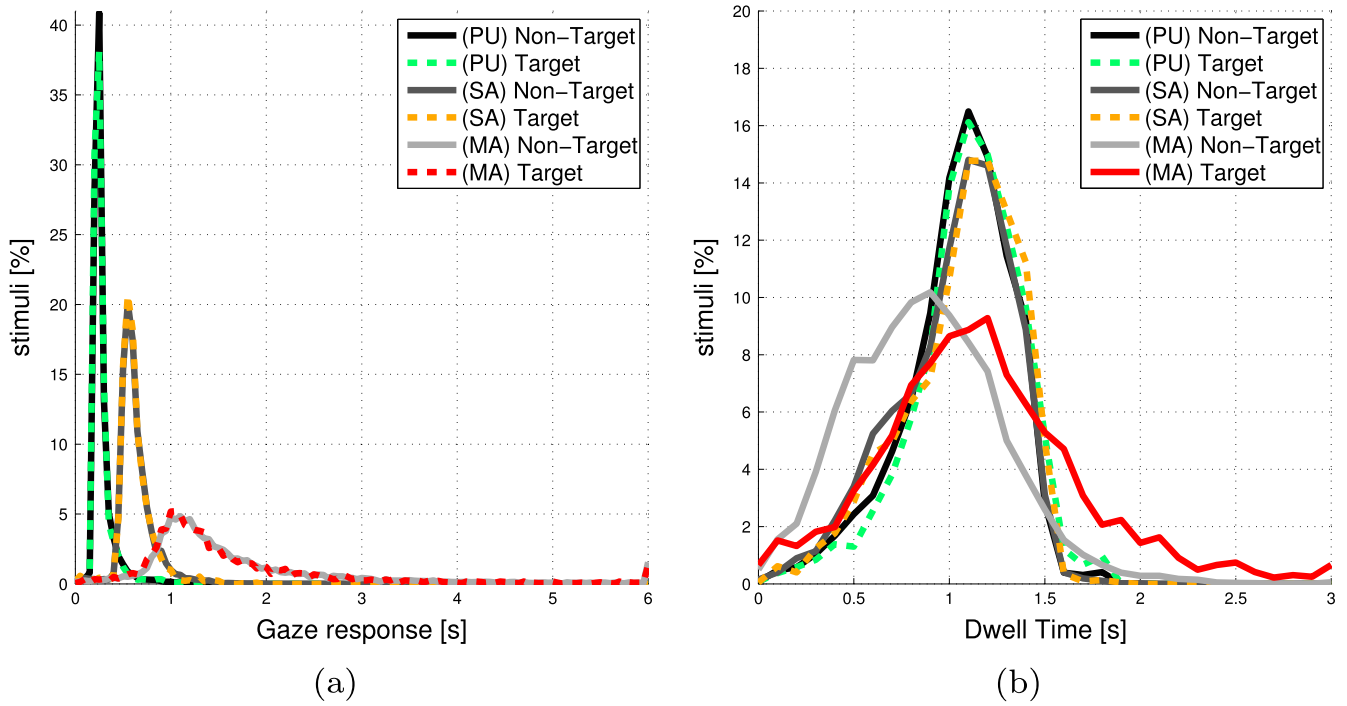
Figure 3. Ocular data (target versus non-target): (a) gaze response (GR) distribution—the time needed for orienting the gaze to the ROI of the new stimulus in the scene. Median GR values across protocols were 0.24 s (PU), 0.59 s (SA) and 1.32 s (MA); (b) dwell time distribution—the time eye-gaze resides within the ROI of the presented stimulus.
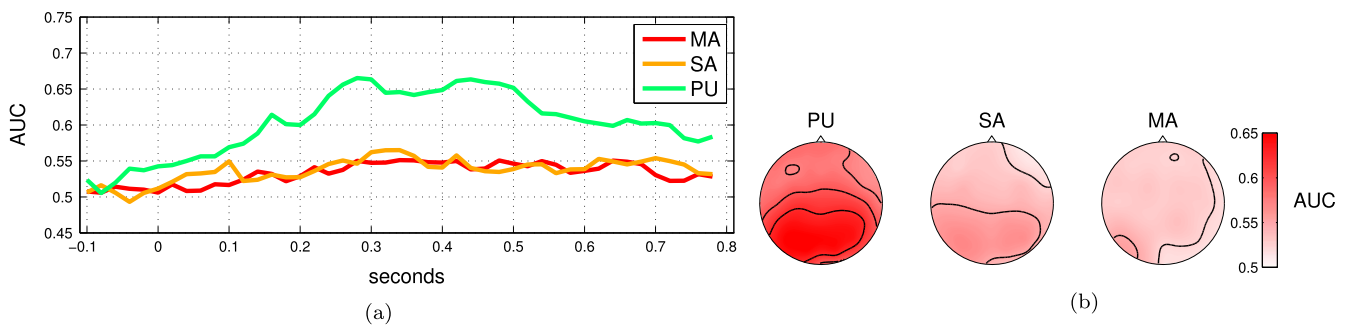

