Impact of Virtual Environment Fidelity on Behavioral and Neurophysiological Response

Anna Skinner	Chris Berka	Lindsay Ohara-Long	Marc Sebrechts
AnthroTronix, Inc.	Advanced Brain Monitoring, Inc.	Clemson University	The Catholic University of America
Silver Spring, MD	Carlsbad, CA	Clemson, SC	Washington, DC
askinner@atinc.com	chris@b-alert.com	lindsay.ohara.long@gmail.com	sebrechts@cua.edu

ABSTRACT

Visual fidelity is a critical element in designing cost-effective training simulations. Methods for quantifying the impact of fidelity components such as resolution on the training experience could optimize cost-benefit analyses in the development of simulated training environments. An ongoing research effort seeks to demonstrate the technical feasibility of a Perceptually-informed Virtual Environment (PerceiVE) Design Tool, capable of using operator behavior and physiology to provide a novel approach to simulation design and assessment. This study reports the results of a basic research experiment examining neural signatures based on event-related potentials (ERPs) that vary as a function of stimulus resolution and are related to performance in a militarily-relevant simulation training task involving vehicle classification. The results of this study demonstrate that ERPs varied across four classes of vehicles and were sensitive to changes in the fidelity of the vehicles within the simulated task environment. While performance, measured by accuracy and reaction times, distinguished between the various stimulus resolution levels and between classes of vehicles, the ERPs further highlighted interactions between resolution and class of vehicle, revealing subtle but critical aspects affecting the perceptual discrimination for the vehicles within the training environment. The distinctive ERP signatures offer a method to characterize objects within military training scenarios that required higher resolution for effective training, as well as those that could be easily recognized at lower resolutions, thus saving developers time and money by highlighting the most efficient requirements to achieve training efficacy. The ERPs can be measured unobtrusively during training, allowing developers to access a metric that could be used to guide scenario development without requiring repeated transfer of training assessments and without relying solely on performance or subjective responses. This novel approach could potentially be used to determine which aspects of VE fidelity will have the highest impact on transfer of training with the lowest development costs for a variety of simulated task environments.

ABOUT THE AUTHORS

Anna Skinner is a Biomedical Engineer at AnthroTronix, Inc. Ms. Skinner has led a number of research efforts in the areas of human-computer interaction, human-robot interaction, virtual training environments, and augmented cognition. Ms. Skinner received her Bachelor's degree in Biomedical Engineering from The Catholic University of America (CUA) in 2001, and is currently enrolled in the Applied Experimental Psychology PhD program at CUA.

Chris Berka is the CEO and Co-Founder of Advanced Brain Monitoring, Inc., and has 25 years experience managing research, development, and commercialization of new technologies. She co-invented fifteen patented/patent-pending systems, served as PI for $20million NIH/DoD awards, authored 50 EEG/cognition papers, received her B.A. with distinction in Psychology/Biology at Ohio State, and completed graduate studies in Neuroscience at UCSD

Lindsay O'Hara-Long is a PhD student in Human Factors Psychology at Clemson University. While at Clemson, Lindsay has worked with AnthroTronix, Inc. on various projects related to virtual environments and simulator fidelity. Her current research includes enhancing teleoperator perception.

Marc Sebrechts, PhD is Professor of Psychology and Department Chair at The Catholic University of America. His research examines the role of technology in perception and learning. He is founding director of the Cognition and Virtual Reality Laboratory, where he investigates how virtual environments can improve our understanding of spatial cognition.

Impact of Virtual Environment Fidelity on Behavioral and Neurophysiological Response

Anna Skinner	Chris Berka	Lindsay Ohara-Long	Marc Sebrechts
AnthroTronix, Inc.	Advanced Brain Monitoring, Inc.	Clemson University	The Catholic University of America
Silver Spring, MD	Carlsbad, CA	Clemson, SC	Washington, DC
askinner@atinc.com	chris@b-alert.com	lindsay.ohara.long@gmail.com	sebrechts@cua.edu

