SAND2022-2063C

A Method of Developing Video Stimuli that are Amenable to Neuroimaging Analysis: An EEG Pilot Study

Michael C. Trumbo[1*], Aaron P. Jones[1], Bradley M. Robert[1], Derek Trumbo[1], and Laura E. Matzen[1]

[1] Sandia National Laboratories, Albuquerque, USA
{mctrumb; ajones3; bmrober; dtrumbo; lematze} @sandia.gov

Abstract. Creation of streaming video stimuli that allow for strict experimental control while providing ease of scene manipulation is difficult to achieve but desired by researchers seeking to approach ecological validity in contexts that involve processing streaming visual information. To that end, we propose leveraging video game modding tools as a method of creating research quality stimuli. As a pilot effort, we used a video game sandbox tool (Garry's Mod) to create three steaming video scenarios designed to mimic video feeds that physical security personnel might observe. All scenarios required participants to identify the presences of a threat appearing during the video feed. Each scenario differed in level of complexity, in that one scenario required only location monitoring, one required location and action monitoring, and one required location, action, and conjunction monitoring in that when an action was performed it was only considered a threat when performed by a certain character model. While there was no behavioral effect of scenario in terms of accuracy or response times, in all scenarios we found evidence of a P300 when comparing response to threatening stimuli to that of standard stimuli. Results therefore indicate that sufficient levels of experimental control may be achieved to allow for the precise timing required for ERP analysis. Thus, we demonstrate the feasibility of using existing modding tools to create video scenarios amenable to neuroimaging analysis.

Keywords: EEG; ERP; Ecological Validity; P300.

1 Introduction

Traditionally, visual search experiments rely primarily on static images of stimuli such as letters and basic shapes [1] which may not be ecologically valid for analysts who examine streaming data feeds such as security personnel monitoring video feeds of security cameras. Behavioral research using stimuli in motion has suggested that factors such as the number of moving objects, the number of video feeds, the motion of both the camera and of the target, the scene complexity, and the area of coverage influence the percentage of missed targets [2-4]. In addition, biological motion appears to have a

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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.

particular influence on salience and interpretation of the scene – including discernment of the intentions of the actors [5].

Development of stimuli in this domain should allow for manipulation and implementation of these factors, as well as factors that drive human attention such as 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 [6]. 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 alter, or have a steep learning curve that acts as a barrier to entry.

As a ubiquitous activity, video game playing has attracted the attention of cognitive scientists who seek to understand who plays video games and why they do so [7, 8] as well as the impact of video games on attention, memory, and other cognitive faculties [9-11]. Conversely, other researchers have focused on the opposite direction – how video game players influence the nature of the games they play.

The term "modding" is used as a slang term in reference to the act of making modifications to the existing aesthetics, experience, or structure of a video game [12; see 13 for a history of modding]. The goals of modding may be to improve the interactivity of a game, adjust the graphics or responsiveness, control the difficulty, create additional content, develop programming skills, or to satisfy various psychological needs such as the need for self-expression, co-operation, and involvement in a community with a shared interest [12, 14, 15]. From a game developer perspective, successful engagement with modding communities can result in new features and content that strengthen the brand-name, add to the shelf-life of the game, increase customer loyalty, improve sales of and spark interest in the original game (as mods often require the original software to run), serve as a method of identifying and recruiting skilled developers, and reduce research & development as well as marketing costs by providing insight into the types of features that are desired by and popular with the gaming community [12, 13, 16].

Therefore, many game development companies encourage the gaming community to act as prosumers engaged in participatory design via the release of tools to make modding easier [14, 17, 18]. Thus, there is a demand for tools that assist in game modifications, and developers may benefit from providing such tools. The result of mutual interest between video game players and video game developers in accessible modding presents an opportunity for educators and researchers to leverage readily available tools to adapt game scenarios to serve an educational [19-21] or a research function [22].

One niche within the modding community is the use of game engines to create cinematic productions, referred to as "machinima" – a misspelled portmanteaux of "machine" and "cinema." There are four primary methods of accomplishing this type of production, including reliance on the AI of the game engine, digital puppetry, manipulation of the in-game camera, and – critically for neuroimaging considerations – precise scripting of actions [23]. This has previously been suggested as a method of creating educational materials [24]. We therefore propose to use these tools as a method of creating ecologically valid streaming stimuli amenable to neuroimaging analysis.

In the current study, we consider the feasibility of using modding tools to develop 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) and provide experimenters with full control of variables of interest. Furthermore, we propose that stimuli developed in this fashion can be made amenable to neuroimaging analysis via strict control over the timing and duration of events. Electroencephalography (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 25 and 26 for reviews]. EEG data can be used to determine the depth of encoding of stimuli [27-29], to determine whether the processing was automatic or controlled [30, 31], and even to detect leading indicators of an analyst's decision [32, 33].

A common paradigm designed to elicit an EEG response is the so-called 'oddball' paradigm, in which (typically) two stimuli are presented in a random order with one occurring less frequently than the other (the infrequent stimulus thus being the oddball); participants are required to identify the rare target stimulus. A variation of this paradigm, the three-stimulus oddball, involves the addition of an infrequent non-target stimulus along with the infrequent target stimulus and frequent standard stimulus [34]. We chose to model our creation of streaming video stimuli after this 3-stimulus implementation of the oddball paradigm, reasoning that in the context of physical security there may be instances in which rare non-threatening events occur alongside rare threatening events and frequent banal events. With this consideration in mind, three video scenarios were created, replacing the static-letter stimuli used in previous research [34]. with streaming events that represent more ecologically-valid scenarios for physical security operators. EEG was recorded while participants watched the videos and responded to events they were instructed to view as threatening.