Figure 4. (a) Temporal localization of the discriminant activity—classification performance using the shrinkage LDA in a sliding manner (a time window of 50 ms with a temporal shift of 20 ms), exploiting all the EEG channels. (b) Spatial distribution of the classification performance in ten-fold cross-validation setting (single-channel based classification), exploiting the whole time interval from 100 to 800 ms.

gray bars. The separability measure (signr^2) for each pair channel and time point is presented in figure 5. Scalp maps correspond to the peaks of the separability measures.

The single-trial classification performance in the intra-protocol setting, is presented across presentation conditions and subjects in figure 7(a). Overall, the decoding of the EEG correlates of visual recognition locked to the fixation onset is above the chance level for all the presentation protocols (the Wilcoxon signed-rank test, at a 5% significance level). The highest decoding performance was observed in the PU condition. The AUC values are similar between the other two conditions. The intra-protocol classification performance, estimated in a sliding-window manner, is presented in figure 8.

The single-trial classification performance obtained in the inter-protocol setting (i.e., using the knowledge transfer

approach) is given across subjects in figures 7(b)–(c), for the SA and MA conditions respectively. The hierarchical classifier was trained on the most discriminant time interval from 200 to 600 ms (see figure 4), corresponding to four non-overlapping windows of 100 ms. Along with the final classification results (given in red), these plots contain the intermediate classification results for each time shift of the HDCA classifier (given in black). The peak performance across the time shifts (the intermediate classification results) appears at 100–150 ms, for SA and MA.

