Evoked Neural Responses to Events in Video

Daniel Rosenthal, Paul DeGuzman, Lucas C. Parra, *Senior Member, IEEE*, and Paul Sajda, *Fellow, IEEE*

***Abstract*—In contrast to static imagery, detection of events of interest in video involves evidence accumulation across space and time; the observer is required to integrate features from both motion and form to decide whether a behavior constituents a target event. Do such events that extend in time elicit evoked responses of similar strength as evoked responses associated with instantaneous events such as the presentation of a static target image? Using a set of simulated scenarios, with avatars/actors having different behaviors, we identified evoked neural activity discriminative of target vs. distractor events (behaviors) at discrimination levels that are comparable to static imagery. EEG discriminative activity was largely in the time-locked evoked response and not in oscillatory activity, with the exception of very low EEG frequency bands such as delta and theta, which simply represent bands dominating the event related potential (ERP). The discriminative evoked response activity we see is observed in all target/distractor conditions and is robust across different recordings from the same subjects. The results suggest that we have identified a robust neural correlate of target detection in video, at least in terms of the stimulus set we used—i.e., dynamic behavior of an individual in a low clutter environment. Additional work is needed to test a larger variety of behaviors and more diverse environments.**

***Index Terms*—Video analysis, electroencephalography, brain-computer interfaces, audio-visual systems, machine learning, video surveillance.**

I. Introduction

VIDEO surveillance typically involve multi-camera systems generating streams of data that are monitored by human operators over lengthy shifts. Operator fatigue and lapses of attention can significantly reduce monitoring performance. Additionally, while many operators are well-trained and can be responsive to subtle cues, the sheer throughput of information, whether in the presence of single or multiple video streams, makes it important for each operator to rapidly prioritize his/her attention—e.g., when facing multiple dynamic events and determining whether an observation warrants an alert trigger and/or further review.

Our aim was to investigate whether non-invasive neuroimaging can be used to detect neural correlates of rapid detection and recognition of events of importance in a video stream. Previous work by our group and others has shown that this type of "neural-tagging" is possible for static images when presented via rapid serial visual presentation (RSVP) [1]–[4]. In the static image case, images containing objects of interest evoke responses in the operator's electroencephalogram (EEG), and these evoked responses could be decoded and used to construct a "neural-tag" for prioritizing the data stream. In addition, neural tagging was seen to be more consistent than a behavioral response, such as a button press, particularly in cases where exploitation time was long [5].

We previously demonstrated that neural signals associated with detection of static images could be used to accelerate broad-area search in geospatial imagery [3]. To do so, we developed an EEG-based brain computer interface (BCI) which coupled the neural evoked responses and computer vision. Using the RSVP paradigm, images were presented to viewers at high frame rates (5–10 frames per second) and subsequently ranked based on a classification score.

In the case of video, one may be interested in not just objects, but events defined by how objects move over time. For example, when monitoring activity at a busy train station, one may be interested in events where a person wanders in the station and then drops a bag and leaves. The event is defined not simply by the presence of a person, but also their dynamic behavior as captured by their motion in the video.

There has been relatively little research done using EEG to investigate neural signatures of target and/or anomaly detection in either real-world or simulated video imagery (for exceptions see [6]–[8]). A far greater amount of work has been conducted using functional magnetic resonance imaging (fMRI) imaging [9], [10]. However fMRI is an expensive and non-portable imaging modality, which is likely to have minimum utility in practical and operational environments. In addition, the activity measured by fMRI is a sluggish (i.e., low pass) indirect representation (i.e., blood flow/volume) of neural activity in the brain and, as a modality on its own, is less well matched for studying the constituent processes of rapid decision-making.

We specifically focused on an initial investigation of whether neural correlates of these complex dynamic events of interest can be measured and decoded from EEG. Compared to static imagery, video involves evidence accumulation across space and time, thus making signatures of targets and anomalies more complex while also providing additional sources of sensory context in the input stream. In addition, we aimed to build a plat-

Manuscript received September 13, 2013; revised January 22, 2014; accepted March 07, 2014. Date of publication March 21, 2014; date of current version May 13, 2014. This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Hari Kalva.

D. Rosenthal and P. DeGuzman are with Neuromatters LLC, New York, NY 10005 USA (e-mail: drosenthal@neuromatters.com; pdeguzman@neuromatters.com).

L. Parra is with the Department of Biomedical Engineering, City College of New York, New York, NY 10031 USA (e-mail: parra@ccny.cuny.edu).

P. Sajda is with Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (e-mail: ps692@columbia.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2014.2313022

Fig. 1. Three frames taken from stimuli from three selected video events. In "Leave Without Bag" videos, actor enters the frame from one of four directions (inside, left, right, front) and places the bag in center of the scene. A target event is defined by the actor then leaving the bag behind and exiting the frame in one of the four directions. In "Leave With Bag" videos, the actor performs the same actions from the prior case, but picks the bag up again before leaving the scene—i.e. this is a distractor event. For all non-bag target events (e.g. "Look At Watch" shown above), the actor enters without a bag, performs a given action, and then leaves the scene.

form for conducting future experiments in which video feeds could be tagged with neural labels, and where such labels could be ingested by a machine vision system to provide a synergistic brain-machine interface for video analysis. It is important to note that the scope of this project was not to perform an exhaustive evaluation of the types of actions or spatio-temporal events that may be reliably detected from neural signatures, but rather, to provide an initial feasibility assessment on a few examples.