BACKGROUND

Virtual environments (VE's) and simulations are currently employed for training across a wide variety of domains, both military and civilian. Technological advances are enhancing the ability of developers to create VE's with visual, auditory, haptic, and even olfactory realism. Such VE's allow the military to train skills that are too costly, too dangerous, or otherwise impractical to rehearse in real world operational environments. While significant research has been conducted examining the transfer of training from VEs to real world tasks (for example, Lathan, Tracey, Sebrechts, Clawson, & Higgins 2002; Sebrechts, Lathan, Clawson, Miller, & Trepagnier, 2003), objective metrics of transfer are limited. The common assumption is that the more realistic the VE (i.e., the higher the fidelity), the better the transfer of training to real world tasks. However, fidelity components (e.g., display resolution, frame rate, texture mapping, physics modeling, etc.) may vary in their importance for transfer of training within a given task or domain.

One of the primary questions that simulation designers must address is, "what components of fidelity have the greatest impact on transfer of training?" Fidelity is defined as the degree to which features (e.g., visual, auditory, etc) in the VE match features in the real environment. Assuming that matching the real world optimizes transfer, one can argue that maximizing VE fidelity would result in transfer of training equivalent to real-world training since, in the limit case, the two environments would be impossible to differentiate (Waller, Hunt & Knapp, 1998; Martin, 1981). However, developers are limited by practical restrictions such as cost, time, and development resources. Thus, trade-offs are necessary. However, there is currently a limited understanding of the specific trade-offs between increases in simulation fidelity and operator behavior, and essentially no guarantee to developers that a particular level/area of simulation fidelity is sufficient to provide effective transfer of training.

Fidelity requirements have traditionally been determined by performance measurements compared before and after design iterations. With each design modification, end users are tested using the VE and their performance is compared to performance on the prior VE design. Improved performance is assumed to be related to improved design and fidelity. However, it is often difficult to identify the specific design components that directly relate to transfer of training improvements. Furthermore, this method of design focuses on trial and error, and is therefore time consuming, undirected, and may result in false associations between performance and VE characteristics. Thus, a more comprehensive and objective assessment of the quality of interaction with a VE is needed to effectively identify the specific components of simulation that bear relevance to real world operational tasks.

Vice, Lathan, Lockerd, and Hitt (2007) outlined a novel approach to determining fidelity simulation design using neurophysiological measures. Vice et al hypothesized that a physiologically-based system capable of detecting changes in operator behavior and physiology throughout a VE experience and comparing those changes to operator behavior and physiology in real-world tasks, could potentially determine which aspects of VE fidelity will have the highest impact on transfer of training.

Electroencephalogram (EEG) and event related potential (ERP) approaches offer excellent temporal resolution for tracking of neural activity representing the flow of information from sensory processing, detection and identification of relevant objects and decision-making. ERP signature components associated with the identification of target stimuli were first reported in 1965 and named "P300s or P3b or Late Positivity" (for example, Sutton, Braren, Zubin, & John, 1965) because target stimulus presentations are associated with large positive potentials maximal over the parietal cortex with peak latencies ranging from 300-800 ms after presentation of the target stimulus.

The P300 is generally accepted to be a post-sensory signal elicited when subjects attend and respond to target stimuli and is believed to be related to higher cognitive processes including updating working memory (Donchin & Coles, 1988). Several reports suggest that when target stimuli are degraded, obscured or difficult to recognize, the amplitude of the P300 is decreased (Kok, 1985, Verleger, 1988). In addition to the extensive work on describing the P300, a growing body of ERP evidence reveals ERP neural signatures of target recognition and discrimination as early as 150-200 milliseconds post-stimulus (Thorpe, 1996).

A pilot study was conducted (Skinner, Vice, Lathan, Fidopiastis, Berka & Sebrechts, 2009) in which variations in the fidelity of the stimuli (high versus low polygon count) in a visual search/identification task did not result in performance changes; however, consistent and distinguishable differences were detected in ERP early and late components, including decreased amplitude p300 components in the low fidelity condition. The current study sought to expand upon this research, examining the effects on EEG response for a slightly more complex VE task in which performance is impacted significantly by changes in fidelity.