Finally, for all three conditions (PU, SA, MA) no significant difference in the classification performance was found when the very frontal EEG channels were excluded from the analysis (the Wilcoxon Signed-Rank test at the 1% significance level).
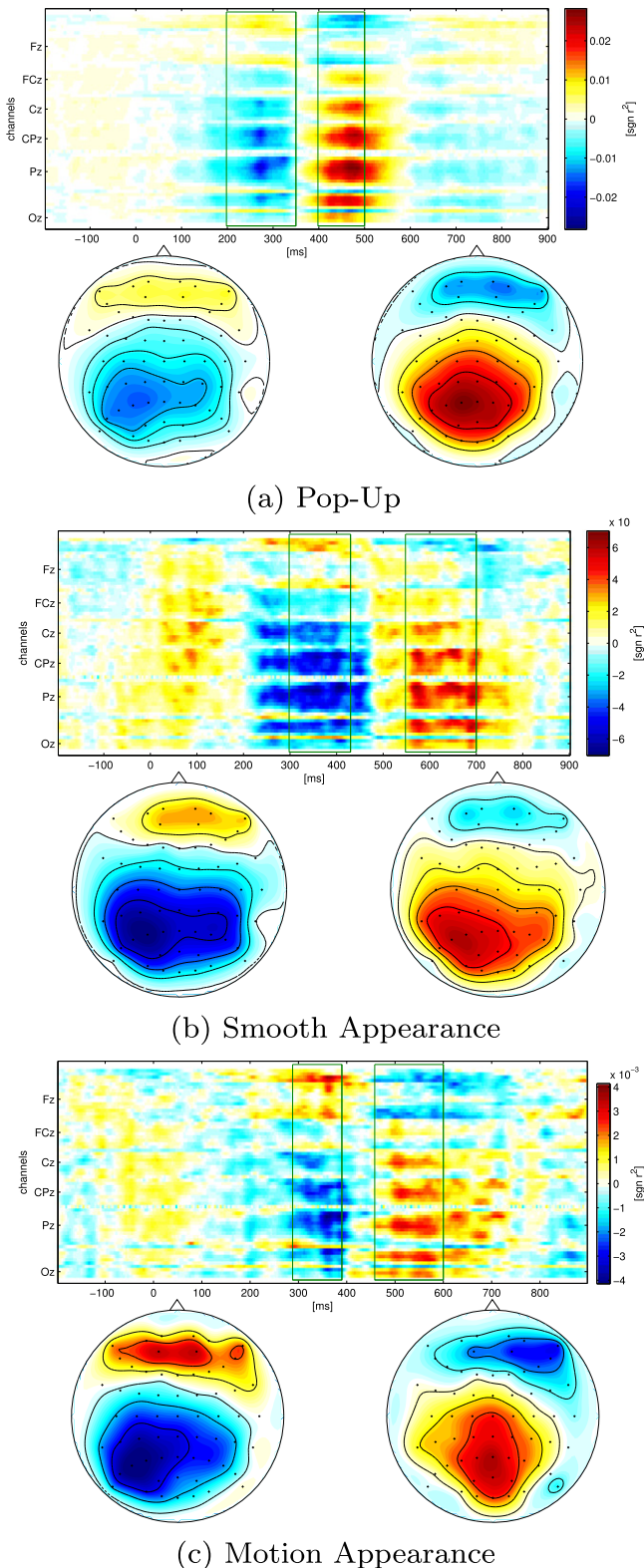
(a) Pop-Up

(b) Smooth Appearance

(c) Motion Appearance

Figure 5. Spatial distribution of sign r^2 values. Note, that the subplots have individual ranges for the colormap. The absolute value of r^2 cannot be compared across conditions. The temporal distribution of discriminative information is quite focused in PA and more scattered in SA and MA. Interestingly, the maps of discriminative information are quite similar across the three conditions.

4. Discussion

Static images are an inadequate replica of our visual surroundings. By way of illustration, HCI involves a dynamic visual content—the screen is constantly updated with the new information using various visual effects (e.g., changes in its opacity and size). Likewise, in driving scenarios, objects enter our visual scene by gradually becoming visible as we pass them. Additionally, beyond objects in reality we often look for events that extend in time (e.g., actions).

In our study on active visual searching, we investigated the EEG correlates of decision making about the content in dynamic scenes. It should be recognized that the aim of our experimental design was not to study the EEG correlates for isolated visual effects (e.g., the *fading−in effect* alone). On the contrary, we selected stimuli with composite appearance styles and design the protocols to represent several effects that may appear in real-world applications. Our research concerns whether scene dynamics might intensify the temporal ERP's uncertainty in active search with respect to the fixation onset.

4.1. Behavioral data

Our study is based on three experimental protocols: 'PU', 'SA' and 'MA'. All three protocols were designed in such a way that stimuli appearance guides shifts of visual attention and the gaze. In contrast with the 'PU' protocol, the other two include composite visual effects comprising *fading-in*, *enlarging* and *motion*.

The detected fixations are predominantly shorter in the MA condition (figure 2(c)) while in the PU and SA conditions a significant part of the fixations were of longer duration (see the peak in distributions around 1 s in figures 2(a) and (b)). This may be partly a result of a slow disclosure dynamic (i.e., fading—in and enlarging effects) and motion in the MA condition. Namely, in this scenario subjects overtly track stimuli upon their disclosure, which requires smooth-pursuit eye movements.