2 Method

2.1 Scenario Development

Given the above requirements regarding need for experimental control and the limitations of previous methods, we used "Garry's Mod" [35]. Garry's Mod is a physics-based sandbox game with no set objectives that allows users to create simulated environments that contain both static and dynamic elements. Simulations take place on maps that define the physical space for the simulation, where terrain has been sculpted, buildings and structures have been placed, and other objects have been arranged in the environment. Several default maps and numerous other objects are provided with the initial install of the game, but it allows custom maps and objects to be created and loaded as well. The game allows players to spawn and manipulate elements in the environment ranging from furniture, weapons, vehicles, and other non-player controlled (NPC), AI-driven characters. Game engines, including the one selected for this study, often include a physics engine [16], and games are often designed to mimic our intuitive mental representations of how physical objects should move and interact with each other, further allowing for the development of ecologically valid stimuli [36].

Critically, Garry's Mod allows programmatic control over actions of characters, such as spawn location, movement path, and movement speed, enabling precise control over timing of situations. There were three scripts required to create a framework to automate scenarios:

1. A script to start a scenario.
2. A script to stop a scenario.
3. A script representing a custom NPC type that executes a series of tasks each with a given duration.

Beyond that, for each scenario we tested in this experiment, we had to create a single scenario definition script. This script defines how many NPCs will be in the scenario, what models are used to represent them visually, and the tasks each NPC is to execute during that scenario. These scripts were developed to match the experimental design and are specific to a given map.

To start a scenario, a given scenario definition script that corresponds to the current map is loaded and then the start scenario script is executed. The latter reads the scenario definition, sorts the defined NPCs by when they should first appear in the simulation, and then uses what is called a *hook* in Garry's Mod to spawn the NPCs at the times requested in the scenario's definition. Hooks allow the scripts to respond to certain events that happen in the simulation, like the user pressing a key. The hook used here is Garry Mod's "Think" hook, which fires on every game frame, allowing our script to repeatedly check whether the next NPC should be spawned yet.

These NPCs are defined by a script based off the "NextBot" entity type available in Garry's Mod. Once spawned, each NPC will follow the list of tasks given to it. Each task defined in the scenario definition script will contain parameters specific to that task's type. For example, movement tasks will have destination coordinates, speed, and acceleration provided.

Finally, to stop a scenario the stop scenario script is executed. This script removes the "Think" hook and removes all custom NPCs that are still loaded in the scenario. This small handful of Garry's Mod Lua scripts allowed us to quickly and flexibly define the scenarios placed before our subjects. Within this development framework, video can be captured as scenarios play out, and subsequently can be presented as stimuli, as in the current pilot experiment.

2.2 Scenarios

Stimuli EEG signals can be time-locked to events such as the onset of a stimulus, resulting in event-related potentials (ERPs) that provide information about the brain's processing in relation to those events [37]. The so-called P300 ERP refers to activity that occurs roughly 300 ms following an event and is thought to reflect processes such as attention allocation and categorization [38]. The P300 is often studied in the context of the 'oddball' paradigm we modeled the video scenarios after for this study [34]. With Three video scenarios were created to reflect common types of monitoring tasks in the physical security domain, including identification of a hazardous situation (Scenario 1:

Hallway), potential theft (Scenario 2: Parking Lot), and suspicious behavior (Scenario 3: Fence).

Scenario 1 (Hallway): In the first scenario, participants were presented with scientist character models entering a fictional research facility (see Figure 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 (e.g., perhaps they are just late to a meeting). Scientists running out of the building should be considered a threat as they may be fleeing a hazardous scenario, and the participant should sound an alarm 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 runners co-occurred with walkers, the total duration of the video scenario was 10 minutes (100 walkers with an average of 6 seconds in between).

