SANDIA REPORT
SAND2018-11341
Unlimited Release
Printed September 2018

Patterns of Analyst Attention

Laura E. Matzen, Mallory C. Stites, Michael C. S. Trumbo & Aaron Jones

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology and Engineering Solutions of Sandia, LLC.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	https://classic.ntis.gov/help/order-methods/

SAND2018-
113411134111341113411134111341113411134111341113411134111341113411134111341113
411134111341113411134111341113411134111341113411134111341113411134111341113411
134111341113411134111341113411134111341113411134111341113411134111341113411134
11134111341113411134111341113411134111341113411134111341
Printed September 2018
Unlimited Release

Patterns of Analyst Attention

Laura E. Matzen, Mallory C. Stites, Michael C. S. Trumbo & Aaron Jones

Applied Cognitive Science
Sandia National Laboratories
P. O. Box 5800
Albuquerque, New Mexico 87185-MS1327

Abstract

This report describes a seedling project in which we developed experimental paradigms for studying patterns of analyst attention to streaming data. The project identified key structure features that can be used to generate appropriate stimuli for nearly any mission domain.

ACKNOWLEDGMENTS

We would like thank Brad Aimone and Christina Ting for their support of this project. We would also like to thank Gina Kuperberg and Phil Holcomb for sharing the stimuli from Sitnivoka et al., 2008, which formed the basis for Experiment 1.

TABLE OF CONTENTS

1. Background 6
   1.1. Project Goals 6
   1.2. Developing Experimental Analogs of Streaming Sensor Data 8
   1.3. Combining EEG and Eye Tracking to Assess Analyst Attention 8
2. Experiment 1 10
   2.1. Participants 12
   2.2. Materials 12
   2.3. Apparatus 13
   2.4. Procedure 13
3. Experiment 2 15
   3.1. Participants 15
   3.2. Apparatus 15
   3.3. Materials and Procedure 16
4. Results and Conclusions 21
5. References 23

FIGURES

Figure 3.1. A scenario in which scientist characters are seen entering and exiting a research facility 16
Figure 3.2. A parking lot scenario in which civilians are seen entering and exiting a convenience store 17
Figure 3.3. Depiction of a military installation with soldiers inside the fence and security guards and civilians appearing outside 18
Figure 3.4. Security guard models used in the third video scenario 19
Figure 3.5. Civilian models used in the third video scenario 19
Figure 3.6. Six simultaneous video scenarios 20
Figure 4.1. Image of one participant's EEG data from the six video surveillance task. The green bars in the EEG data indicate regions that are time-locked to the participant's responses and corresponding eye tracking data 21

1. BACKGROUND

In many sensor systems, particularly systems operating on streaming data, the volume of data acquired outstrips computational and analyst capacity. Analysts are confronted with a deluge of data, much of which falls arbitrarily to the cutting room floor. New approaches are needed to triage, or compress, these massive streams of data so that the most decision-relevant segments are retained for use by the analysts or by subsequent automated processing systems.

The goal of this seedling project was to develop the basic science associated with understanding the patterns of analyst perception and attention when processing streaming sensor data. We aimed to test the feasibility of developing lossy compression algorithms for streaming data that are based on patterns of human attention. In many existing lossy compression algorithms, knowledge about human perception is used to drop nonessential data so that file sizes are reduced without impacting how a user perceives the information. We hypothesize that an analogous approach could be used to triage streaming sensor data by modeling patterns of human attention and identifying data features that draw attention. If this is possible, it could highlight features of interest while compressing or de-emphasizing data that is irrelevant to decision making, resulting in improved human-system performance.

In support of this overarching idea, we proposed to develop a new methodology for using cognitive neuroscience techniques to inform the development of compression algorithms. Specifically, we used a combination of eye tracking and electroencephalography (EEG) methods to characterize patterns of human attention when viewing streaming data. This seedling project supported a pilot study in which we developed and tested this methodology with a small group of participants. We developed stimuli that have an attentional profile similar to streaming sensor data, such as full-motion video (i.e., a continuous stream of irrelevant events with important events interspersed at unpredictable intervals). We collected a pilot dataset using these experimental paradigms and stimuli to assess their effectiveness. These pilot results served as a proof-of-concept and will be used to support future proposals that will expand upon this preliminary work.

1.1. Project Goals

This project had three main goals:

1. Develop experimental analogs of streaming sensor data that mimic the attentional demands of streaming data but provide experimenters with full control of variables of interest.