An important aspect of our previous work with static imagery is the rapid presentation of images which accomplishes two important goals: 1) it significantly accelerates the inspection of the data without significant loss in detection performance, and 2) more clearly defines the moment of visual detection and thus strengthens the signal-to-noise ratio for the neural signals detection task. Similarly, we centered our work around the accelerated presentation of stimuli to the viewer. Accelerated video playback shortens the time required to review video footage (already an accepted approach in video surveillance). In addition, we hypothesized that neural signals are more salient as the moment of subjective event detection is better delineated in time.

II. Methods, Assumptions and Procedures

A. Stimulus Set Design

Experimental stimuli were created using the Object Video Virtual Video Tool (OVVVT) [11], a software package that places virtual actors in realistic simulated environments. The OVVVT was chosen not only because it offered surveillance-related assets and maps, but also because it gave experimentalists strict control over video content and actor behavior. However, our approach and platform (see below) are agnostic to the simulation environment chosen.

In the course of the study, several iterations of stimuli were developed. Specifically, early stimuli included multiple actors in a wide context to enable the simultaneous analysis of saccadic activity and neural activity. However, there was a need and opportunity to constrain the experimental paradigm and elucidate neural components in an incremental fashion. Thus, the stimulus set evolved to focus on a narrower field of view requiring minimal gaze changes, reducing eye-movements as a confounding factor.

The finalized stimulus set focused on a single actor that would enter and leave the scene from any of several predefined directions in the context of a train station entrance (Fig. 1). In addition to leaving or entering the scene, the actor could perform one of several activities which occurred in the center of the scene. The activity occurred in the center of the scene. As a result, gaze was relatively constrained and ocular activity minimized to facilitate subsequent data analysis, though we still recorded electrooculogram (EOG) for additional removal of eye movement related confounds. The actor could be any of 7 predefined characters, each of which differed in physical attributes. Two characters were similar in their appearance while others varied more widely. Videos were annotated manually using the ANVIL Video Annotation Research tool [12].

An important aspect of the presentation was the increased presentation speed, with the dual goal of increasing viewing efficiency and more clearly defining the relevant event time. Defining events are relatively brief in this fast playback $(<1\ s)$, but they are flanked by extended periods of actors waiting, or entering and exiting the scene. Importantly, without these flanking periods, the short event may not be detectable by the viewer. Thus, while the event is brief, the defining characteristic is still extended in time and required for reliable detection. Nevertheless, the brief events are expected to elicit time-locked evoked activity in the EEG.

B. Experimental Procedures and Tasks

Two data collections were performed with the same subjects ($\mathrm{N} = 8$) for each collection. This enabled data analysis to be performed across two time points; however one subject was unable to return for the second experimental session. For the purpose of cross-session analysis, this subject's data was removed. During each data collection, and for each participant, the EOG was recorded to measure eye movements that might produce artifacts in the data analysis. For the EOG, sensors were placed below the eyes and near the temples to record horizontal and vertical eye movements as well as blinks.

The first experimental session used two identical stimuli sets with the same target activity but involved two different reporting tasks. Each participant viewed 25 blocks of randomly selected clips, where each block contained exactly ten 30 second clips and the prevalence of target clips was a random 3 to 5 per block. During breaks between blocks of videos, subjects gave a response that depended on the task they were performing (Table I) and then immediately continued with the next block.

The stimulus set consisted of "Leave Without Bag" and "Take Bag" clips, and participants were instructed to internally identify—no explicit behavior responses such as a button press or eye movement—the moment in time when the actor left the scene without the bag (i.e., leaving the bag behind was the target event). In the first session of the data collection, subjects were instructed to count the number of target events (Task A). In the second part of the data collection, the video clips could involve any of 6 actors. At the end of each block, the full bodies of 4 characters were presented for a brief time interval, interspaced with a mask consisting of a grey screen. One of these actors had not performed in a target video during the entire block. Participants were asked to report which of the 4 presented was this character—the subject had to identify the "innocent" actor (Task D). This latter task was more complex in that it required a subject to hold the actors in memory so as to determine which actor did not perform a bag-drop.

The second experimental session involved three stimulus sets. The first set was the bag drop vs. bag pick up set used in the first data collection session (Task B). This set served to analyze intra-subject reproducibility across two time points. The second stimulus set consisted of the first set, expanded with additional distractor clips of a single actor walking through the scene with or without the bag (Task E). The final set consisted of entirely different video clips, each still 30 s long, involving an actor performing one of five tasks: walking (without a bag), stretching, waving, looking at his watch, tying his shoe laces. The purpose of the second and third set was to analyze the extent to which neural components observed in the first data collection remained stable in the presence of additional distractor events, and with different activities as targets (Task C).

TABLE I
TARGET AND DISTRACTOR EVENTS USED IN ANALYSES

Task	Target Event	Distractor Event	Rationale
A	Leave Bag	Take Bag	Base Comparison (Session 1)
B	Leave Bag	Take Bag	Repeatability (Session 2)
C	Look at Watch *or* Wave	Stretch, Tie Shoe and either Look at Watch *or* Wave	Generalization Across Different Target Events (Session 2)
D	Leave Bag	Take Bag	Generalization Across Different Task Conditions (Session 1)
E	Leave Bag	Take Bag + Additional Distractors	Generalization Across Wider Variety of Distractor Conditions (Session 2)

For all sets in session two, participants were asked to count the number of bag drop events. For Task C, four participants were asked to pay attention to events where the actor looks at his watch, and the remaining subjects were asked to pay attention to events where the actor waves. Waving and looking at one's watch involve similar movements in early parts of the activity, requiring the participant to maintain attention. As in the first data collection, all clips were presented at an accelerated pace of five times the original video capture speed.