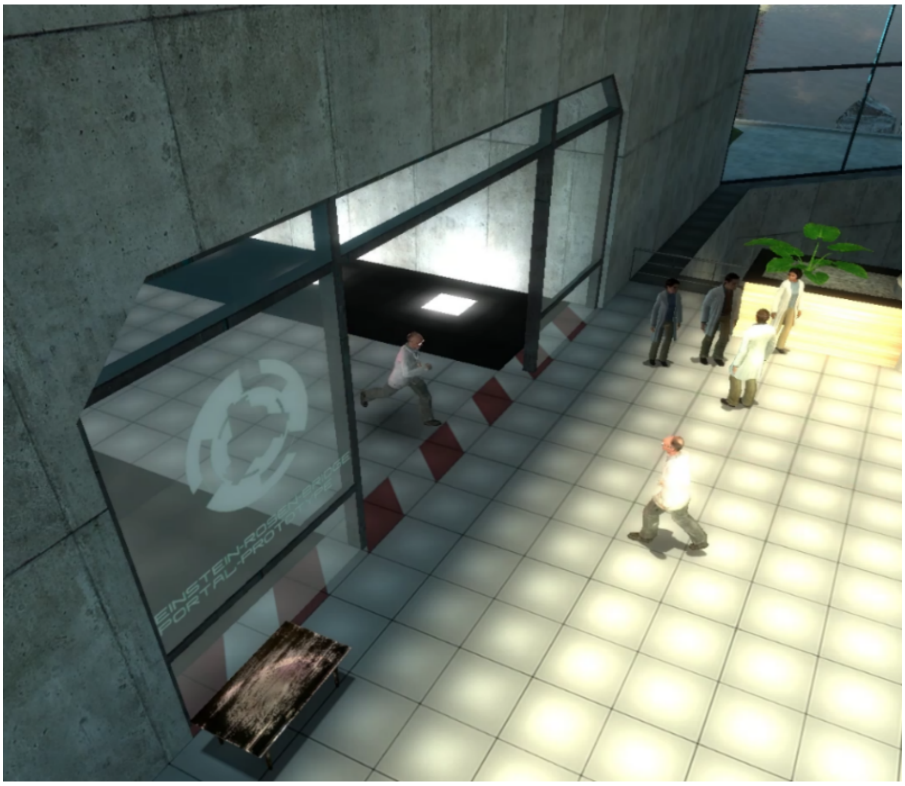

Fig. 1. An image of Scenario 1 (Hallway). Scientist character models are seen entering and exiting a research facility.

Scenario 2 (Parking Lot): In the second video scenario, civilian character models were seen entering or exiting a convenience store (see Figure 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.

Fig. 2. An image of Scenario 2 (Parking Lot). Civilian character models are seen entering and exiting a convenience store.

Scenario 3 (Fence): 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 below). 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 4 and 5 below for guard and civilian models, respectively).

Fig. 3. An image of Scenario 3 (Fence). A military installation is displayed with soldier character models inside the fence and security guard and civilian models outside of the fence.

Fig. 4. Security guard models used in the third (Fence) video scenario.

Fig. 5. Civilian models used in the third (Fence) 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.

2.3 Participants

Eight employees (five female; age 24-59) of Sandia National Laboratories participated in data collection.

2.4 Procedure

EEG data were collected using an Advanced Neuro Technologies (ANT) system with a 128-channel, Duke layout cap and digitized at 250 Hz. Participants were tested individually in a sound-attenuated booth. Participants sat 90 cm away from the computer monitor. Scenarios were presented electronically in a random order using the E-Prime 3.0 software [39]. Scenarios were presented in randomized order as three separate video files (i.e., one video file of approximately 10-minute duration each for Scenario 1, 2, and 3), with a short break given after completion of each scenario. E-Prime 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. For Scenario 1 (Hallway), trigger timing corresponded with the first frame at which pixels of a scientist character model became visible, as participants were able to determine the categorization of stimuli based on the screen location onset of the character model. In Scenario 2 (Parking Lot), triggers for standard stimuli were sent corresponding to first frame of onset of the character model. For rare threat and rate non-threat stimuli, triggers were sent corresponding to a critical action. Rare non-threat and rare threat character models would both stop near a car, then either continue walking (non-threat) or duck near the vehicle door (threat). The trigger was sent at the first frame of this critical action, as that represented the point at which the participant was able to categorize the stimuli. For Scenario 3 (Fence), triggers were sent in a similar fashion as in Scenario 2 (Parking Lot); standard models generated triggers at onset, while rare threat stimuli and rare non-threat stimuli generated triggers at the first frame of a critical decision point (continuing to walk past a security gate or stopping at the gate).

Participants responded using button presses on a controller that were recorded via E-Prime. As in previous ERP research using the oddball paradigm, participants were asked to respond to both rare types of stimuli (threat and non-threat) in order to avoid motor contamination [40]. This allowed for a clean comparison between rare events that were framed as threats and rare events that were framed as benign.

EEG Preprocessing/P300 Measurement: EEG data were preprocessed in EEGLAB v2019.1 [41], using the FASTER toolbox [42], all in MATLAB 2017b [43]. Raw data were bandpass filtered from 1-50 Hz, and Independent Components Analysis (ICA; runICA.m function) was used for artifact rejection. Automatic artifactual rejection was accomplished within the FASTER toolbox using a 3 Z threshold for median gradient, spectral slope, spatial kurtosis, Hurst exponent, and EOG correlation. Data were then re-referenced to the average of all channels. After running ICA artifact rejection procedures, no trials were rejected based on artifact, using a 3 Z threshold for deviation from mean, variance, and amplitude range. Using ERPLAB [44], the data were then epoched into 1 sec bins, from -200ms to 800ms post-stimulus, and baseline corrected from -100ms to 0ms pre-stimulus. Averaged event-related potentials (ERPs) were then generated and lowpass filtered to 30Hz for analysis. P300 amplitude values were calculated as the peak amplitude between a 250ms to 500ms post-stimulus latency range over 15 channels (45; see Figure 6 for the analogous channels used from the ANT Duke layout). Finally, ERPs were grand averaged across subjects for visualization in the form of topographical scalp maps. In addition, a representative waveform was produced for each scenario separately for visualization purposes.