2. Combine EEG and eye tracking data to track the time course of analysts' attention to features in streaming data.

3. Collect a small EEG/eye tracking dataset to assess the success of goals 1 and 2.

One of the key elements that is required for studying patterns of analyst perception and attention in streaming data is a well-defined stimulus set that will allow experimenters to control factors that drive human attention. These factors include the frequency of events of interest and distractor events, the timing and duration of the events, the perceptibility of the events, and the overall length of the task. The stimulus set must be an effective analog for streaming sensor data while providing experimenters with control over each of the relevant attentional variables. Our first goal was to develop such a stimulus set, drawing on the existing scientific literature on human attention and visual cognition.

Our second goal was to integrate the stimulus set into a test environment that can combine eye tracking and EEG data collection modalities. Eye tracking provides information about what an analyst looked at in a streaming data set, and EEG provides information about whether and how that information was processed. EEG records millisecond-level information about the electrical activity of the brain, and the relationships between specific patterns in EEG data and neural processes related to attention are well established in the cognitive neuroscience literature (see Fabiani, Gratton & Federmeier, 2007 and Herrmann & Knight, 2001 for reviews, see also Harmony et al., 1996; Klimesch et al., 1998; Müller, Gruber & Keil, 2000; Sauseng et al., 2005). EEG data can be used to determine the depth of encoding of stimuli (Hanslmayr, Spitzer & Bäuml, 2009; Haass & Matzen, 2011; Rugg, Allan & Birch, 2000; Rugg & Curran, 2007; Sanquist et al., 1980), to determine whether the processing was automatic or controlled (Hoffman, Simons & Houck, 1983; Schneider & Shiffrin, 1977; Strayer & Kramer, 1990), and even to detect leading indicators of an analyst's decision (Gratton et al., 1990; Leuthold, Sommen & Ulrich, 1996). Human attention waxes and wanes over time (Busch & VanRullen, 2010; Egeth & Yantis, 1997; Sarter, Givens & Bruno, 2001), so the combination of eye tracking and EEG will allow us to assess what visual features draw the participants' attention and the level at which that information is processed in the brain (i.e., automatic versus controlled and shallow versus deep encoding). Integrating eye tracking and EEG is crucial for characterizing human attention to streaming data, but it is also a non-trivial technical challenge.

After developing an integrated EEG/eye tracking test environment and a series of experiments in which variables related to attention can be manipulated, we collected a small dataset of EEG and eye tracking data from individuals completing our experimental paradigm. This dataset was intended to test the validity of the design of the streaming data analog as well as the feasibility of integrating eye tracking and EEG for this application.

This project was intended to lay the groundwork for future research on analyst attention and for the development of new data compression algorithms that are based on human attention. Ultimately, our vision is to track attention and processing to different features in the data set and map their contributions to the analyst's decision making process. The resulting mapping could be used to develop data compression and data visualization algorithms that retain and highlight information that is likely to be central to an analyst's decision making process while de-emphasizing information that is unlikely to be relevant.

1.2. Developing Experimental Analogs of Streaming Sensor Data

One of the main challenges in this project was developing experimental paradigms that can be adapted to multiple types of streaming data. To address this challenge, we focused on identifying cognitive aspects of scene perception that would apply to any domain, regardless of the specific details of the data or analytic process.

Wolfe and Horowitz (2017) reviewed the existing literature on the factors that drive visual attention in natural scenes. They identified several key factors that have been shown to drive attention across multiple studies. The factors that were most relevant to our investigation were the following:

1. Top-down drivers of attention, such as the viewer's task and the relative value of search targets
2. Bottom-up drivers of attention, such as the physical features of the stimuli
3. Syntactic constraints within scenes, such as physical constraints about where certain items could appear
4. Semantic constraints regarding the likelihood of different objects appearing in different parts of a scene

By focusing on these four factors, we developed an experimental framework that can be generalized to any domain. Top-down and bottom-up drivers of attention are well-characterized and can be defined for different mission areas after an understanding of the analysts' data sources, goals, and analytic process has been developed. By framing data streams in terms of their syntactic and semantic constraints, we can further characterize the patterns of attention that any given data stream is most likely to elicit.

The sections below describe the two experimental paradigms that we created to manipulate these parameters and test their impact on participants' patterns of attention. Experiment 1 used still images, presented in sequences that had congruent or incongruent endings, where the incongruent endings could be either semantic or syntactic incongruities. Experiment 2 used animated videos that were created using video game world and character models. The tasks in Experiment 2 mimicked streaming data in the form of physical surveillance videos, and each scenario has semantic and syntactic constraints that can be manipulated by the experimenter.