METHODS

A static, VE-based visual search task was developed using 3D models of vehicles from the Virtual Battlespace 2 (VBS2) simulation platform. Stimuli included 4 vehicle types (car, SUV, truck, van), 3 colors (red, blue, white), 8 orientations (north, northeast, east, southeast, south, southwest, west, northwest), 3 sizes (emulating effects of distance from target), and 3 levels of fidelity (resolution, based on pixels/inch). A variety of sample stimuli are shown in Figure 1.

Figure 1. Sample stimuli representing a variety of the vehicles, orientations, colors, sizes, and fidelities.

Individual task stimuli consisted of one vehicle presented in the center of the screen for 250ms, as shown in Figure 2. A blank screen was then presented during a response period of up to 2 seconds. Participants were instructed to indicate vehicle type (car, SUV, truck, or van) using the arrow keys on the keyboard. An inter-stimulus interval of 1 second was provided, and a visual plus-sign cue was displayed in the center of the screen for 250ms immediately preceding each new stimulus.

Figure 2. Sample low fidelity stimulus.

At the start of the experiment, participants were informed that they would be shown a series of graphics of varying detail. Individual images were displayed in random order on a 19-inch monitor. Participants were positioned 30 inches from the display. A total of 25 participants each performed two 10-minute consecutive sessions, consisting of 432 stimuli presentation trials each. Trials in which the participant did not provide a response within the 2-second response period were considered missed trials.

Neurophysiological response was assessed using electroencephalogram (EEG) measures acquired on the surface of the scalp. The B-Alert® wireless Sensor Headset from Advanced Brian Monitoring (ABM) was used to acquire EEG data from 9 sites (F3, F4, C3, C4, P3, P4, Fz, Cz, and POz), referenced to linked-mastoids (A1 and A2), as shown in Figure 3. A DLL was implemented to allow the EEG signal to be synchronized with the task stimuli, which were presented within a custom program using Psychology Software Tools' E-Prime® experiment management software.

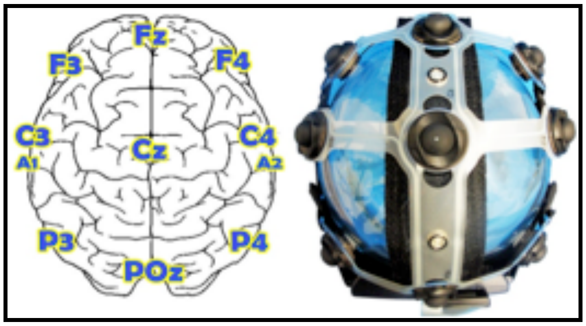

Figure 3. EEG electrode placement and cap

The independent variables for this task included fidelity level, vehicle type, orientation, color, and size. The dependent variables included the neurophysiological response data, as well as the task performance data (i.e., accuracy and reaction time).

RESULTS

Performance and neurophysiological response data were successfully collected for 25 participants. However, the EEG data for 6 participants contained excessive eye movement artifact; the data for these 6 participants was not included in the performance or neurophysiological response analysis. The following provides a summary of the data analysis and results for the remaining 19 participants.

Performance Results

Analysis of incorrect trials for each condition revealed which of the independent variables contributed most to behavioral performance. The overall percentage and distribution of incorrect trials was analyzed using a series of repeated measures Analysis of Variance (ANOVA) computations. These analyses revealed main effects for fidelity (F=5.62, p=.01), vehicle type (F=57.36, p<.01), and orientation (F=115.06, p<.01). The distribution of incorrect trials across vehicle type and fidelity is shown in Figure 4, though no significant interactions were found.

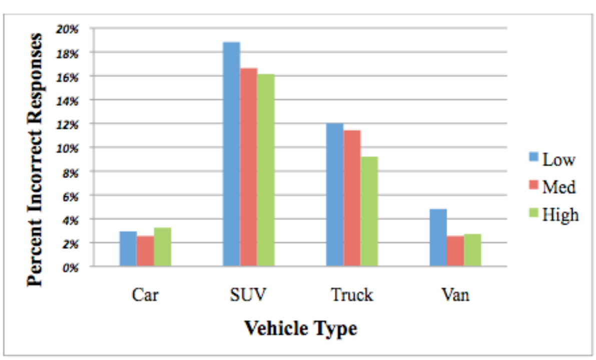

Figure 4. Incorrect trials distribution

No significant differences were found for the number of incorrect trials across the dimensions of size (F=1.02, p=.35) or color (F=1.35, p=.27). As a result, only fidelity, vehicle type, and orientation were considered in further analyses of the correct trials.