A summary of the different tasks given to the subjects is given in Table I. In the simplest case (Task A) subjects were instructed to count the number of occurrences of a bag drop against a background of a majority of bag pick-ups. The counting intends to assure that subjects remain attentive for the duration of the trial. This experiment was repeated in a second recording (Task B) to test for within-subject reproducibility, i.e., how robust are the features and results. Note that the activities performed by the actors differ, by definition, in their low-level movement characteristics (the different ways in which the actor moves when the different actions are performed). Previous work with EEG evoked responses suggested that the low-level movement features were not expected to be discriminative. Nevertheless, such low-level differences were specifically controlled for by using a more diverse set of targets and distractors in Task C and E, respectively. Finally, one may argue that counting is not a particularly meaningful task in the context of surveillance. In Task D subjects were instructed to remember which actor did not leave a bag behind. Thus, a more operationally meaningful assignment would be tested for its effects on the results. Similar to counting, this task requires one to maintain a memory of the past events, but in this case, the memory content is arguably more meaningful.

Note that all task definitions allow verification that the subjects detected all relevant events i.e., reported the correct number of target events (Task A, B, C, E) or reported the correct actor (Task D). Initially various presentations speeds were explored and eventually converged towards 5 times real-time, which still permitted nearly perfect subject performance. Near perfect, but not 100% perfect performance is also important to maintain constant attention, as subjects are prone to lapses in

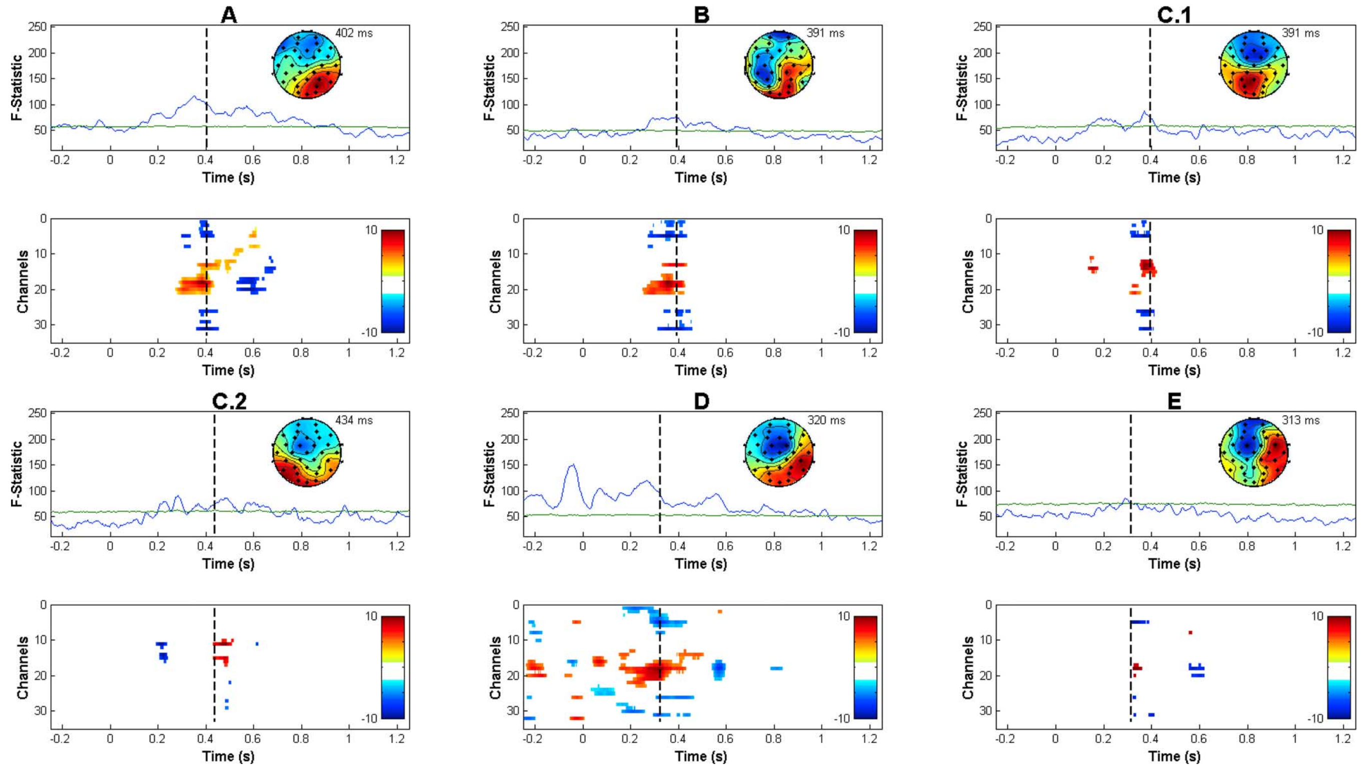


Fig. 2. Group results ($N = 8$) across different target vs. distractor events (panels are labeled by the event definition defined in Table II). Each panel represents an analysis from 250 ms pre-stimulus to 1250 *ms* post-stimulus (stimulus onset at time 0 *ms*). The first and third rows represent the F-statistic computed across all channels and subjects. The green line represents statistical significance at $p = 0.01$, estimated via shuffle statistics. Inset represents average scalp difference activity at times indicated by dotted vertical lines. Second and forth rows represents z-scores as a function of time plotted for each channel. All z-scores representing $p < 0.01$ (two-tail) are shown in color. Figures are labeled to correspond to conditions described in Table I.

attention if the task is too easy or too hard. Other than that, the approximately even rate of occurrence of targets assured that targets and distractors receive an equal amount of attention. Importantly, this was achieved without requiring a button response every time an event occurred. A button press is known to elicit a strong evoked response in the EEG [13] and thus would severely confound the detection performance based on cognitive processes alone, which was the primary goal of this research effort.