1.3. Combining EEG and Eye Tracking to Assess Analyst Attention

One of the technical challenges involved in this project was integrating the stimulus presentation, eye tracking, and EEG systems to allow for analysis of multiple data streams. Each of these aspects of the data collection is controlled by a different software system running on separate computers. In order to accurately analyze patterns of attention to streaming data, we must be able to insert time-locked triggers, encoding specific events that happen in the stimuli and the participants' responses to those events, into the eye tracking and EEG data streams with high temporal precision. This is a non-trivial challenge.

We tested a variety of setups for accomplishing the required data synchronization. Ultimately, we used the software package E-Prime 3 to send triggers to both data collection systems. E-Prime recorded the participants' behavioral responses and presented the stimuli such that the onset of each trial or video was synchronized with a refresh of the stimulus presentation monitor. E-Prime sent triggers related to the onset of specific stimuli and participants' responses to the EEG amplifier via parallel port. At the same time, it sent triggers to EyeWorks Record, the eye tracking data collection software, by sending codes over the network to the eye tracking computer. These two sets of triggers, sent simultaneously, allowed us to time-lock both data streams to events of interest and to synchronize them with one another.

2. EXPERIMENT 1

In Experiment 1, we tested how humans process semantic and syntactic violations in scene sequences. Semantic violations refer to violations of meaning-related aspects of a scene, whereas syntactic violations refer to violations of structural or functional elements of a scene (more examples provided below). This experiment will help characterize patterns of human attention to these two types of anomalies occurring in naturalistic scene sequences, data which could then be used to inform compression algorithms for streaming data as to which elements of scenes are and are not important for human viewers to process in order to correctly interpret the scene. Specifically, semantic and syntactic violations scene could have different meanings for the analyst and could help distinguish between a threat or non-threat anomalous event. Although computer vision algorithms could likely learn to distinguish low-level cues from a scene (i.e., is there a person present in the scene), the ability to identify whether an object or event is *unexpected* given the context is still a uniquely human skill.

There is a large body of work that has examined the brain electrical activity elicited by semantic and syntactic violations in linguistic stimuli. Much of this work has been done by collecting event-related brain potentials (ERPs), which are portions of the ongoing electroencephalogram (EEG) recorded at the scalp and time-locked to stimulus events of interest. ERPs are a multi-dimensional data source, and provide information about the timing, amplitude, and polarity of brain activity elicited in response to specific types of stimulus events. Importantly for our purposes, they are automatically elicited even in the absence of a task, and can even show effects that participants cannot report behaviorally (e.g., responses to grammatical violations in second language learners; Tokowicz & MacWhinney, 2005). In this way, ERPs collected to streaming data could potentially reveal anomalies in the data stream that the human analysts may not be able to report behaviorally.

There are two distinct event-related potential (ERP) components that are elicited to semantic and syntactic anomalies: the N400 and the P600. The N400 is a negative-going ERP deflection that is elicited by all meaningful or potentially meaningful stimuli, peaks 300-500 ms post-stimulus onset, and is largest over centro-parietal electrode sites (for review, see Kutas & Federmeier, 2011). Its amplitude is largest for anomalous or totally unexpected stimuli, and is reduced proportionally relative to an items predictability given the preceding context. On the other hand, the P600 is a positive-going ERP deflection elicited by syntactic violations (Osterhout & Holcomb, 1992), as well as a wide range of difficult-to-process syntactic structures and other types of grammatical errors. It is broadly distributed over posterior channels, and tends to peak around 600 ms post-stimulus onset, although its timecourse can vary from 500-100ms post-stimulus onset.

Although the study of syntactic and semantic violations has been most often studied using linguistic stimuli, other work has begun to extend this work into non-linguistic settings (for review, see Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008), demonstrating that these mechanisms reflect how people generally make sense and meaning from the world, as opposed to processes that are strictly linguistic in nature. This experiment is a direct extension of Sitnikova et al. (2008), and as such, we will briefly describe their experimental paradigm and results. Sitnikova et al. (2008) showed participants videos of people performing everyday actions (e.g., turning on water in a bathroom sink, putting shaving cream on face) that could end with either an expected action (e.g., shaving face with razor), a semantic violation (e.g., the scene jumps to the same actor performing a completely unrelated action, like putting food in the microwave), or a syntactic violation (e.g., the person rolling a rolling pin on their face). Across two separate experiments, they found that semantic violations elicited an N400 relative to the congruent condition, whereas syntactic violations elicited a P600 component relative to the congruent condition. They interpreted these findings as showing that the comprehension of real-world events is mediated by two largely distinct processing mechanisms: one, indexed by the N400, that evaluates new information based on the strength of their semantic relationships with previously viewed information, and a second mechanism, indexed by the P600, that evaluates new information based on how it fits the physical constraints of the scene. This is consistent with other work that has also found N400-like components to objects containing incongruent scenes (e.g., Demiral, Malcolm, & Henderson, 2012; Ganis & Kutas, 2003; Mudrik, Lamy, & Deouell, 2009; Sitnikova, Kuperberg, & Holcomb, 2003), although under situations where prediction is made more difficult, P600 effects have been observed instead (Demiral et al., 2012).