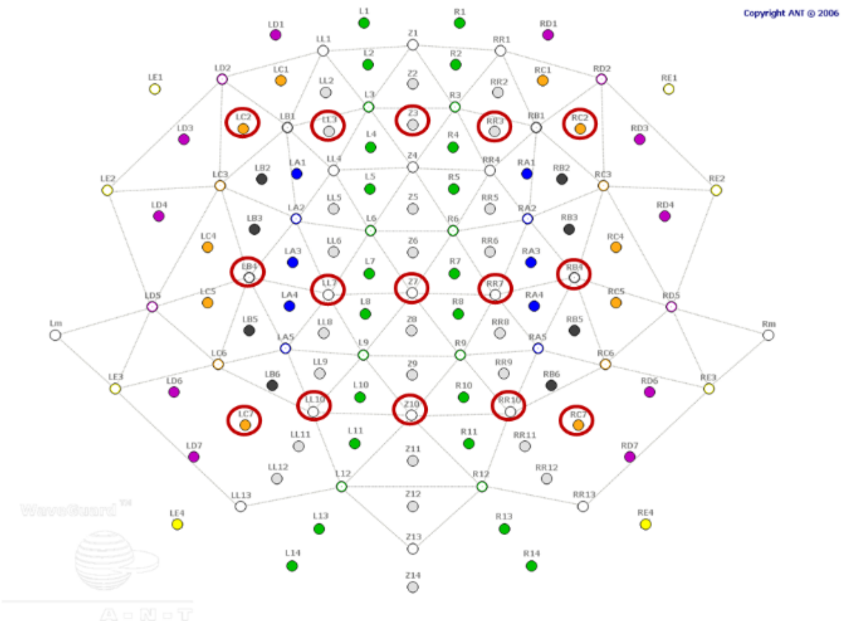

Fig. 6. Advanced Neuro Technologies (ANT) system 128-channel, Duke layout. Channels used for P300 analysis are circled in red and correspond to the 15 channels used in [45].

3 Results

3.1 Behavioral

Target accuracy data for each scenario (1 – Hallway; 2 – Fence; 3 – Parking Lot) were investigated using a mixed effects ANOVA (subject entered as a random factor and scenario [3-levels; hallway, fence and parking lot] entered as a fixed factor). Data were subjected to outlier analysis. Data for one participant was removed from the Fence and the Parking Lot scenarios due to confusion regarding the correct response button, resulting in abnormally low accuracy, leaving a total of eight participants for the hallway scenario and seven participants each for the Fence and Parking Lot scenarios. Analysis of hit rate did not reveal any significant effect of scenario ($F_{(2\ 19)} = 1.24$, $p = 0.313$, see Figure 7).

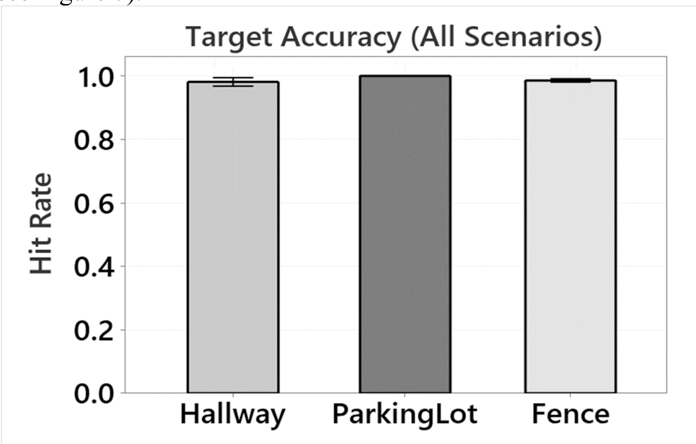

Fig. 7. Average target accuracy reported as hit rate for each scenario; no significant effects were observed. Error bars = +/- 1 SEM.

3.2 P300 Amplitude

EEG data from one participant was removed due to poor data quality leaving seven participants to be included in the P300 analysis. The participant for whom behavioral data was excluded was included in the EEG analysis because they understood the task, but merely pressed the wrong response button, presumably leaving the P300 intact for this participant. The data for each scenario (1 – Hallway; 2 – Fence; 3 – Parking Lot) were investigated individually using a linear mixed effects model, with subject specified as a random factor, and stimulus type (3 levels – standard, rare threat, rare no-threat) as a fixed factor. Amplitude data from the 15 EEG channels were averaged, such that each participant had one value for each stimulus type. Variance estimation was accomplished with restricted maximum likelihood (REML) estimation. Post-Hoc pairwise comparisons between stimulus types were Bonferroni corrected. Data were analyzed in Minitab 19.2020 [46].

Scenario 1 (Hallway): The effect of stimulus type was significant, $F_{(2, 12)} = 37.09$, $p < 0.001$, where amplitudes for standard stimuli ($M = 0.826$ uV) were significantly lower than rare threat ($M = 1.828$ uV) and rare no threat ($M = 1.713$ uV) stimuli (see Figure 8). No significant difference was found between rare stimulus types. The topographical maps for the Hallway scenario can be found in the top row of Figure 11, and a waveform from a representative participant can be found in Figure 12a.

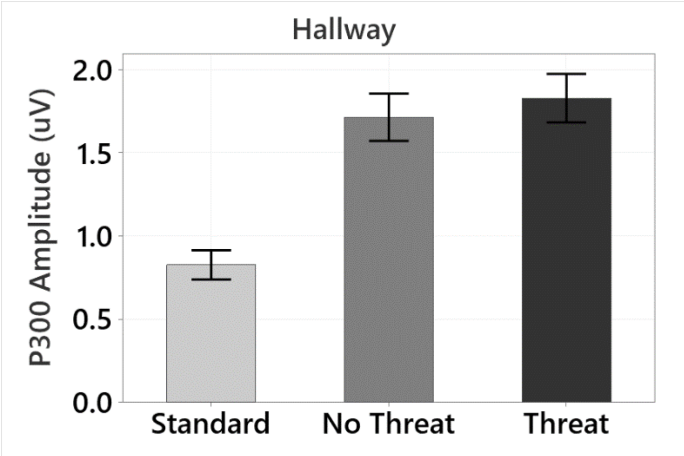

Fig. 8. Average mean amplitude for the Hallway scenario P300 for Standard, Rare No Threat, and Rare Threat stimuli. Rare stimuli showed significantly greater amplitude compared to standard, but no difference between rare stimulus types. Error bars = +\- 1 SEM.