C. Data Analysis Methods

The EEG was recorded using a BioSemi ActiveTwo system. Subjects wore a standard 32-electrode cap configured in the international 10/20 system. In analysis, EEG was downsampled from 2048 Hz to 256 Hz, high-pass filtered at 0.5 Hz to remove baseline drift and notch filtered at 60 Hz and 120 Hz to remove line noise. Electrooculogram (EOG) was recorded with six auxiliary electrodes and eye-movement artifacts were linearly regressed out of the processed EEG. Epochs from each video were extracted from each video clip, time-locked to the specific salient events (Table I).

Both event-related responses as well as oscillatory features in the EEG that would be discriminative of a target event from a non-target event condition were investigated. Several definitions of the target and non-target events were considered, all of which were tested with the same data analysis methodology. To explore single-trial ERP differences for each case, hierarchical discriminant component analysis (HDCA) was used. To briefly summarize previous work, classifiers are trained by combining spatial and temporal weights to maximize differences between the putative current sources detected in EEG during targets and non-target epochs. Firstly, spatial weights are computed as to maximally separate a weighted average of electrode potentials between the two conditions. Once spatial coefficients are found, optimal temporal weight vectors for temporal windows of 100 *ms* can be computed. In other words, HDCA estimates a series of linear spatial filters that are discriminating at specific times in the epoch relative to the stimulus and linearly integrate these across time within an epoch. These epochs used only post-stimulus activity, with epochs classified between 100 *ms* and 1100 *ms*. This methodology is well established, particularly with respect to its utility in identifying discriminative components in the EEG [3], [14], [15].

III. Results and Discussion

A. Analysis of Trial-Averaged ERPs

Differences in event related potentials (ERPs) between target events and distractor events were analyzed. Fig. 2 shows these results averaged across the group of eight subjects. A clear differential activity was found in the base condition (Task A in Table I) which peaks at 300 *ms* post-stimulus and is right lateralized in occipito-parietal electrodes. This differential activity is robust across recording sessions (compare Fig. 2(A) and (B)). Interestingly, this activity is very similar in timing and scalp topology seen in a static image case, though spreading slightly