The current experiment expands Sitnikova et al. (2008) in two ways. First, we will ask whether the N400 and P600 components to the two types of violations are elicited within the same individual when they encounter intermixed semantic and syntactic violations. Secondly, we will present participants with two scenarios to track simultaneously, one in each visual field (biasing processing to the contralateral hemisphere initially), which will allow us to ask two related questions: 1) can people track two ongoing scenarios sufficiently so as to elicit the canonical ERP effects associated with these violations, and 2) do the two hemispheres respond to violations in scenes differently (from both each other, and from the pattern observed at central fixation)? Federmeier (2007) has shown that the left hemisphere tends to do more predictive, top-down processing, whereas the right hemisphere tends to do more veridical, integrative, bottom-up processing. Given these findings, it is possible that the two hemispheres will show different effects to these violations, especially if predictive processing is necessary to elicit the N400 effect.

These questions are also important to inform the creation of compression algorithms for streaming data, for several reasons. First, it will often be the case that the same analyst will view data streams that contain multiple types of anomalies, and so it is critical to understand how the same person processes several, interleaved types of anomalies. Secondly, analysts will typically view multiple data streams simultaneously, and those presented to their different visual fields will necessarily elicit different neural mechanisms, so it is crucial to characterize these differences in order to incorporate them appropriately into data compression algorithms based on patterns of human attention.

2.1. Participants

Eight employees (five female) of Sandia National Laboratories participated in the pilot data collection. The participants ages ranged from 24-59.

2.2. Materials

Stimuli were adapted from Sitnivoka et al. (2008). The original stimulus set consisted of color movie clips that depicted an actor performing an everyday activity (e.g., getting out a loaf of bread and putting it on a cutting board). Each video was constructed to have three possible endings: congruent, semantically incongruent, and syntactically incongruent. In the congruent condition, the final scene of the clip depicted the individual using an object in a congruent manner that followed from the semantic and syntactic constraints of the scene (e.g., uses a knife to cut the bread). In the syntactically incongruent condition, the actor was depicted using an object in an incongruent way that was inconsistent with meeting their goal-directed behavior in the scene (e.g., using an iron to unsuccessfully cut the loaf of bread). In the semantically incongruent condition, a congruent scene from a different movie clip was spliced onto the end of the original video, to depict the same actor performing an action that was not related to the preceding context, but in which they were using a tool in a realistic way (e.g., ironing pants). The wholly unexpected nature of both movie endings was confirmed by a norming study with 18 separate individuals (see Sitnikova et al., 2008, for details).

In the current study, the videos were adapted in order to be shown as four consecutive still images (to standardize the length of each scenario and enable clear time-locking to events of interest). To do so, individual frames were saved from the beginning, middle, and end of the initial set-up of each video, as well as a single frame from each of the three ending conditions that reflected the action of the scene. For the semantically incongruent endings, the same video frame was used as when the scene appeared as a congruent ending to ensure that these remained physically identical. Participants viewed the scenarios in two different configurations: a 1-scenario condition, in which only one scenario was presented at central fixation, and a 2-scenario condition, in which two scenarios were presented simultaneously (with one image in the left visual field, and one image in the right visual field).