Only correct trials were used to analyze reaction time. Median reaction time values were used to control for outliers and non-normal distributions typical of reaction time data. To assess the effects of stimulus fidelity, vehicle type, and orientation on median reaction time, a 3 (fidelity level) x 4 (vehicle type) x 8 (orientation) repeated measures ANOVA was conducted. A main effect was found for fidelity level, F (2, 36) = 14.07, p<.001; high and medium fidelity stimuli resulted in significantly lower reaction times than low fidelity stimuli (p<.005), although medium fidelity reaction times were closer to those found for low fidelity for SUV stimuli. A main effect was also found for vehicle type, F (3,54) = 15.72, p<.001. Car stimuli resulted in significantly lower reaction times than any other vehicle (p<.001 for all individual comparisons). In contrast, SUVs and trucks resulted in significantly higher reaction times than cars (p<.01) or vans (p<.05. There was no main effect for orientation, and no significant interaction between any factors. Figure 5 displays the median reaction times for each fidelity level by vehicle type, averaged across all 19 participants.

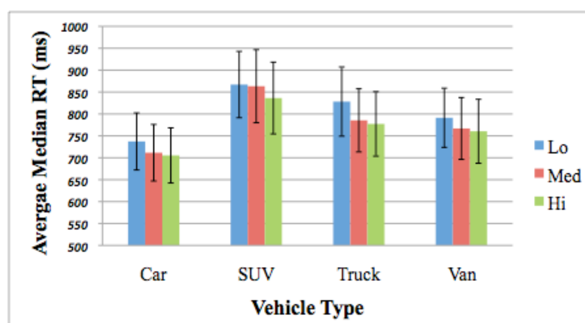

Figure 5. Reaction times by fidelity level and vehicle type

In an effort to provide an equivalent analysis to the neurophysiological data assessment, which is based on grand means across trials, fidelity level and vehicle type were also analyzed using mean reaction times rather than medians. A 3 (fidelity level) x 4 (vehicle type) x 8 (orientation) repeated measures ANOVA was conducted and supported results using median reaction times. A main effect was found for fidelity, F (2, 36) = 11.88, p <.001, such that high and medium fidelity stimuli both resulted in significantly lower reaction times than low fidelity stimuli (p<.005). There was also a main effect for vehicle type, F (3, 54) = 13.19, *p*<.001. Car stimuli resulted in the lowest reaction times, followed by vans and trucks; SUV stimuli resulted in the longest reaction times. There was no main effect for vehicle orientation, nor any significant interactions between the three factors.

Thus, the performance results revealed significant effects for fidelity, vehicle type, and orientation for accuracy (% correct), though only effects of fidelity and vehicle type were found for reaction time (for correct trials); however, no significant interaction effects were found across these stimulus characteristics.

Neurophysiological Results

EEG data were analyzed for the 19 participants that did not display excessive eye-movement artifact. One-and-a-half second stimulus-locked Event Related Potentials (ERPs) were averaged across all participants to analyze the differences between fidelity & stimulus type. Figure 6 provides the grand mean ERP data at all 9 electrode sites for the low, medium, and high fidelity conditions averaged across all 19 participants. Fidelity differentially impacted ERPs for the low fidelity condition with effects similar to those observed in the reaction time results; Low fidelity stimulus ERPs showed greater early positivity than the medium and high fidelity ERPs in the range of 200ms-650ms at all sites. The ERP differences as a function of fidelity were maximal for the frontal-central regions and more distinct over the left hemisphere than right.

ERP grand means were also calculated by vehicle type for all 19 participants. As shown in Figure 7, differentiation was demonstrated for the SUV stimuli from the remaining vehicle types, indicating that SUV stimuli had the highest amplitude positivity throughout the trials. This amplitude difference is maximal at approximately 400ms as well as between 600 and 1000ms post-stimulus onset.