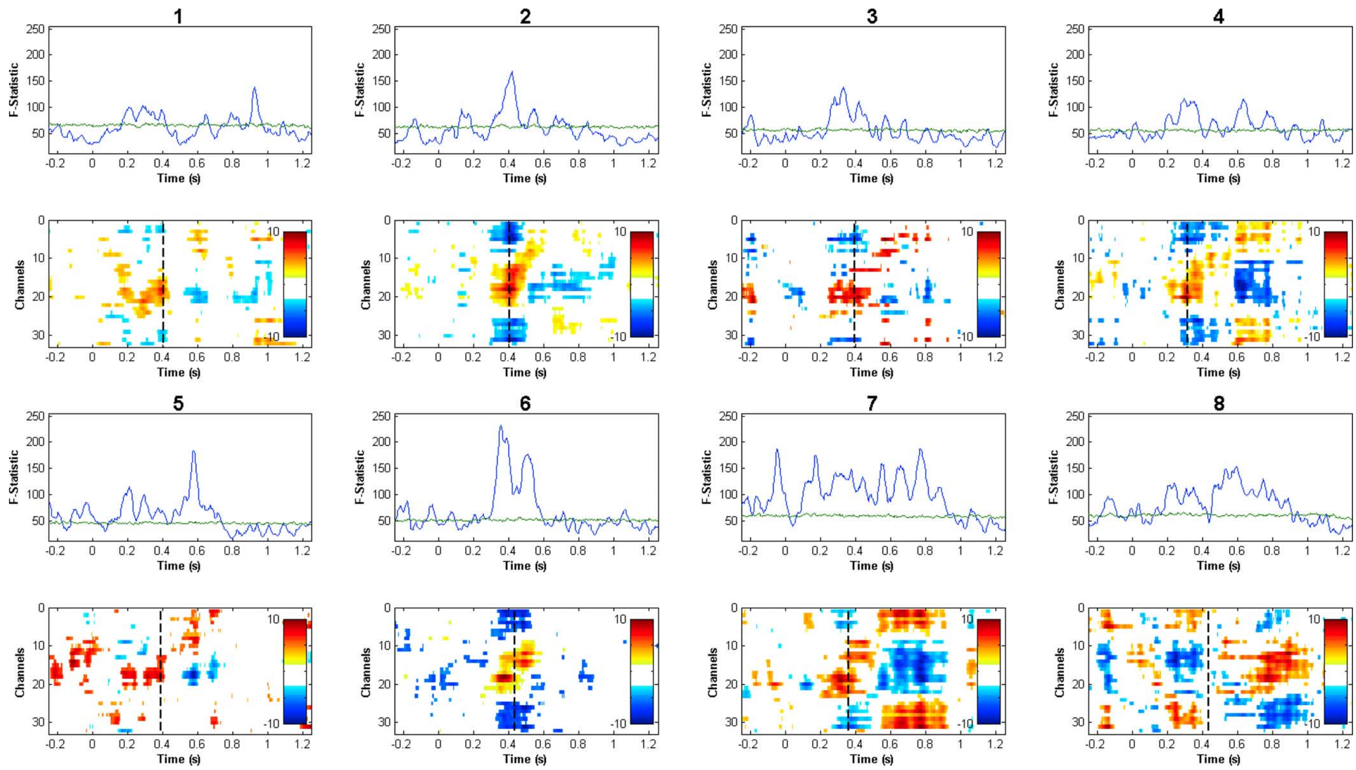


Fig. 3. Individual subject results ($\mathrm{N} = 8$) across target vs. distractor events for Task A (labeled provided in Table I). This figure follows directly from Fig. 3, where the first and third throws represent the F-statistic and second and fourth show corresponding z-scores for each channel.

more dorsally (compare to figures shown in [16]). Since the processing of visual motion would typically result in differential activity in more dorsal areas of the occipital lobe, this finding might reflect an integration of both form and motion information in detecting the target behavior. When the target behavior is changed from a bag drop case to another behavior, such as looking at one's watch or waving (Fig. 2(C.1) and (C.2), respectively), the lateralization was less pronounced, though differential activity is still clearly occipito-parietal. When the task changes, and instead of simply counting the number of bags left behind the subject must report which actor did not leave a bag behind (Task D), the same right lateralized occipito-parietal activity is seen, also at approximately 300 *ms* post-stimulus. Intriguingly, there is pre-stimulus activity that is discriminative. However, this result is completely consistent with the task, since the subjects did not need to count bags left behind, instead they only needed to pay attention and remember whether an actor ever left a bag behind. Thus as the experiment progressed in time there was potentially discriminative information as soon as the actor entered the screen. This is reflected in pre-stimulus differential activity (Fig. 2(D)). Finally, in terms of an ERP analysis, the weakest differential activity was observed when additional distractors were added (Fig. 2(E)). It is interesting, however, that when all sensors were integrated in the single-trial analysis, single-trial discrimination accuracy for the added distractor case was not significantly different from the base condition (see Fig. 4). Note also that Task E cannot be solved based on simple low-level movement features, ruling out the possibility that the detected neural signals reflect simple stimulus differences. In summary, we conclude from these trial and group averaged results that there is potentially information in the EEG that could be decoded single-trial for labeling the video.

Fig. 2 shows a summary of the activity for all tasks by averaging across subjects for each task—this is the conventional "grand average" taken in EEG in particular if one is interested in the spatial distribution of evoked responses, which are highly variable across individuals due to variations in skull anatomy (see e.g., [17]). To provide an appreciation of the robustness/variability of this activity across subjects we display in Fig. 3 the same data for individual subjects for one of the tasks (Task A). Considering typical across-subject variability these spatio-temporal patterns of evoked responses are remarkably similar across subjects.

B. *Analysis of Single-Trial ERP Differences Using HDCA*

On the whole, the single-trial classification accuracy was substantially above chance (Fig. 4). While average classification performance decreased between sessions for the first experiment, it did not do so significantly (A vs. B). This suggests that the evoked responses classified are repeatable, given that two experimental sessions were conducted several weeks apart. The "Who Is Innocent" task had the strongest classification of all groups. It is most likely the case that the task was more engaging to viewers, leading to more reliable evoked responses across the subject group. Furthermore, this performance was independent of the pre-stimulus activity discussed earlier (A vs. D). The inclusion of additional distractors from the task performed in A and B did not significantly impact the results of the classification, though it did result in a small improvement in average AUC (B vs. E). Finally, the non-bag events did *not* yield as high clas-