Scenario 2 (Parking Lot): The effect of stimulus type was significant, $F_{(2, 12)} = 17.80$, $p < 0.001$. P300 amplitude for standard stimuli ($M = 0.777$ uV) were significantly lower than rare no threat ($M = 1.385$ uV) or rare threat ($M = 1.883$ uV) stimuli (see Figure 9). A trend effect ($t_{(12)} = 2.68$, $p = 0.06$) was found between rare stimulus types, where amplitude for threat stimuli was larger than no threat stimuli. The topographical maps for the Parking Lot scenario can be found in the middle row of Figure 11, and a waveform from a representative participant can be found in Figure 12b.

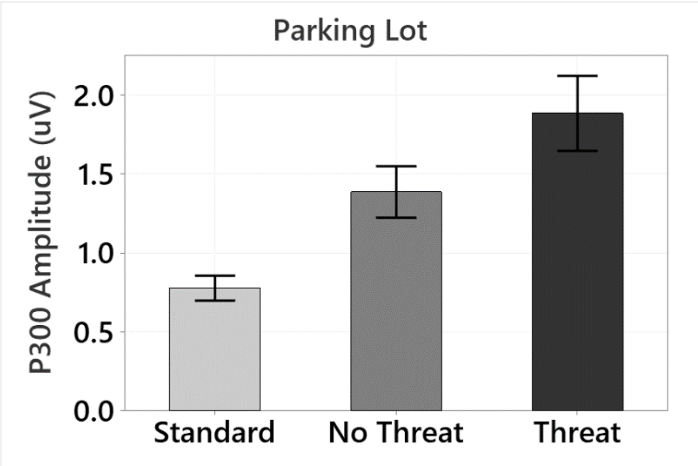

Fig. 9. Average mean amplitude for the Parking Lot scenario P300 for Standard, Rare No Threat, and Rare Threat stimuli. Rare stimuli showed significantly greater amplitude compared to standard, and a trend effect (p = 0.06) was found between rare stimulus types. Error bars = +\- 1 SEM.

Scenario 3 (Fence): The effect of stimulus type was significant, $F_{(2, 12)} = 31.15$, $p < 0.001$. P300 amplitude was significantly lower for standard stimuli (M = 0.917 uV), than rare no threat (M = 1.702 uV) or rare threat (M = 1.559 uV) stimuli (see Figure 10). No significant difference was found between rare stimulus types. The topographical maps for the Fence scenario can be found in the bottom row of Figure 11, and a waveform from a representative participant can be found in Figure 12c.

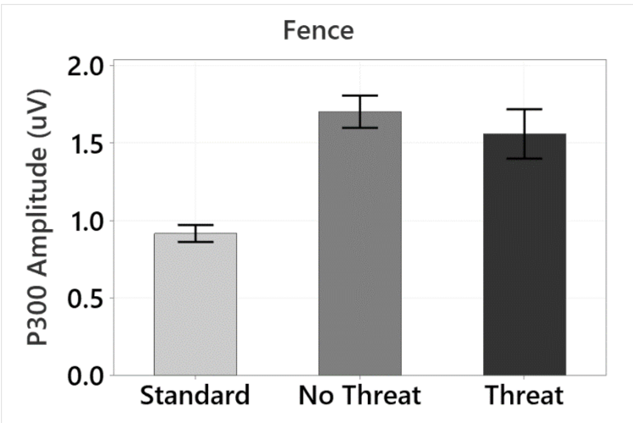

Fig. 10. Average mean amplitude for the Fence scenario P300 for Standard, Rare No Threat, and Rare Threat stimuli. Rare stimuli showed significantly greater amplitude compared to standard, but no difference between rare stimulus types. Error bars = +\- 1 SEM.

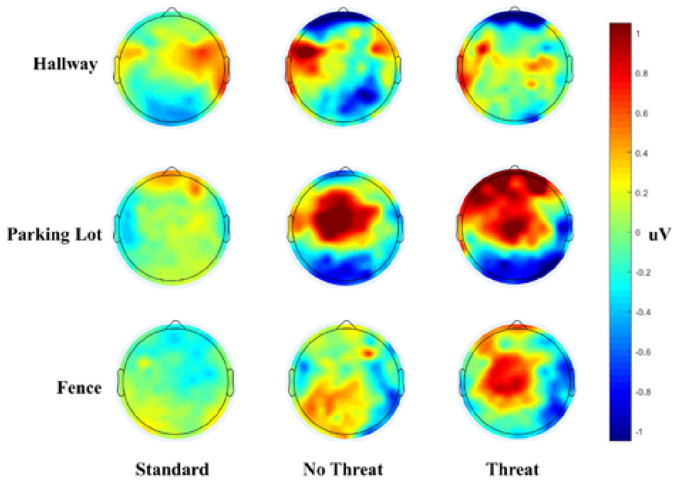

Fig. 11. Topographical maps for each scenario and stimulus type. Maps were derived by taking the mean peak amplitude across participants for each stimulus type in each scenario occurring between 250-500ms post-stimulus. Note in all three scenarios, there is a much larger response to both rare (threat and no threat) stimuli compared to the standard stimuli. Specifically, for the Parking Lot scenario, a trend level effect was observed between rare no threat and rare threat stimuli, where a stronger response was observed for threat stimuli. The scale is +/- 1 uV.