The time distributions of gaze responses (figure 3(a)) confirm that humans react faster to instant changes in the scene (such as in the PU condition) than to transient ones (such as in the SA and MA conditions). In the MA condition, a delayed gaze response due to slow stimuli disclosure results in the presence of multiple unattended objects in the scene. The next shift of attention is challenged by this situation, causing more variability in gaze responses. The lack of difference in gaze response between the stimuli types in all three conditions suggests that foveal vision is needed to judge the stimuli.

In the MA condition, a difference in dwell time between the target and non-target stimuli was found for the majority of subjects, in contrast with the other two conditions. We cannot exclude silent counting of targets as a potential origin of the longer mean dwell time of target stimuli. Although the task was identical in all three conditions, the explanation could be a sufficient inter-stimuli time to complete the task in the PU and SA conditions. Thus, in these two conditions, the

observed dwell time might not reflect the difference in the processing time. At the same time, in the MA condition, as already mentioned, multiple as-yet-unattended stimuli could be simultaneously present in the screen competing for attention. The observed difference between conditions implies that the potential of exploiting ocular data for discrimination between the target and non-target content might be restrained by scene content (i.e., its dynamic, as well as sparsity regarding task-relevant content). Finally, the larger variance in dwell time for the MA condition may be explained by subjects' behavior, which ranges from either repositioning the gaze to a stimulus as soon as it is visible in the scene and overtly tracking it until enough evidence is accumulated for making a decision to repositioning the gaze when the stimulus is already mostly uncovered.

The trend of the average error rate of the subjects' responses with regard to the presentation protocols indicates the 'MA' protocol as the most challenging (see section 3.1). We explain this by the motion effect that makes the distance between the subsequent stimuli to change over time. The changes in relative positions between the stimuli can escalate the challenge to track and attend all stimuli in the sequence, likely causing the low percentage of sequences in which all the target stimuli were detected. Importantly, the percentage of the missed stimuli is still relatively low although higher than in the PU and SA presentation protocols (see section 3.1).

4.2. Discriminant EEG analysis

The single-trial classification performance significantly above the chance level (AUC = 0.5) supports the presence of the recognition-related EEG activity when fixating on the target stimuli in any of the conditions.

In the PU condition, the most discriminative activity corresponds to the centro-parietal positivity elicited by target stimuli around 450 ms of the fixation onset (figures 4(a) and 6(a)). This EFRP's component resembles the spatio-temporal signature of the P300 activity in the fixed-gaze visual oddball paradigm [31, 32]—the target-related positivity over the centro-parietal scalp area at ∼400 ms. This positivity is preceded by the discriminative negative EEG activity at ∼250 ms within the same scalp region. In the MA condition, a similar pattern (i.e., the centro-parietal positivity) is also seen, but shifted in time for ∼100 ms, peaking at ∼550 ms of the fixation onset. This centro-parietal positivity is visible in 'Smooth appearance' around ∼600 ms, although more spread out in time. The strong resemblance between the discriminant EEG activity across different conditions becomes more obvious when inspecting signr^2 measure (see figure 5). The peaks of signr^2 measure within a trial correspond to a centro-parietal positivity preceded by a centro-parietal negativity in all the conditions. The temporal location with respect to the fixation onset and their relative appearance time varies across conditions as discussed above.