There were 80 critical stimuli, each of which had three possible endings, for a total of 240 possible unique scenarios. Participants saw each stimulus four times total across the 1-and 2-scenario conditions: twice in the 1-scenario condition (once as congruent, once as incongruent), and twice in the 2-scenario condition (once as congruent, once as incongruent). Specifically, every participant saw every stimulus in its congruent form in both the 1- and 2-scenario conditions. However, if they saw it as semantically incongruent in the 1-scenario condition, they would see it was syntactically incongruent in the 2-scenario condition, and vice-versa. Eight counterbalanced lists were created, which ensured that across participants, stimuli were rotated through conditions equally often in the 1-scenario and 2-scenario conditions, and that within the 2-scenario condition, a stimulus' appearance as congruent and incongruent was balanced across visual fields. A set of 80 filler scenarios was created in order to fill in the second position of the 2-scenario condition by drawing from freely available

YouTube videos depicting everyday events similar to those in the critical scenarios (e.g., cooking, cleaning, yard work). All of the filler scenarios consisted of four individual video frames, and all endings were semantically and syntactically congruent. Each of the four instances of a critical stimulus was paired randomly with a filler scenario, and this pairing of scenarios rotated through conditions together (to ensure that the physical characteristics of the screen that appeared alongside each critical scenario were held constant across conditions).

Participants saw 320 trials total, 160 in the 1-scenario condition, and 160 in the 2-scenario condition. Within these conditions, half of all trials contained a congruent ending, and half contained an incongruent ending (half of which were semantically incongruent, and half were syntactically incongruent). Trials were presented in a fixed random order . Order of the 1- and 2-scenario conditions was counterbalanced, such that half of participants started in the 1-scenario condition, and half started in the 2-scenario condition.

2.3. Apparatus

EEG data were collected using an Advanced Neuro Technologies (ANT) system with a 128-channel, Duke layout cap and digitized at 250 Hz. Stimulus presentation was controlled by E-Prime software. For each event within Experiment 1, such as the presentation of an image or a button press made by a participant, a trigger was sent to the EEG amplifier via parallel port.

2.4. Procedure

Participants were tested individually in a sound-attenuated booth. Participants sat 90 cm away from the computer monitor, such that videos were viewed at a size of 4 degrees of visual angle horizontally (as in Sitnikova et al, 2008). In the 1-scenario condition, scenarios were viewed at central fixation. In the 2-scenario condition, participants maintained fixation on a central fixation cross to minimize eye movement artifacts and help ensure that participants were attending to both scenarios simultaneously. Trial sequences went as follows. Each trial began with a fixation cross for 1500 ms (jittered by 10-100 ms across trials). In the 1-scenario condition, the images replaced the fixation across. Each image of the sequence was visible for 1500 ms with a 300 ms inter-stimulus interval . After the final image, a red question mark appeared, to which participants responded as to whether the final image of the scene was congruent with the preceding context (“yes”) or not (“no”). The yes response was always made by pressing a trigger on the left side of a game pad, and the no response was always made by pressing a trigger on the right side . The question mark remained on the screen until the participant responded, at which point the next trial began. Participants were instructed that if they

In the 2-scenario condition, the same stimulus timing was used. The only difference was that the fixation cross remained on the screen (to make it easier for participants to maintain central fixation), and the two scenarios appeared simultaneously on the screen immediately adjacent to the fixation cross (beginning approximately 1 degree of visual angle from fixation, and each image subtending four degrees of visual angle horizontally). At the end of the trial, participants were instructed to respond “no” if either of the final images was incongruent with its preceding scenario (although only the critical scenarios had incongruent endings).

The 1-scenario and 2-scenario blocks were each subdivided into four blocks of work, each consisting of 40 trials and which took approximately 5-7 minutes to complete. Participants received a self-timed break between each block, as well as a slightly longer break between the 1- and 2-scenario blocks.

3. EXPERIMENT 2

In Experiment 2, we developed a method of creating streaming (video) stimuli. Traditionally, visual search experiments rely primarily on static images of stimuli such as letters and basic shapes (Wolfe et al., 2011) which may not be ecologically valid for analysts who examine streaming data feeds such as guards monitoring video feeds of security cameras. EEG research has been conducted with stimuli in motion (e.g., Hirai et al., 2006; Steel et al., 2016), and a number of databases of motion stimuli exist (e.g, Mandery et al., 2015). In addition, behavioral research using stimuli in motion has suggested that the number of moving objects influences the percentage of missed targets (Sulman et al., 2008). There are several limitations to this previous work. For instance, Sulman and colleagues (2008) used stimuli in motion, but the stimuli were shapes. Biological motion appears to have a particular influence on salience and interpretation of the scene – including discernment of the intentions of the actors (Steel et al. 2014). Therefore, the use of human actors instead of abstract shapes is critical for research in security domains. It can be difficult, however, to develop such stimuli. The use of motion-capture, professional animations, or live-action video can be expensive, time-consuming, difficult to modify, and have a steep learning curve that acts as a barrier to entry.