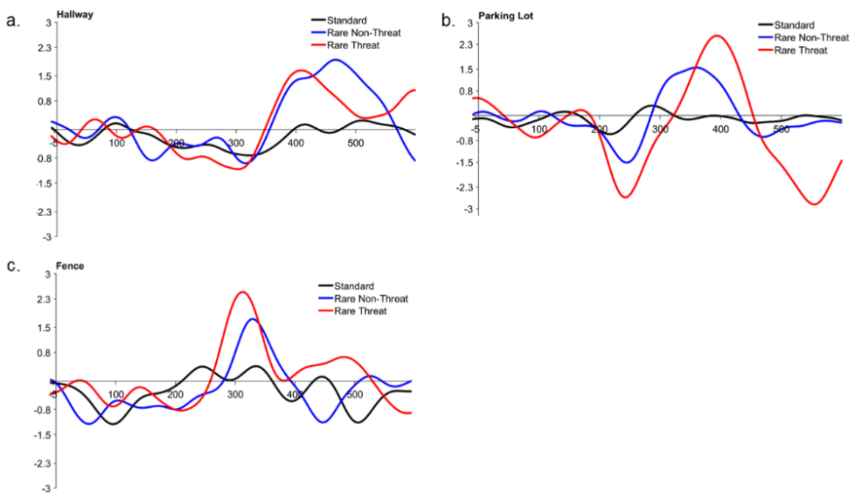

Fig. 12. Waveforms for a representative participant and channel for the Hallway Scenario (a), Fence Scenario (b), and Parking Lot Scenario (c). P300 Responses to standard stimuli are in black, responses to rare non-threat stimuli are in blue, and responses to rare threat stimuli are in red. ERPs were filtered (infinite impulse response) from 1-10Hz for visualization.

4 Discussion

Creation of streaming video stimuli that allow for strict experimental control while providing ease of scene manipulation is difficult to achieve. In this effort, we propose leveraging video game modding tools as a method of creating research quality stimuli. To this end, three steaming video scenarios were created, following the three-stimulus oddball paradigm of frequent non-targets, rare non-targets, and rare-targets [34], with the targets and non-targets having equal probability. In the current design, participants saw frequent non-targets, and two types of rare target stimuli (threat and non-threat). These scenarios were designed to mimic situations that operators monitoring video surveillance feeds for physical security purposes might experience. In Scenario 1 (Hallway; scientists entering and exiting a facility) location information was sufficient to categorize stimuli. In Scenario 2 (Parking Lot; convenience store patrons passing through a parking lot) participants had to monitor a particular location for a particular action by a character model in order to distinguish between target and non-target stimuli. In Scenario 3 (Fence; security personnel and civilians passing alongside a fence within which military exercises were occurring), participants had to monitor a particular location (a gate) for a particular action (pausing or continuing to walk without stopping) that was only considered a threat if performed by a certain character model. In all scenarios we found evidence of a P300 for comparisons between standard stimuli and both types of rare stimuli (threat and non-threat), though the amplitudes observed were on average smaller than other studies.

The amplitude of the P300 can be influenced by numerous variables, including age, where a reduction in amplitude was observed as participants advanced in age [47], as well as attentional load (the tasks in the current experiment likely require a higher degree of attention than more traditional P300 experiments which could lead to reduction in amplitude; see [48]). In addition, the P300 is known to exhibit significant inter-individual variability in amplitude [49]. Any of these factors, and likely others, could have contributed to the lower-than-average amplitudes observed in the current experiment. Though this finding suggests that the proposed method of stimuli creation results in sufficient experimental control to allow for ERP analysis, there are several limitations to this pilot work that should be addressed in subsequent efforts. As a pilot study emphasizing the method of stimuli creation, the number of participants is quite small and therefore results should be interpreted with caution.

For Scenario 2 (Parking Lot), there was an additional finding of a trend effect for a larger P300 when comparing rare threat stimuli to rare no-threat stimuli. This is consistent with a prior finding in the context of a three-stimulus oddball paradigm in which targets elicited a larger P300 than non-targets when probability of occurrence was equal [45]. It is possible that a similar effect was not observed in the other two scenarios due to the ambiguous nature of the so-called "threatening" stimuli; a model crouched behind a vehicle may be more clearly identified as suspicious than a person exiting a building quickly or the impact of a character model pausing or not pausing at a gated entrance. Future research could benefit from inclusion of ratings of stimuli to determine threat ambiguity and intensity.

Prior work has established that P300 amplitude is inversely related to stimulus probability [45]. In the current work, we did not vary the probability of occurrence for the different stimuli types, and the probability of targets and non-targets was equal. Other factors found to influence P300 amplitude, such as the number of non-targets preceding a target and the target-to-target interval [50] were also not systematically manipulated in this study. This is notable because while this method *could* theoretically produce scenarios that are ecologically valid, the target-to-target intervals, number of targets, and ratio of targets to non-targets used in this pilot study were almost certainly not an accurate representation of what an analyst monitoring a security feed experiences in the world. While the current study likely overrepresented number of targets within a brief time frame due to time considerations, the P300 amplitude has previously been found to be sensitive to level of fatigue, and may therefore serve as an indication of flagging attention in tasks that require sustained attention [38]. Our method of creating streaming video stimuli could therefore be used to emulate the number and types of video feeds that operators in domains such as physical security may be exposed to and thereby characterize the time course of fatigue onset.