Orientation was also examined for the correct ERP trials. Specifically, differentiation between canonical (i.e., East) representations of the vehicles was compared to the most difficult orientations (i.e., South) in which the vehicle is directly facing the observer head-on, providing fewer distinct visual cues. As shown in Figure 8, ERP grand means indicate that the most difficult orientation showed greater positivity at 200ms – 650ms. This finding may be a reflection of increased attentional and cognitive processing demands for the more difficult stimuli (Polich, 1997).

The SUV stimulus ERPs were observed to be the most distinct as compared to the other vehicle types. These results may reflect the challenge presented by the SUV class of vehicles as confirmed by the significantly higher percentage of incorrect SUV responses and longer reaction times in comparison to the other three vehicle types (see Figures 4 and 5). To further explore this class the SUV ERPs were analyzed in more detail. Figure 9 displays the SUV stimulus grand mean ERPs across all 19 participants for low, medium, and high fidelity conditions. High fidelity (shown in green) resulted in higher positivities than the low and medium fidelity conditions across most electrode sites with a maximal difference in the range of 500-800ms. This is contrary to the patterns observed for low, medium, and high fidelity ERPs when collapsed across all 4 vehicle types. These results, in combination with the performance, suggest that the difficult process of identification of the SUVs resulted in large positivities across all levels of fidelity. This finding may be in part due to the fact that there were fewer correct responses for the SUVs and thus few ERPs available for averaging. This may also reflect less confidence in responses.

These data demonstrate an interaction effect of vehicle and fidelity, which was not detected within the performance data analysis. Fidelity requirements within a virtual environment may differ across specific stimuli in order to achieve comparable physiological responses.

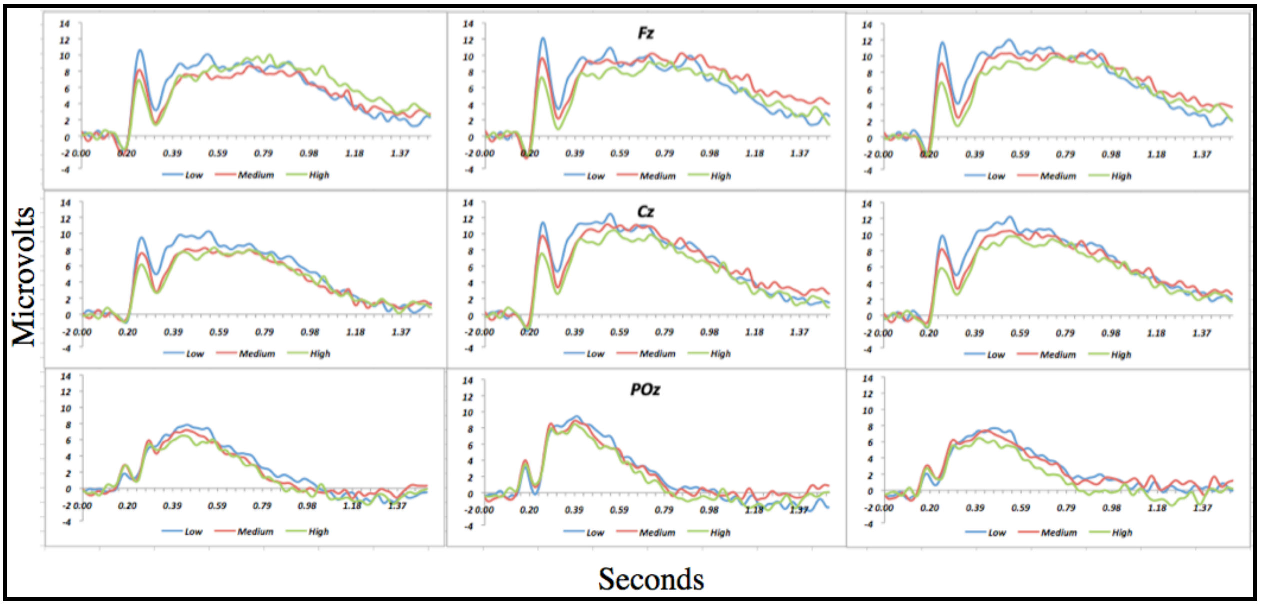

Figure 6. ERPs (9 electrode sites) for low, medium, and high fidelity conditions.