Therefore, the primary goal of Experiment 2 was to develop a method of creating video stimuli that are analogous to streaming data and contain the desired semantic and syntactic constraints, that can be created and modified easily, and are simple to use and understand. In addition, the stimuli and method of presentation must provide sufficient experimental control such that timing and distribution of stimuli enable rigorous experimentation protocols. Given these requirements, and the limitations of previous methods, we settled on using a sandbox physics game (Garry's Mod, created by Facepunch Studios; https://gmod.facepunch.com/). This software platform was built to allow for easy modification of scenes and characters. Furthermore, actions of characters, such as spawn location, movement path, and movement speed can be scripted via Lua files to provide precise control over situations. Video can be captured as scenarios play out, and subsequently can be presented as stimuli, as in the current proof-of-concept experiment. We developed tasks that mimic streaming data in a physical security scenario, in which a human must monitor video feeds and respond to potential threats. EEG and eye tracking were recorded while participants watched the videos and responded to events they were instructed to view as threatening.

3.1. Participants

Eight employees (five female) of Sandia National Laboratories participated in the pilot data collection (the same participants who were included in Experiment 1).

3.2. Apparatus

EEG data were collected using an Advanced Neuro Technologies (ANT) system with a 128-channel, Duke layout cap and digitized at 250 Hz. Stimulus presentation was controlled by E-Prime software. Eye tracking data were collected using the EyeWorks

Record software produced by Eyetracking, Inc. For each event within Experiment 2, such as the initiation of a video scenario or a button press made by a participant, a trigger was sent to the EEG amplifier via parallel port and to EyeWorks Record over the network.

3.3. Materials and Procedure

Three video scenarios were created to reflect common types of monitoring tasks in the physical security domain. In the first scenario, participants were presented with scientist character models entering a fictional research facility (see Figure 3.1, below). Participants were told that scientists walking into the building was a normal activity that did not require a response. Scientists running into the building represented an anomaly worth noting (via a button press), but did not constitute a threat (perhaps they are just late to a meeting). Scientists running out of the building should be considered a threat as they may have stolen sensitive material, and the participant should alert security by pressing a “threat detected” button. There were 140 stimuli total; 100 common stimuli (walkers), 20 non-threat distractors (runners into the building), and 20 threats (runners out of the building). Non-threat and threat stimuli could co-occur with common stimuli, but did not co-occur with each other. Common stimuli were spaced an average of 6 seconds apart, with up to 500 ms of jitter on either side (5.5 s – 6.5 s); the 40 runners were spaced an average of 15 seconds apart, with up to 1250 ms of jitter on either side (13.75 s – 16.25 s). As runner co-occurred with walkers, the total duration of the video scenario was 10 minutes (100 walkers with an average of 6 seconds in between).

Figure 3.1. A scenario in which scientist characters are seen entering and exiting a research facility

In the second video scenario, civilian character models were seen entering or exiting a convenience store (see Figure 3.2 below). Participants were instructed to ignore characters entering the store, but to indicate via a non-threat button press characters who exited the store, went into the parking lot, and passed between vehicles. They were asked to press the “threat detected” button when characters whom paused to peer into a car window, as this could indicate a potential car theft. Distribution (100 entering the store, 20 non-threats exiting, 20 threats exiting) and timing (an average of 6 s with up to 500 ms jitter in between common stimuli; an average of 15 s with up to 1250 ms of jitter for the uncommon stimuli) was the same as in the hallway scenario.

Figure 3.2. A parking lot scenario in which civilians are seen entering and exiting a convenience store

In a third scenario, participants were presented with a depiction of a military installation with a fence separating the installation from public space (see Figure 3.3 below).

Figure 3.3. Depiction of a military installation with soldiers inside the fence and security guards and civilians appearing outside

Participants were told that soldiers were performing their morning exercises, and soldiers running inside of the fence could be safely ignored. The fence may be approached from the outside by both civilians and security guards. Civilians walking by the fence without stopping were a notable (button-press) non-threat event, as were security guards stopping to check the fence as per their duties. Conversely, participants were told that civilians stopping at the fence were a threat, as were security guards failing to stop at the fence (see Figures 3.4 and 3.5 below for guard and civilian models, respectively).

Figure 3.4. Security guard models used in the third video scenario

Figure 3.5. Civilian models used in the third video scenario

Therefore, this scenario required a conjunction consideration (character model + action) in order for a given character to be deemed a threat or non-threat. Timing was the same as the first two scenarios (6 s on average in between with up to 500 ms jitter on either side for the common soldier stimuli; an average of 15 s in between for civilian and guard stimuli with up to 1250 ms jitter on either side). Distribution of stimuli was similar as well, though with conjunction considerations, as follows: 100

common soldier stimuli, 10 civilian threats, 10 civilian non-threats, 10 guard threats, 10 guard non-threats.