Additionally, Garry's Mod was released in 2006, making it a relatively old game with dated graphics at the time of this study. Over time, video game realism has increased via improved graphics and game engines more capable of paralleling human mental representations [51]. With advances in virtual reality gaming tools, increasing levels of immersion may be possible and offer much greater ecological validity as the ability to approximate realistic environments increases. Therefore, this pilot study does not constitute an endorsement of Garry's Mod in particular, and does not include a direct comparison to other modding tools or methods of creating streaming stimuli. Popular games such as Roblox and Minecraft offer modding capabilities, but these were deemed not realistic enough for the physical security scenarios desired in the current study. While modding may be possible for a number of games, support and accessibility might differ dramatically (e.g., Sims 4, while offering powerful modding abilities, is sometimes derided by community members as lacking support from the developer [52]. Other tools should therefore be evaluated in future work to determine their pros and cons and allow researchers to select the optimal tool for their needs.

It is also worth noting that creating video stimuli using video game tools can range from simple to complex contingent on the tool selected and the nature of the desired scenario. There are four main methods of creating machinima [23]. From most simple to most difficult these are: using the game's inherent AI to control actions, digital puppetry (capturing the manual manipulation of digital characters/objects – "playing" the scenario), recamming (adjusting camera locations), and precise scripting of actions. It is likely that the most difficult method – precise scripting of actions – is necessary if the study goal involves the timing necessary for ERP analysis. Another option, as per [16], is to take an existing scenario and adjust it to suit the goals of the research. For instance, adjustment of the character models used in the scenarios described in the current work may be accomplished via a trivial replacement of filenames that could be performed in a manner of minutes, allowing adjustment along dimensions such as the size, gender, and skin color of character models.

Regarding analysis considerations, often P300 paradigms ask participants to fixate their gaze on the center of a screen while static stimuli are presented, allowing researchers to be relatively certain about what participants are looking at and when [53]. However, when using streaming stimuli with the possibility of co-occurrence of different stimuli types as in the current study, researchers may know when certain events are presented by virtue of having scripted the timing of those events, but it may not be clear where participants are directing their gaze at a given time. To this end, in the current study P300 amplitude values were calculated as the peak amplitude during a latency range of 250ms to 500ms [45] following an event of interest. This assumes that the highest amplitude wave in this timeframe reflects participant processing of the event of interest (e.g., a character model ducking near a car in the Parking Lot scenario), but without eye tracking data to verify gaze location this remains an assumption. The addition of eye tracking data would allow for calculation of fixation-related potentials that clearly time-lock neuroimaging data to timepoints during which participants were looking at events of interest [53] and is recommended as a future direction for purposes of verifying data quality.

As a general limitation of creating video stimuli that reflect real-world circumstances, allowing for the timing of events to play out (e.g., having character models walk over distances) results in a reduced number of trials over a given timeframe relative to presentation of simple static stimuli that may be flashed on a screen for short durations. This could be mitigated by a longer experimental session (which may risk fatiguing a participant) to accumulate a reasonable number of trials, or by running a greater number of participants (note that the number of participants in the current study is quite small). While these factors may be difficult for researchers to work around, they are necessary if one wishes to study human behavior and brain responses using stimuli that closer approximate real-world situations.

5 Conclusion

Mutual interest between game developers and game players in ease of modding has created a situation that may be leveraged by researchers interested in creating video stimuli. Here we demonstrate the feasibility of using existing modding tools to create video scenarios amenable to neuroimaging analysis. The results indicate that sufficient levels of experimental control may be achieved to allow for the precise timing required for ERP analysis. The variety of tools available allows for a vast range of video stimuli to be created using this method, which also allows for relative ease of adjustment. While the emphasis of the current study was on scenarios relevant to physical security, the general method of stimuli creations could be implemented by researchers operating in a wide variety of domains. This method of creating realistic streaming stimuli may assist in adding to the body of literature representing research with stimuli in motion [54, 55].

Acknowledgements

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. This work was funded by Sandia National Laboratories' Laboratory-Directed Research and Development program. SANDXXXX-XXXX X. *This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government*.

References

1. Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77-84.
2. Eziolisa, O. N. (2014). Investigation of capabilities of observers in a watch window study. (Master of Science in Engineering), Wright State University, Dayton, OH. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=wright1401889055
3. Irvine, J. M., Fenimore, C., Cannon, D., Haverkamp, D., Roberts, J., Israel, S. A., ... & Brennan, M. (2006, May). Development of a motion imagery quality metric. In *Proceedings of the American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Meeting* (pp. 1-5).
4. Sulman, N., Sanocki, T., Goldgof, D., & Kasturi, R. (2008, December). How effective is human video surveillance performance? In *2008 19th International Conference on Pattern Recognition* (pp. 1-3). IEEE.
5. Steel, K., Ellem, E., & Baxter, D. (2015). The application of biological motion research: biometrics, sport, and the military. *Psychonomic Bulletin & Review*, *22*(1), 78-87.
6. Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1-8.
7. Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, *13*(4), 993-1018.
8. Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772-775.
9. Green, C. S., & Seitz, A. R. (2015). The impacts of video games on cognition (and how the government can guide the industry). *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 101-110.
10. Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92-104.
11. Toril, P., Reales, J. M., & Ballesteros, S. (2014). Video game training enhances cognition of older adults: a meta-analytic study. *Psychology and Aging*, *29*(3), 706-716.
12. Bostan, B., & Kaplancali, U. (2010). Explorations in player motivations: Game mods. In *Proceedings of GAMEON-ASIA*.
13. Kücklich, J. (2005). Precarious playbour: Modders and the digital games industry. *Fibreculture*, *5*(1), 1-5.
14. Hofman-Kohlmeyer, M. (2019). Players as a prosumer. Individual motives for engaging in game modding. *World Scientific News*, *133*, 191-203.