Comparable classification performances over time are observed in the SA and MA protocols (figure 4). We found this reduced performance compared to the PU condition to be
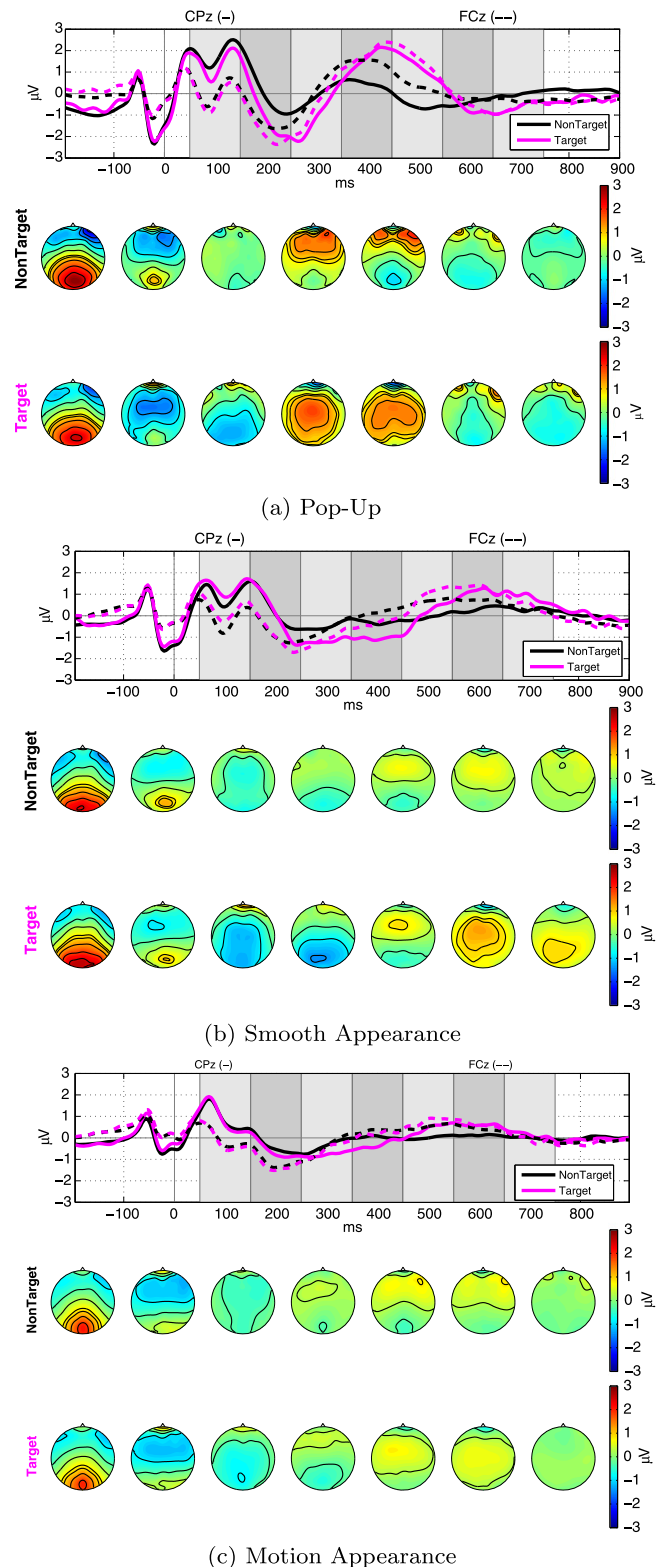

(a) Pop-Up

(b) Smooth Appearance

(c) Motion Appearance

Figure 6. Grand average eye-fixation related potentials.