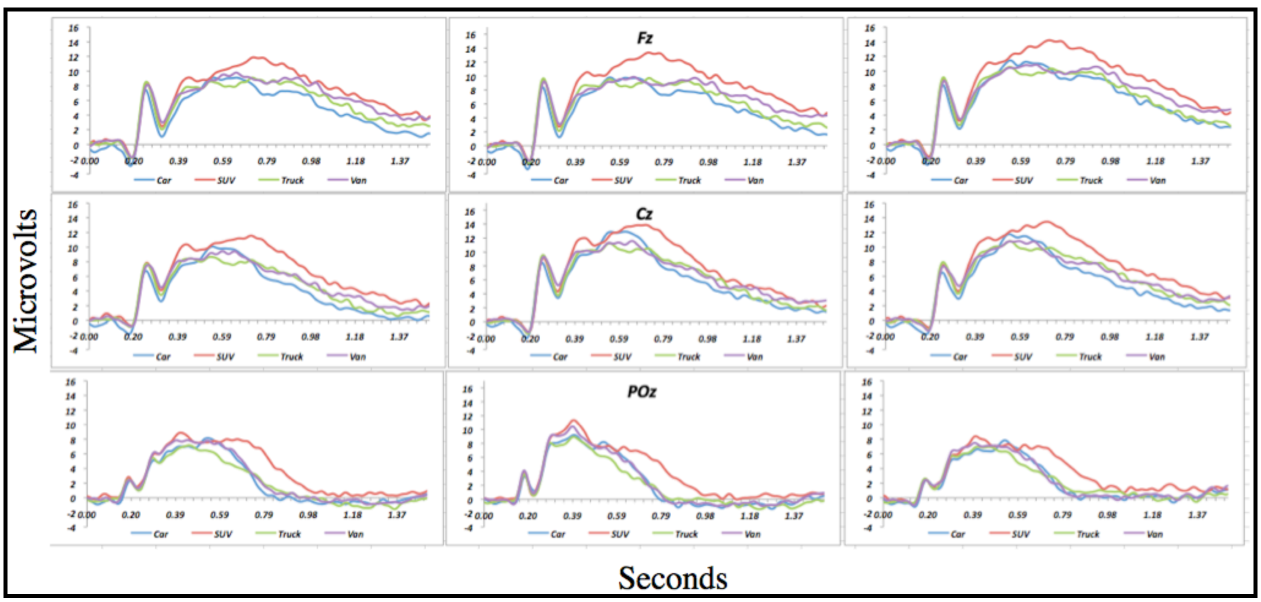

Figure 7. ERPs (9 electrode sites) for 4 vehicle types across all correct trials

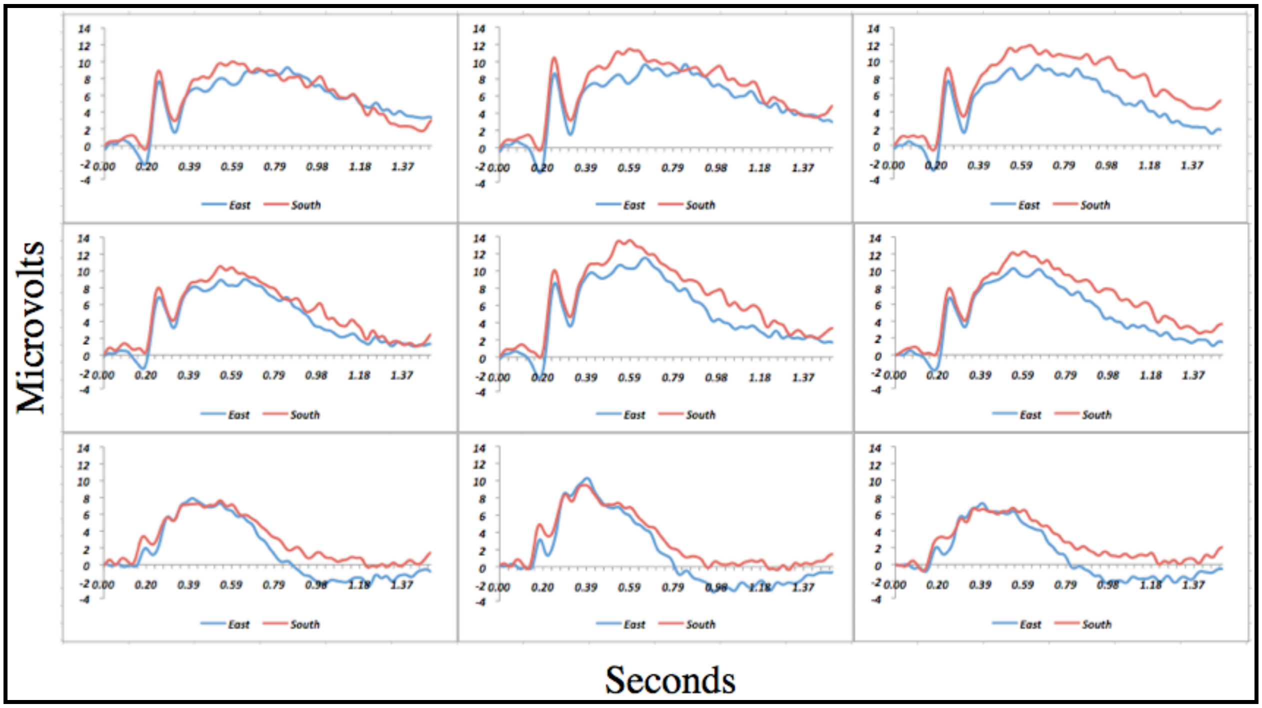

Figure 8. ERPs for head-on (South) versus canonical (East) orientation of vehicles for all correct trials