These three scenarios were presented in a randomized order following completion of Experiment 1. Finally, participants were presented with six video stimuli at one time – the three video scenarios, each duplicated once with a different offset in timing and different character models for the scenario 1 and 2 duplicates (see Figure 3.6 below).

Figure 3.6. Six simultaneous video scenarios

This scenario involved looping the 10-minute scenarios once, such that total presentation time was 20 minutes. This was designed to evaluate the effects of multiple presentations of stimuli on ability of participants to detect threats and notable non-threats (they were asked to respond to the same types of events as they had during initial presentation of the three scenarios) and to examine the effects of fatigue on performance.

4. RESULTS AND CONCLUSIONS

As the goal of this project was the collection of a pilot dataset that will support future proposals, we will not present a detailed analysis of the data in this report. However, preliminary analyses of the data indicate that we succeeded in creating a series of tasks that progressively increases the fidelity of the stimuli and the difficulty of allocating attention effectively to one or more scenes. In addition, we demonstrated that we can effectively combine our eye tracking and EEG systems by sending simultaneous triggers to both recording systems from a single stimulus presentation system (see Figure 4.1 for an example). This combination will benefit any future projects that would like to collect simultaneous EEG and eye tracking data.

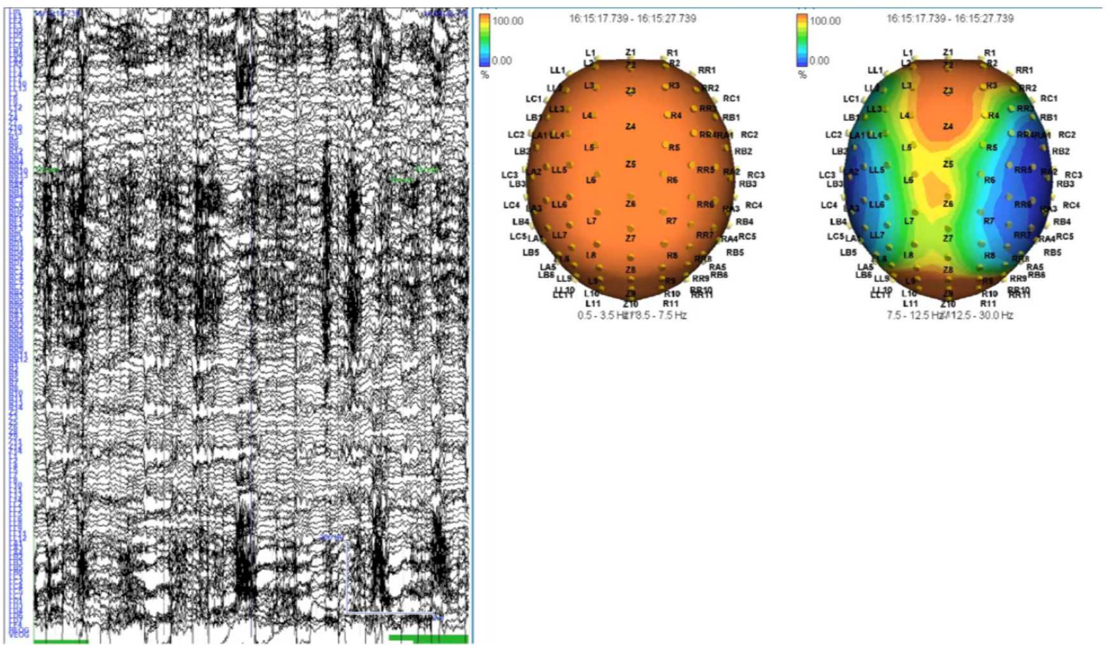

Figure 4.1. Image of one participant's EEG data from the six video surveillance task. The green bars in the EEG data indicate regions that are time-locked to the participant's responses and corresponding eye tracking data

As outlined in the introduction, the goal of this project was to develop experimental paradigms that lay the groundwork for future studies that will investigate the feasibility of developing compression algorithms based on patterns of human attention. We extended the existing research on visual attention in natural scenes and bridged that work with the literature on the event-related potentials elicited by semantic and syntactic anomalies. In addition, we explored new methods for experiment stimulus creation using video game engines and scripting. We found that this approach can produce realistic tasks in which experimenters can manipulate the semantic and syntactic constraints within one or more scenes. Factors that drive top-down and bottom-up attention, such as the participants' task and the visual characteristics of the scenes can be controlled to study the dynamics of human attention in the context of monitoring streaming data. The basic characteristics that were included in the exemplar tasks can be manipulated to match the constraints of any mission area or data type.