indicative of temporal uncertainty of the recognition-related EEG components with reference to fixations onset. Interestingly, additional evidence supporting our hypothesis is found in the results of the proposed inter-protocol classification method (see section 2.4), where we transferred the knowledge about the recognition-related EEG activity from the 'PU' to
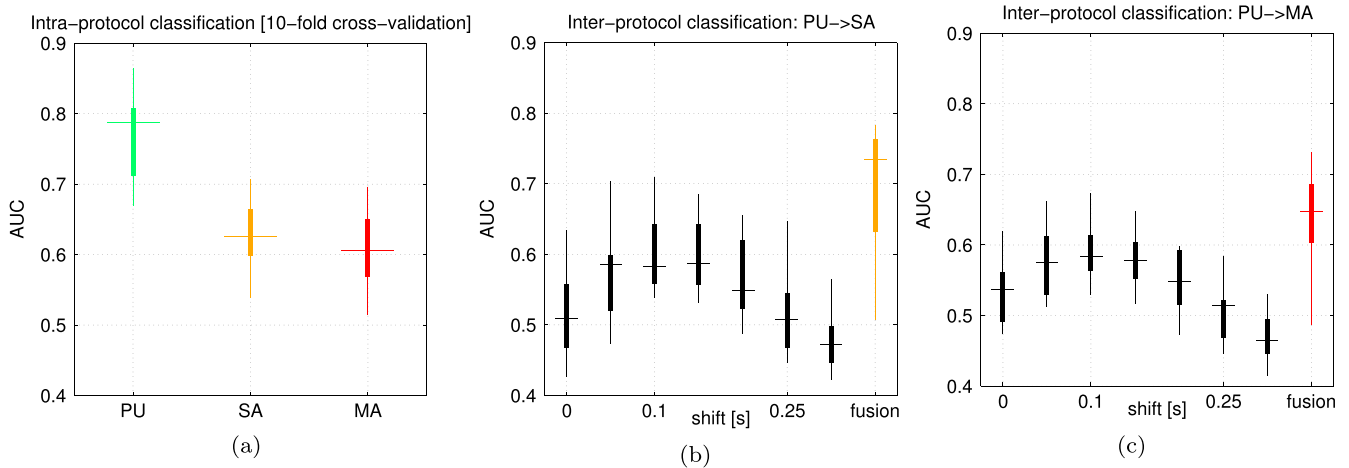
Figure 7. Classification performance (target versus non-target): (a) intra-protocol classification performance; (b) and (c) inter-protocol single-trial classification performance; the results for SA and MA protocols respectively. Note that zero shift value corresponds to 200 ms of fixation onset.
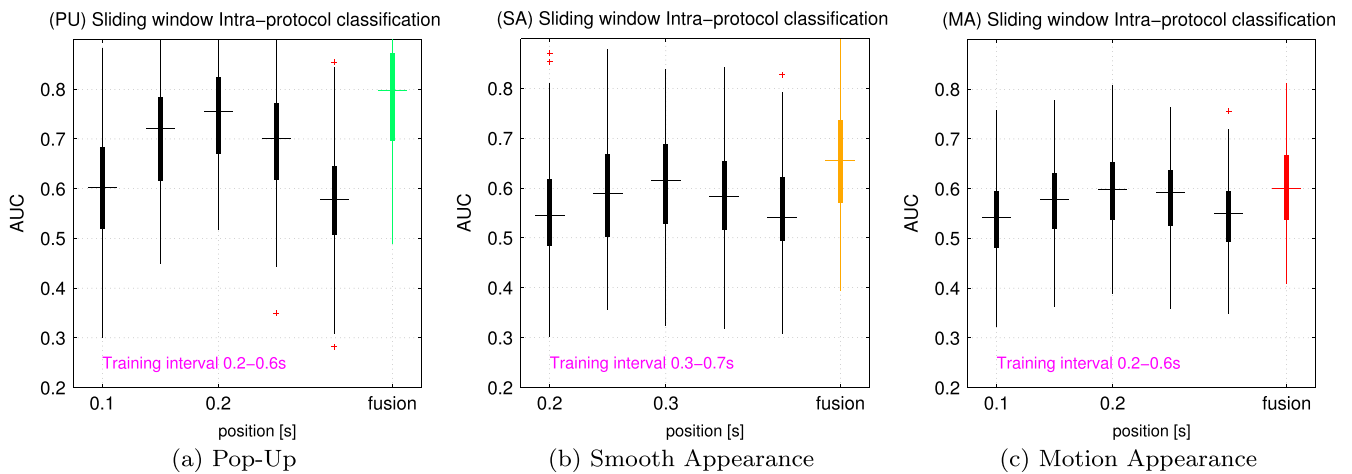

Figure 8. Intra-protocol classification with an HDCA, applied in a sliding manner on the testing dataset. The HDCA was trained on the training data on the fixed time interval as indicated in each plot. The intermediate results of the testing dataset, obtained for different positions of the sliding window, are presented with reference to the starting time of the window (i.e., time positions in the plot).

the other two presentation protocols. The proposed method improved the single-trial classification performance with respect to the case when the data are treated individually for each condition (figures 7(b) and (c)). Furthermore, the trends of the intermediate classification results in the proposed method, peaking at 100–150 ms, indicate the delays in the most discriminant activity compared to the PU condition.