15. Sotamaa, O. (2010). When the game is not enough: Motivations and practices among computer game modding culture. *Games and Culture*, *5*(3), 239-255.
16. Wallace, R. (2014). Modding: Amateur authorship and how the video game industry is actually getting it right. *BYu L. Rev, 219.*, 219-256.
17. Bostan, B. (2005). Game Modding and TES: New Way to Design Virtual Worlds, 3rd International Symposium of Interactive Media Design, Yeditepe University, Turkey.
18. Newman, J. (2008). *Playing with videogames*. Routledge.
19. McClarty, K. L., Orr, A., Frey, P. M., Dolan, R. P., Vassileva, V., & McVay, A. (2012). A literature review of gaming in education. *Gaming in Education*, 1-35.
20. Moshirnia, A. (2007). The educational potential of modified video games. *Issues in Informing Science & Information Technology*, *4,* 511-521.
21. Paul, N. & Hansen, K. (2006). Disaster at Harperville: The modding of Neverwinter Nights to teach journalism students the strategic steps in information gathering. In E. Pearson & P. Bohman (Eds.), *Proceedings of ED-MEDIA 2006--World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 1954-1959). Orlando, FL USA: Association for the Advancement of Computing in Education (AACE).
22. Mohseni, M. R., Liebold, B., & Pietschmann, D. (2015). Extensive modding for experimental game research. In P. Lankoski & S. Björk (Eds.), *Game Research Methods* (pp. 323-340). Pittsburgh, PA: ETC Press.
23. Kelland, M., Morris, D., & Lloyd, D. (2005). *Machinima: Making Movies in 3D Virtual Environments*. Cambridge: The Ilex Press.
24. Gawlik-Kobylińska, M. (2017). Machinima in Education for Security and Safety. *IJRET: International Journal of Research in Engineering and Technology*, *6*, 42-46.
25. Fabiani, M., Gratton, G., & Coles, M. G. H. (2000). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 53–84). Cambridge, England: Cambridge University Press.
26. Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465-476.
27. 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.
28. 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.
29. Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251-257.
30. Hoffman, J. E., Simons, R. F., & Houck, M. R. (1983). Event-related potentials during controlled and automatic target detection. *Psychophysiology*, *20*(6), 625-632.
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), 67–82.
32. 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.
33. 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), 307-323.

34. Katayama, J. I., & Polich, J. (1999). Auditory and visual P300 topography from a 3 stimulus paradigm. *Clinical Neurophysiology*, *110*(3), 463-468.
35. Facepunch Studies, Inc. [Garry's Mod]. (2006). Retrieved from https://gmod.facepunch.com/
36. Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, *21*(9), 649-665.
37. Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
38. Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, *41*(2), 103-146.
39. Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://www.pstnet.com.
40. Yuan, J., He, Y., Qinglin, Z., Chen, A., & Li, H. (2008). Gender differences in behavioral inhibitory control: ERP evidence from a two-choice oddball task. *Psychophysiology*, *45*(6), 986-993.
41. Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.
42. Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods, 192*(1), 152-162.
43. MATLAB. (2017). version 9.3.0.713579 (R2017b). Natick, Massachusetts: The MathWorks Inc.
44. Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213.
45. Katayama, J. I., & Polich, J. (1996). P300, probability, and the three-tone paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *100*(6), 555-562.
46. Minitab 19 Statistical Software (2020). [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com).
47. van Dinteren, R., Arns, M., Jongsma, M. L., & Kessels, R. P. (2014). P300 development across the lifespan: a systematic review and meta-analysis. *PloS One*, 9(2), e87347.
48. Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998-1003.
49. Li, F., Tao, Q., Peng, W., Zhang, T., Si, Y., Zhang, Y., ... & Xu, P. (2020). Inter-subject P300 variability relates to the efficiency of brain networks reconfigured from resting-to task-state: evidence from a simultaneous event-related EEG-fMRI study. *NeuroImage*, *205*, 116285.
50. Gonsalvez, C. J., & Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology*, *39*(3), 388-396.
51. Barlett, C. P., & Rodeheffer, C. (2009). Effects of realism on extended violent and nonviolent video game play on aggressive thoughts, feelings, and physiological arousal. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *35*(3), 213-224. 10.1002/ab.20279
52. Hanabi, J. (2020, September 20). The Sims 4 modern Python modding. Gitconnected. https://levelup.gitconnected.com/the-sims-4-modern-python-modding-part-1-setup-83d1a100c5f6
53. Brouwer, A. M., Reuderink, B., Vincent, J., van Gerven, M. A., & van Erp, J. B. (2013). Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision*, *13*(3), 17-17.

54. 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.
55. 55. Steel, K. A., Baxter, D., Dogramaci, S., Cobley, S., & Ellem, E. (2016). Can biological motion research provide insight on how to reduce friendly fire incidents? *Psychonomic Bulletin & Review, 23*(5), 1429-1439.