In summary, this project successfully developed an experimental framework that is both novel and highly adaptable. This work leaves us well positioned to study the dynamics of analyst attention in any mission domain that involves dynamic streaming data.

5. REFERENCES

1. Bigdely-Shamlo, N., Vankov, A., Ramirez, R. R., & Makeig, S. (2008). Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *16*(5), 432-441.
2. Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048-16053.
3. Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia*, *50*(7), 1271-1285.
4. Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual review of psychology*, *48*(1), 269-297.
5. Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications.
6. Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.
7. Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123-144.
8. Gratton, G., Bosco, C. M., Kramer, A. F., Coles, M. G., Wickens, C. D., & Donchin, E. (1990). Event-related brain potentials as indices of information extraction and response priming. *Electroencephalography and clinical neurophysiology*, *75*(5), 419-432.
9. Haass, M. J., & Matzen, L. E. (2011, July). Using computational modeling to assess use of cognitive strategies. In *International Conference on Foundations of Augmented Cognition* (pp. 77-86). Springer Berlin Heidelberg.
10. Hanslmayr, S., Spitzer, B., & Bäuml, K. H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral cortex*, *19*(7), 1631-1640.
11. Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., ... & Rodríguez, M. (1996). EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *International journal of psychophysiology*, *24*(1), 161-171.
12. Hirai, M., & Hiraki, K. (2006). The relative importance of spatial versus temporal structure in the perception of biological motion: an event-related potential study. *Cognition*, *99*(1), B15-B29.
13. Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465-476.
14. Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). Event-Related Potentials During Controlled and Automatic Target Detection. *Psychophysiology*, *20*(6), 625-632.

15. Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., & Hild, K. E. (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, *74*(12), 2041-2051.

16. Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., & Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience letters*, *244*(2), 73-76.

17. Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

18. Leuthold, H., Sommer, W., & Ulrich, R. (1996). Partial advance information and response preparation: inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *125*(3), 307.

19. Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N., & Asfour, T. (2015, July). The KIT whole-body human motion database. In *Advanced Robotics (ICAR), 2015 International Conference on* (pp. 329-336). IEEE.

20. Matzen, L. E., Haass, M. J., Tran, J., & McNamara, L. A. (2016). Using eye tracking metrics and visual saliency maps to assess image utility. *Electronic Imaging*, *2016*(16), 1-8.

21. Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*(2), 507-517.

22. Müller, M. M., Gruber, T., & Keil, A. (2000). Modulation of induced gamma band activity in the human EEG by attention and visual information processing. *International Journal of Psychophysiology*, *38*(3), 283-299.

23. Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

24. Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, *12*(4), 664-678.

25. Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences*, *11*(6), 251-257.

26. Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K., & Lindsley, D.B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. Psychophysiology, 17, 568–576.

27. Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain research reviews*, *35*(2), 146-160.

28. Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., ... & Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience*, *22*(11), 2917-2926.

29. Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review*, *84*(1), 1.

30. Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of cognitive neuroscience*, *20*(11), 2037-2057.
31. Strayer, D. L., & Kramer, A. F. (1990). Attentional requirements of automatic and controlled processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 67.
32. Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real–world events: An electrophysiological investigation. *Psychophysiology*, *40*(1), 160-164.
33. Sulman, N., Sanocki, T., Goldgof, D., & Kasturi, R. (2008, December). How effective is human video surveillance performance?. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1-3). IEEE.
34. Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in second language acquisition*, *27*(2), 173-204.
35. Trumbo, M. C., Matzen, L. E., Silva, A., Haass, M. J., Divis, K. & Speed, A. (2015). Through a scanner quickly: Elicitation of P3 in Transportation Security Officers following rapid image presentation and categorization. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition. Lecture Notes in Artificial Intelligence, 9183*, 348-360.
36. Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650.
37. Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058.
38. Womelsdorf, T., & Fries, P. (2007). The role of neuronal synchronization in selective attention. *Current opinion in neurobiology*, *17*(2), 154-160.

DISTRIBUTION

1	MS0152	Phil Bennett	9131
1	MS0899	Technical Library	9536 (electronic copy)
1	MS0359	D. Chavez, LDRD Office	1911

Sandia National Laboratories