In the intra-protocol setting, the classification using the sliding window improved performance only slightly (figure 8). This result shows that the performance improvement in the inter-protocol setting compared to the intra-protocol setting mainly originates in the transference of knowledge from PU to the other paradigms.

A relatively low classification gain in the inter-protocol approach for the MA condition suggests the presence of some additional phenomena besides stronger temporal variability. One perspective on decoding challenges in our MA condition comes from perceptual decision-making theory, which posits that the rate of stimuli disclosure may affect the accumulation of evidence by directly influencing the build-up rate of the
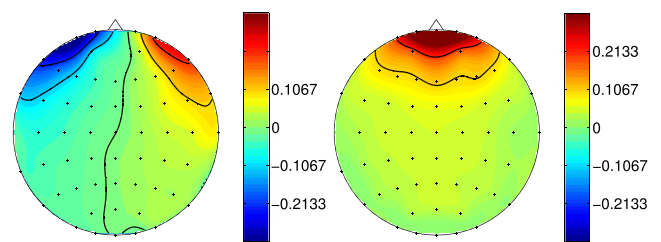

Figure 9. The estimated scalp projection (a.u.), averaged across subjects, of sources for horizontal (left) and vertical right eye-movements.

centro-parietal positivity corresponding to the recognition event [18–20]. Hence, the classifier trained on the ERPs evoked in the PU condition might not optimally explain the ones in the MA condition even if an exact timing would be known because of the slow disclosure. This concern requires further investigation i.e., a systematic study on the effects of stimuli disclosure rate on the ERP decoding. On the other side, our ocular behavior in free active search is driven by

scene content and dynamic. Thus, the presence of multiple yet-unattended stimuli in the MA condition might influence subjects' behavior by motivating faster judgment about the attended content in order to avoid missing new stimuli. As a result the decision criterion might be changed, which might influence the amplitude of the centro-parietal positivity [18]. Finally, free active viewing permits larger diversity in subjects' ocular behavior (see section 4.1). If subjects gaze at stimuli when their task-relevant features are disclosed, we would expect that the recognition task more closely resembles the PU condition. In contrast, earlier gaze repositioning to the novel stimulus may require longer tracking to accumulate the evidence for making the decision about the stimuli. Thus, the choice of the very first fixations might not be always optimal, considering that the analysis is done in a limited interval in relation to it.

Regarding the slow *motion effect* alone, a recent study suggests that overt tracking itself likely does not influence EEG decoding [33]. The authors demonstrated that if the exact time of the event is known, the overt tracking of objects moving at slow speed does not lessen the EEG decoding performance compared to the fixed-stimuli condition using the state-of-the-art methods.

5. Conclusion

We demonstrated that composite visual effects consisting of *fading-in*, *enlarging*, and *motion* may increase temporal uncertainty of the recognition-related EEG components with respect to the fixation onset in active viewing task. In the context of decoding users' intentions from the EEG, the temporal uncertainty introduces an extra challenge for the classical decoding approaches resulting in decreased classification performance. We showed, however, that the knowledge transferred from the paradigm characterized by less temporal uncertainty can be exploited to boost the EEG decoding performance in the more challenging conditions. We believe that these results point to the desirability of further research on applying knowledge gained from controlled experimental paradigms to real-world scenarios.

Finally, to the best of our knowledge, previous EEG studies on free-view search task have paid no attention on the dynamic of scenes and the presence of objects in motion. Our results show that enhanced scene's dynamic results in a larger variability in both behavioral and neural responses while performing an active visual search task.