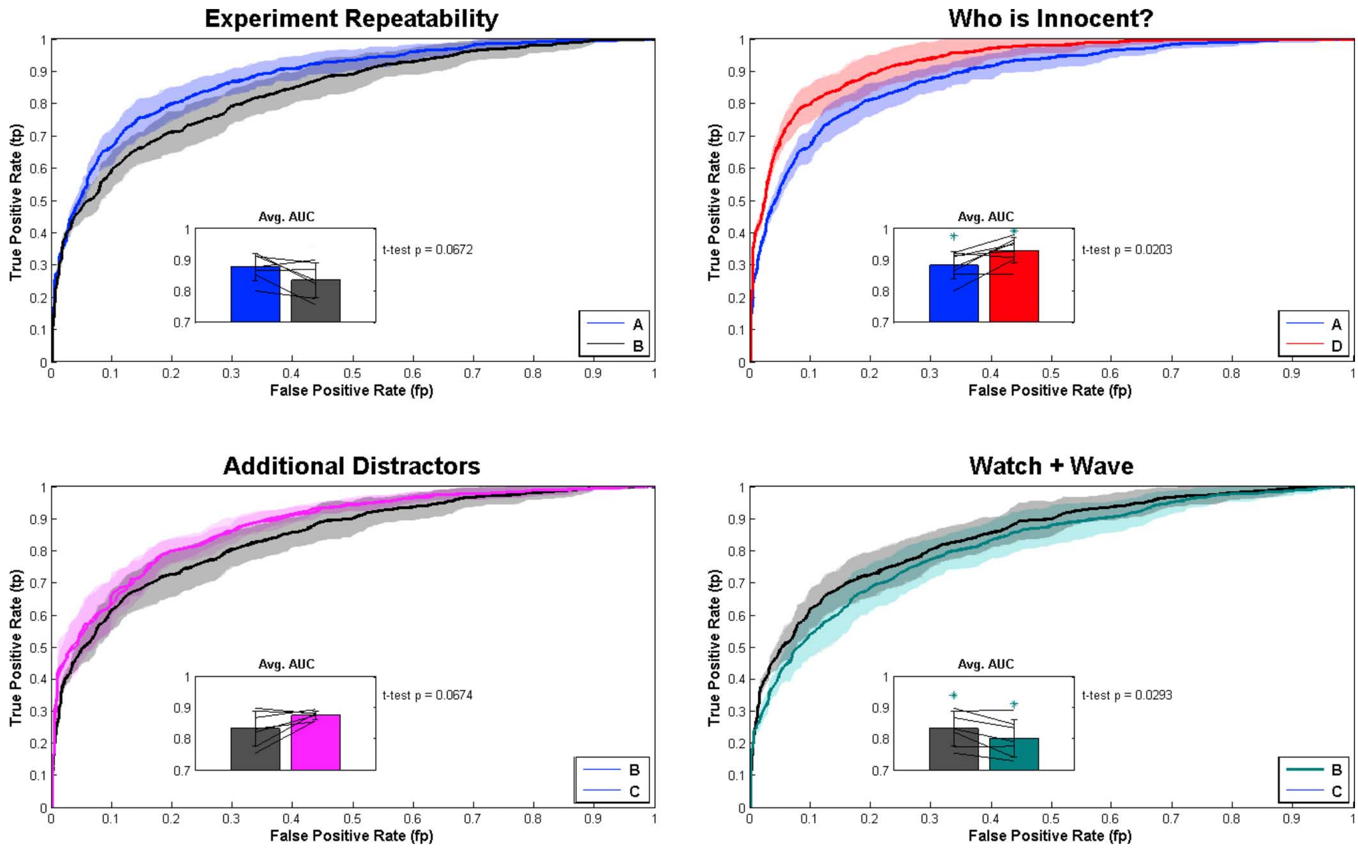


Fig. 4. Receiver Operating Characteristic "ROC" curves showing results of single-trial analysis of evoked responses. All single-trial analyses were done using HDCA. Inset legends indicate the conditions tested as defined in Table I. (Experimental Repeatability): Shown are ROC curves for two different sessions of the same eight subjects (sessions are several weeks apart). In both cases the AUC was >0.80 (i.e., probability of a correct single-trial decoding of the target event was $>80\%$). Though on average the second session had a lower average AUC than the first session, this difference was not significant at $p < 0.05$ (inset shows p-value for paired t-test across two sessions). (Who is innocent): Results show that this task results in better single-trial discrimination relative to the counting tasks (conditions A vs. D) with the "who is innocent" having an average $\mathrm{AUC} > 0.90$. (Additional Distractors): We found no significant difference when additional distractors were added. This is despite that fact that the trial averaged ERP analysis indicated less differential activity between target and distractor events when augmenting with additional distractors. $(\mathrm{Watch} + \mathrm{Wave})$: Discrimination of target events when they were defined by the actor looking at their watch or waving resulted in a statistically significant reduction in accuracy, though this accuracy was still substantially above chance.

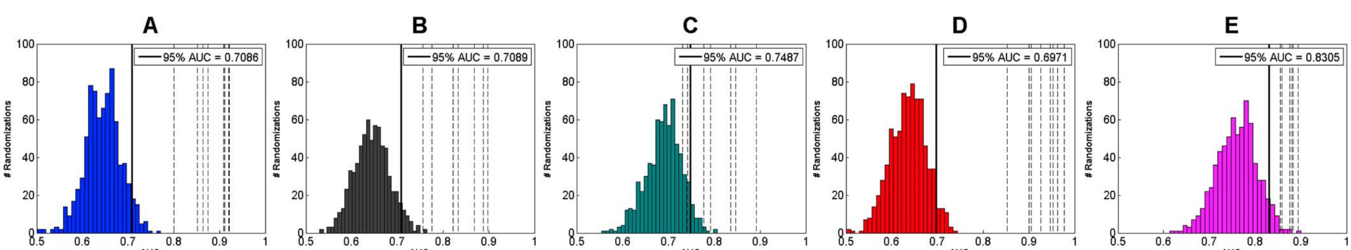


Fig. 5. Histograms showing distribution AUC for HDCA classifiers generated with randomized labeling $(\mathrm{N} = 1000)$. Solid line represents significant AUC (95th percentile) for each given experiment. Dotted lines indicate individual subject classification performance. With the exception of two subjects in condition C, all subjects achieved discrimination above chance.

sification accuracy. This may be a result of greater variability as to where the events were locked in time. This may also be a task-related effect, similarly to the improvements seen in task B.

To assure that the detection performance reported here is statistically significant, and to investigate significance for individual subjects, we performed a bootstrap analysis (Fig. 5). Labels were randomized, classifiers trained, and AUC performance computed using leave-one-out cross-validation in an identical fashion to the original data. This was repeated $(\mathrm{N} = 1000)$ to provide a distribution of AUC values under the null-hypothesis (no difference in the data). The histograms of these shuffled AUCs indicate that in all instances performance is above the 95% confidence level (except for 2 subjects in Task C). Incidentally, note that all the analysis performed with the HDCA was done on data from individual subjects, i.e., data for classification was not combined across subjects.