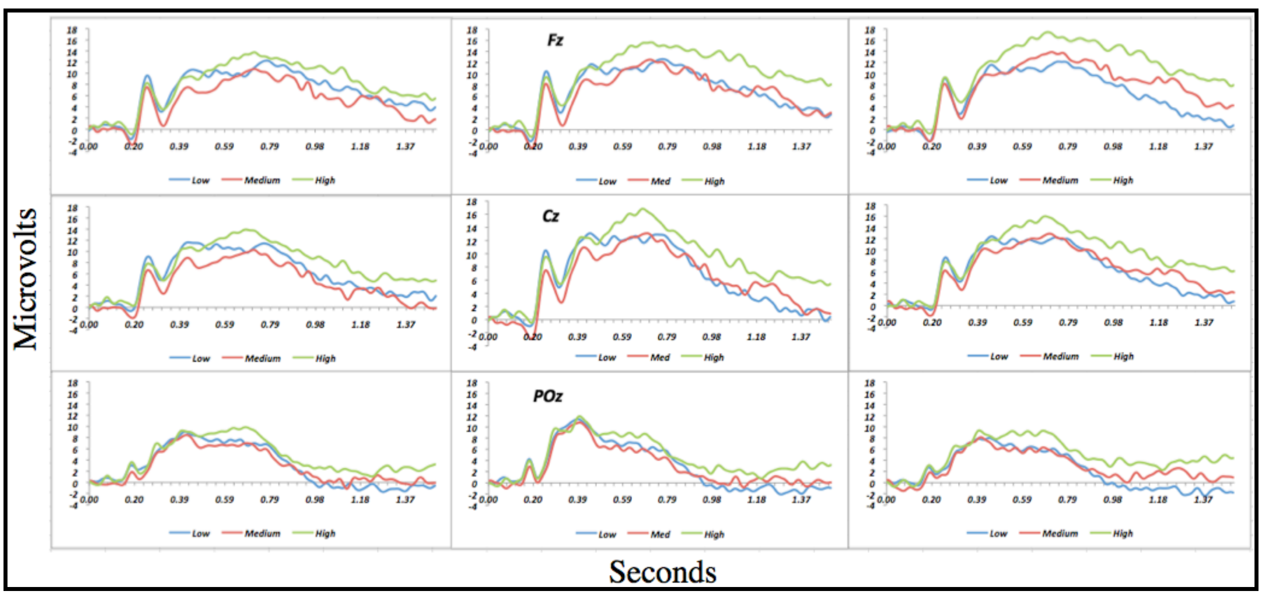

Figure 9. Low, medium, and high fidelity ERPs for correct trials of SUV stimuli

DISCUSSION

Based on these results, accurate identification of low, medium and high fidelity stimuli appear to elicit distinctive ERP components for different stimuli. Consistent and detectable differences in ERP data were observed for variations in fidelity level, vehicle type, and vehicle orientation. Furthermore, a distinct ERP profile was observed for SUV stimuli compared to the other vehicles. While performance, measured by accuracy and reaction times, distinguished between the various stimulus resolution levels and between classes of vehicles, the ERPs further highlighted interactions between resolution and class of vehicle, revealing subtle but critical aspects affecting the perceptual discrimination for the vehicles within the training environment. The distinctive ERP signatures offer a method to characterize objects within military training scenarios that required higher resolution for effective training, as well as those that could be easily recognized at lower resolutions, thus saving developers time and money by highlighting the most efficient requirements to achieve training efficacy. ERPs can be measured unobtrusively during training, allowing developers to access a metric that could be used to guide scenario development without requiring repeated transfer of training assessments and without relying solely on performance or subjective responses. This novel approach could potentially be used to determine which aspects of VE fidelity will have the highest impact on transfer of training with the lowest development costs for a variety of simulated task environments.

These findings will be leveraged under an ongoing research effort to develop and validate a perceptual skills VE task. Performance and neurophysiological data will be collected in both a real world and VE version of the task to further examine the technical feasibility of utilizing neurophysiological measures to assess fidelity design requirements in order to maximize cost-benefit tradeoffs and transfer of training.