Acknowledgments

References

[1] Nikolaev A R, Pannasch S, Ito J and Belopolsky A 2014 Eye movement-related brain activity during perceptual and cognitive processing *Front. Syst. Neurosci.* **8** 62

[2] Rämä P and Baccino T 2010 Eye-fixation related potentials (EFRPs) during object identification *Vis. Neurosci.* **27** 187–92

[3] Healy G and Smeaton A F 2011 Eye fixation related potentials in a target search task *Conf. Proc. IEEE Eng. Med. Biol. Soc.* pp 4203–6

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

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

[6] 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

[7] Krusienski D J, Sellers E W, McFarl D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21

[8] Treder M S and Blankertz B 2010 Covert attention and visual speller design in an ERP-based brain–computer interface *Behav. Brain Functions* **6** 28

[9] Gerson A D, Parra L C and Sajda P 2006 Cortically coupled computer vision for rapid image search *IEEE Trans. Neural Rehabil. Syst. Eng.* **14** 174–9

[10] Bigdely-Shamlo N, Vankov A, Ramirez R R and Makeig S 2008 Brain activity-based image classification from rapid serial visual presentation *IEEE Ttrans. Neural Syst. Rehabil. Eng.* **16** 432–41

[11] Kapoor A, Pradeep S and Desney T 2008 Combining brain computer interfaces with vision for object categorization *Comput. Vis. Pattern Recognit.* **1** 23–8

[12] Pohlmeyer E A, Jangraw D C, Wang J, Chang S F and Sajda P 2010 Combining computer and human vision into a BCI: Can the whole be greater than the sum of its parts? *Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society* pp 138–141

[13] 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 Open Access

[14] 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** e72018–08

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

[16] Parra L C, Spence C D, Gerson A D and Sajda P 2005 Recipes for the linear analysis of EEG *NeuroImage* **28** 326–41

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

[18] O'Connell R G, Dockree P M and Kelly S P 2012 A supramodal accumulation-to-bound signal that determines perceptual decisions in humans *Nat. Neurosci.* **15** 1729–35

[19] Kelly S P and O'Connell R G 2013 Internal and external influences on the rate of sensory evidence accumulation in the human brain *J. Neurosci.:Official J. Soc. Neurosci.* **33** 19434–41

[20] Twomey D M, Murphy P R, Kelly S P and O'Connell R G 2015 The classic p300 encodes a build-to-threshold decision variable *Eur. J. Neurosci.* **42** 1636–43

[21] 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

[22] 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

[23] Devillez H, Guyader N and Guerin-Dugue A 2015 An eye fixation related potentials analysis of the p300 potential for fixations onto a target object when exploring natural scenes *J. Vis.* **15** 20

[24] Rosenthal D, DeGuzman P, Parra L C and Sajda P 2014 Evoked neural responses to events in video *IEEE J. Sel. Top. Signal Process.* **8** 358–65

[25] Marathe A R, Ries A J and McDowell K 2013 A novel method for single-trial classification in the face of temporal variability *Foundations of Augmented Cognition—7th Int. Conf., AC 2013, Held as Part of HCI Int. 2013 Proc. (Las Vegas, NV, USA, 21–26 July 2013)* pp 345–352

[26] Marathe A R, Ries A J and McDowell K 2014 Sliding HDCA: single-trial EEG classification to overcome and quantify temporal variability *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 201–11

[27] Kutas M, McCarthy G and Donchin E 1977 Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time *Science* **197** 792–5

[28] Kutas M and Donchin E 1978 Variations in the latency of P300 as a function of variations in semantic categorizations *Multidisciplinary Perspectives in Event-Related Brain Potential Research* (Chapel Hill, NC: The Office)

[29] Parra L C, Christoforou C, Gerson A D, Dyrholm M, An Luo, Wagner M, Philiastides M G and Sajda P 2008 Spatiotemporal linear decoding of brain state *IEEE Signal Process. Mag.* **25** 107–15

[30] Marathe A R, Ries A J and McDowell K 2013 A novel method for single-trial classification in the face of temporal variability *Foundations of Augmented Cognition* (*Lecture Notes in Computer Science* vol 8027) ed D D Schmorrow and C M Fidopiastis (Berlin: Springer) pp 345–352

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

[32] 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**

[33] Ušćumlić M, Hägele M and Blankertz B 2015 EEG correlates of visual recognition while overtly tracking a moving object *Proc. of the 4th Int. Workshop, Symbiotic 2015 (Berlin, 7–8 October 2015)* pp 166–71