C. *Analysis of Trial-Averaged Spectral Power Differences*

A frequency domain analysis was performed for the different conditions outlined in Table I. In general, we find little consistency in the differential activity computed via differences in oscillatory power for target vs. distractor events (Fig. 6). The possible exception is in the low frequency bands (1–3 Hz delta and 4–7 Hz theta). Cortical theta has been implicated in perceptual

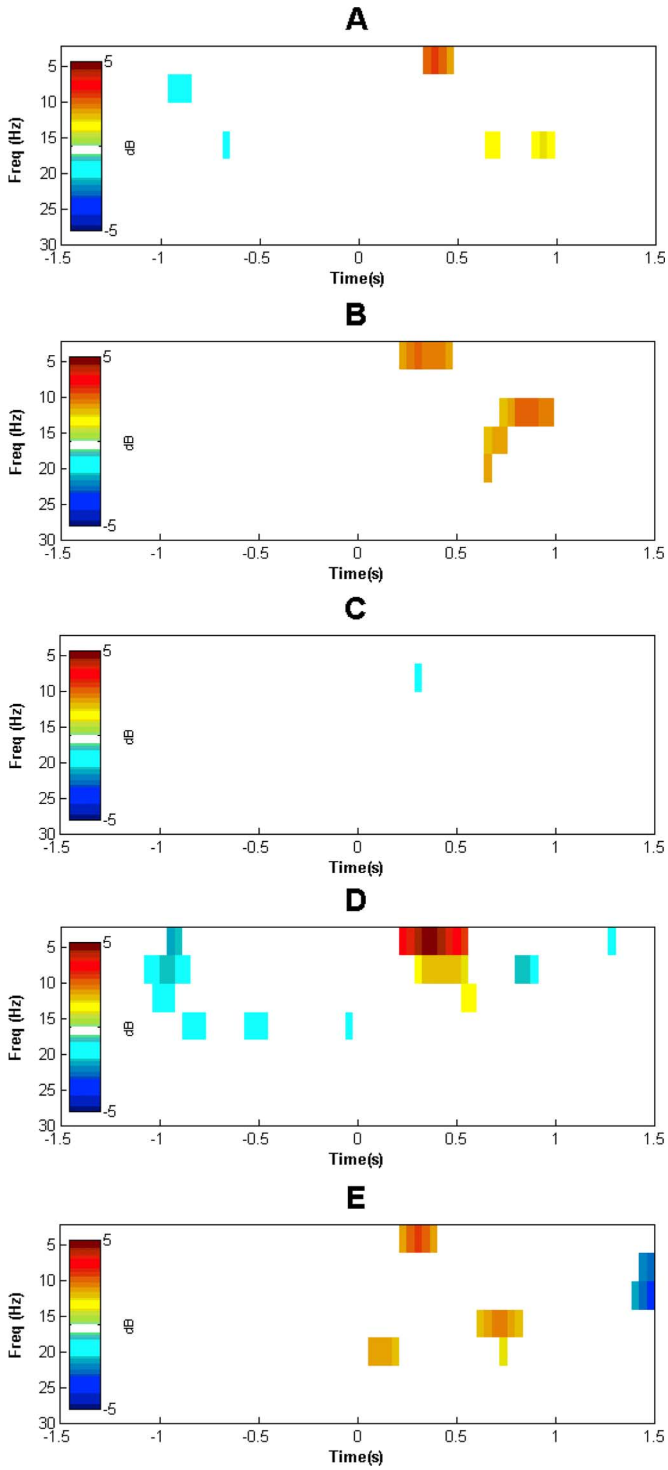


Fig. 6. Group analysis ($N = 8$) showing significant ($p < 0.01$) spectral power differences (ERSP) between target and non-target events. Panels are labeled with conditions as defined in Table I.

decision making as well as memory and retrieval [18] and it is possible that these differences reflect this type of activity. Particularly in Task D, where the subject must remember actors who left bags behind, we see differential activity in the theta band at 300 *ms* to 500 *ms* post-stimulus, which may reflect this element of the task—i.e., memory. Although there were spectral power differences in higher bands (alpha, beta, etc.), they were not reproducible, even across tasks A and B, which consisted of the same stimulus set. By and large the differential activity in the oscillatory power is less robust than in the ERPs.

Additionally, Common Spatial Pattern (CSP) [5] analysis was performed on the data. However, no stable components were found across any of the frequency bands that could reliably differentiate target from distractor events.

IV. Conclusion

We found that neural activity discriminative of target vs. distractor events defined by actor behavior in simulated video sequences. Discriminative activity is largely in the time-locked evoked response and not in oscillatory activity, with the exception of very low EEG frequency bands such as delta and theta, which likely simply represents those bands which dominate the ERP or are simply artifacts. The discriminative evoked response activity we see is observed in all target/distractor conditions and is robust across different recordings from the same subjects. This suggests we have identified a robust neural correlate of target detection in video, at least in terms of the stimulus set we used—i.e., dynamic behavior of an individual in a low clutter environment. It is important to note that the target is defined not just by static features, as in previous work [1]–[3], but an integration of static and dynamic features which implies an integration of form and motion in accumulating evidence to "decide" if/when a target behavior is present.