ACKNOWLEDGEMENTS

This research was funded under an Office of Naval Research (ONR) Small Business Technology Transfer Research (STTR) grant.

REFERENCES

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences, 11, 357-374.

Kok, A. (1986). Effects of degradation of visual stimuli on components of the event related potential (ERP) in go/nogo reaction tasks. Biological Psychology. 23, 21-38.

Lathan, C.E., Tracey, M.R., Sebrechts, M.M., Clawson, D. M. & Higgins, G. (2002). Using virtual environments as training simulators: Measuring transfer. In K. Stanney (Ed.), Handbook of Virtual Environments: Design, Implementation, and Applications (pp. 403-414) Mahwah, NJ: Lawrence Erlbaum Associates.

Polich, J., Alexander, J.E., Bauer, L.O., Kuperman, S., Rohrbaugh, J., Morzorati, S., O'Connor, S.J., Porjesz, B. & Begleiter, H. (1997). P300 topography of amplitude/latency correlations. Brain Topography. 9: 275– 282.

Sebrechts, M.M., Lathan, C.E., Clawson, D.M., Miller, M.S. & Trepagnier, C. (2003). Transfer of training in virtual environments: issues for human performance. In L.J. Hettinger& M.W. Haas (Eds.), Virtual and Adaptive Environments; Applications, Implications, and Human Performance Issues (pp. 67-90) Mahwah, NJ: Lawrence Erlbaum Associates.

Skinner, A., Vice, J., Lathan, C., Fidopiastis, C., Berka, C., & Sebrechts, M. (2009). Perceptually-informed Virtual Environment (PerceiVE) Design Tool. In Lecture Notes in Artificial Intelligence (5638). Proceedings of the 5th International Conference on Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience: Held as Part of HCI International 2009. 650-657. Doi 10.1007/978-3-642-02812-0_74.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381, 520-522.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. Behavioral and Brain Sciences, 11, 343-427.

Vice, J.M., Lathan, C., Lockerd, A.D., and Hitt II, J.M. (2007). Simulation fidelity design Informed by physiologically-based measurement tools. Proceedings the Human Computer Interaction International Conference.

Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. Presence: Teleoperators and Virtual environments, 7(2), 129-143.