Though the discriminative activity in the EEG is significant, it is less clear what the neurological source of this evoked response is. One explanation, at least given the timing of the component activity, is that we are observing a response similar to the P300 [19], often typical in these types of target vs. distractor scenarios. More recent work has shown that the P300 may in fact represent a more general process of evidence accumulation in decision-making [20]. However the scalp topology is not typical of the P300 and instead is more consistent with activity that one would see in a face discrimination task, that involves activation of the fusiform face area, where discriminative activity is usually observed in ERPs over electrode PO8 (right lateralized parietal-occipital activity) [16], [21]. Interestingly, the activity is also slightly more dorsal and includes motion selective areas such as MT/V5. One hypothesis is that the activity we see represents the integration of motion (MT/V5) and form (person/face) during the evidence accumulation process. Further work, using neuroimaging modalities with a better combined spatial- temporal resolution, such as simultaneous EEG/fMRI, would be needed to test this hypothesis.

References

[1] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans Neural Syst Rehabil Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

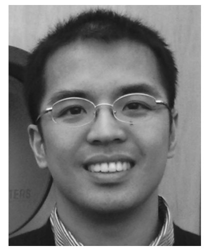
[2] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, "Brain state decoding for rapid image retrieval," in *Proc. ACM MultiMedia*, Beijing, China, 2009.

[3] P. Sajda, E. Pohlmeyer, J. , W. , L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and C. S.-F, "In a blink of an eye and a switch of a transistor: Cortically coupled computer vision," *Proc. IEEE*, vol. 98, no. 3, pp. 462–478, Mar. 2010.

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

[5] P. Sajda, M. G. Philiastides, and L. C. Parra, "Single-trial analysis of neuroimaging data: Inferring neural networks underlying perceptual decision-making in the human brain," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 97–109, 2009.
[6] T. Sitnikova, G. Kuperberg, and P. J. Holcomb, "Semantic integration in videos of real-world events: an electrophysiological investigation," *Psychophysiology*, vol. 40, pp. 160–164, 2003.
[7] A. Luo and P. Sajda, "Comparing neural correlates of visual target detection in serial visual presentations having different temporal correlations," *Front. Hum. Neurosci.*, vol. 3, p. 5, 2009.
[8] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated components of ongoing EEG point to emotionally laden attention—A possible marker of engagement?," *Front. Hum. Neurosci.*, vol. 6, p. 112, 2012.
[9] U. Hasson, O. Furman, D. Clark, Y. Dudai, and L. Davachi, "Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding," *Neuron*, vol. 57, pp. 452–462, 2008.
[10] S. Nishimoto, A. T. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. L. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Curr. Biol.*, vol. 21, pp. 1641–1646, 2011.
[11] G. R. Taylor, A. J. Chosak, and P. C. Brewer, "OVVV: Using Virtual Worlds to Design and Evaluate Surveillance Systems," presented at the IEEE Conf. Comput. Vis. Pattern Recogn., 2007 [Online]. Available: http://nguyendangbinh.org/Proceedings/CVPR/2007/data/papers/workshops/VisualSurveillance/poksters.htm
[12] M. Kipp, "Spatiotemporal coding in ANVIL," presented at the 6th Int. Conf. Lang. Resources Eval., 2008 [Online]. Available: http://www.lrecconf. org/proceedings/lrec2008
[13] L. Eimer, "The lateralized readiness potential as an on-line measure of central response activation process," *Behav. Res. Meth., Instrum., Comput.*, vol. 30, no. 1, pp. 146–156, 1998.
[14] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *Neuroimage*, vol. 28, pp. 326–341, 2005.
[15] A. D. Gerson, L. C. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.
[16] M. G. Philiastides and P. Sajda, "Temporal characterization of the neural correlates of perceptual decision making in the human brain," *Cereb. Cortex*, vol. 16, pp. 509–518, 2006.
[17] J. Rice, C. Rorden, J. S. Little, and L. C. Parra, "Subject position affects EEG magnitudes," *NeuroImage*, vol. 64, pp. 476–484, 2013.
[18] J. Jacobs, G. Hwang, T. Curran, and M. J. Kahana, "EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision-making," *NeuroImage*, vol. 32, pp. 978–987, 2006.
[19] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clinical Neurophysiol.*, vol. 118, pp. 2128–2148, 2007.
[20] R. G. O'Connell, P. M. Dockree, and S. P. Kelly, "A supramodal accumulation-to-bound signal that determines perceptual decisions in humans," *Nat. Neurosci.*, vol. 15, pp. 1729–1735, 2012.
[21] S. Bentin *et al.*, "Electrophysiological studies of face perception in humans," *J. Cognitive Neurosci.*, vol. 8, no. 6, pp. 551–565, 1996.

Daniel Rosenthal received the B.S.E. degree in bioengineering from the University of Pennsylvania (UPenn). Dan researched seizure localization techniques and helped develop real time processing software for a new generation of flexible subdural electrodes. Since graduating in 2011, he has been with the Neuromatters team assisting in the design of BCI software and performing data and algorithm analysis.



Paul Deguzman received the B.S. degree in biomedical engineering from Boston University and the M.S. degree in biomedical engineering from the City College of New York (CCNY). Paul researched neural correlates of phantom auditory percepts as part of his thesis and specializes in neural signal processing and psychoacoustics.

Lucas C. Parra received the Ph.D. degree in physics from the Ludwig Maximilian University, Germany. He is currently Professor at City College of New York and specializes in signal and image processing, auditory and visual psychophysics, computational and mathematical modeling. He is a co-Founder of Neuromattters.

Paul Sajda has a B.S. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, and the M.S. and Ph.D. degrees in bioengineering from the University of Pennsylvania (UPenn). He is currently Professor at Columbia University, New York and specializes in neurocomputational modeling and neuroengineering, pattern recognition, adaptive image and signal processing. He is a co-Founder of Neuromattters.