Human Interaction and Emerging Technologies (IHIET-AI 2024), Vol. 120, 2024, 464–470
https://doi.org/10.54941/ahfe1004598

Development of a System for Real-Time Assessment and Modification of Team-Based Exercises

Alan Liu[1] and Richard Wainess[2]

[1]Val G. Hemming Simulation Canter, Uniformed Services University, Bethesda MD, 20814, USA
[2]Richard Wainess LLC, USA

ABSTRACT

This project seeks to develop a Multimodal Performance Evaluation System (MPES), a system for real-time evaluation of individual and team stress levels suitable for a wide array of military medical training and assessment domains and conditions. This system was designed to identify changes in stress levels during training scenarios and determine whether changes to environmental or task stressors are needed. MPES capitalizes on the predictive abilities of pretest data (cognitive and affective) along with real-time physiological and sociometric data, to make claims about individual team member stress levels during training and aggregating that data into a team stress-level score. Our overarching goal was to improve the quality of training for medical teams, by adapting scenario-related characteristics to maximize stress levels, while avoiding overstressing to the point of performance failure. MPES is designed to be utilized in a variety of training environments of varying complexity, including screen-based, head-mount, immersive VR theatre, and open field exercises. Independent variables include heart rate, skin temperature, vocal characteristics. Moderating variables are working memory capacity and eight affective variables (e.g., extroversion and emotional stability). A Machine Learning approach using TPOT (a Python Automated Machine Learning Tool) was utilized for validation of the MPES analytic engine. Results indicate that the system is sensitive to individual differences and is capable of reliably reporting on perceived stress level.

Keywords: Immersive virtual environments, Medical team training, Performance assessment

INTRODUCTION

The research question we aimed to answer was what combination of data types, usable in live, virtual and face-to-face (F2F) team events, provide the most reliable prediction of a trainee's (and team's) stress level.

Key Take-aways – We seek to demonstrate that, through the use of machine learning and artificial intelligence, multiple data sources and types can be aggregated to provide sufficient evidence of real-time individual and team stress levels.

There is sufficient research evidence supporting the use of the various modalities behind our proposed model, including research on the use of biometric badges and sensors to identify cognitive and stress conditions, the use of personality tests to predict susceptibility to stress, the relationship between working memory capacity and stress, and the use of pre-event stress triggers to facilitate stress (Alamudun 2012, Baldasaro 2013, Hale 2013, Healey 2000, Kirschbaum, 1993, Schoofs 2009). What was missing was model and analytics to weight and aggregate the various data to reliably predict individuals' and teams' stress levels, in real time, so that the environment or task can be modified, based on those predictions. The goal of this project was to create that missing model and analytics and create a system (hardware and software) that is ready for large scale data gathering.

The personality factors selected for this project were the Mini-IPIP (a highly utilized and validated personality test that examines 5 personality factors, some of which have been shown to influence susceptibility to stress, such as extroversion) and three other personality factors that have been shown to impact susceptibility to stress (persistence/industriousness, sensation-seeking, and toughness).

Building the analytic model required the use of machine learning models and psychometric models to examine a variety of multimodal data sources (i.e., self-reports, biometric data, sociometric data, pretest cognitive skills) to achieve stable and reliable readings. All intended biometric and sociometric devices have been used in real-world settings, albeit maybe not as challenging as a combat environment we were designing for, but there did not seem to be any logical barriers to succeeding at sending relevant and reliable data to the analytic engine. The goal of this project was to develop and bring the analytic engine to a state of readiness for large scale field testing. Figure 1 depicts the flow of the various data types in our model.

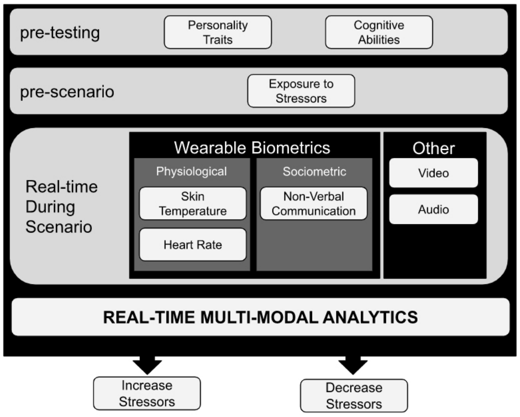

Figure 1: MPES model.

The most challenging task required real-time coding of telemetry and sophisticated approaches to real-time analysis and aggregation with other data. We leveraged work done at Imbellus, Inc. on real-time telemetric and biometric data analysis to develop a model optimized for adaptive task modification. Utilizing members of our development team, we tested the system, in real-time, with multiple participants and have validated that a reasonably accurate "stress level" score was achievable.

APPROACH

We combined a validated personality test (the Mini-IPIP) with items from several other affective tests to be applicable to military medical training. In total, 8 personality factors were selected for inclusion in MPES: extraversion, agreeableness, emotional stability, conscientiousness, openness, sensation-seeking, persistence/industriousness, and toughness.

We created a sensor array (hardware and software) to evaluate the reliability and accuracy of various wearable biometric and sociometric devices (heart rate, skin temperature, voice characteristics) under train-as-you-fight conditions.

We developed an analytic model to evaluate real-time data, weight both real-time data (physiological and sociometric) and pre-test data (working memory capacity and personality traits), and aggregate the various results to represent both individual and team states and determine if a scenario should be modified and in which direction (to increase stress or decrease stress).

For future field testing, prior to exposure to a training scenario, team member's working memory capacity are determined along with eight personality traits, the majority of which are known to be predictors of susceptibility to stress. During the training scenario, changes to heart rate, skin temperature, and vocal characteristics (e.g., pitch and cadence), all of which have been shown to be valid indicators of stress, are monitored and evaluated.

We designed an analytic engine that utilizes Machine Learning (ML) and Artificial Intelligence (AI) to train and improve the system as data are gathered during training sessions. The overall approach of the MPES analytic engine is triangulation of data, where any individual stress indicator is not sufficient to make a claim about stress levels, but 5 indicators are. MPES is designed to integrate with any virtual environment application for medical (and other) team training where awareness of participants' stress levels is valuable. MPES's open specification facilitates integration with any current and future application. The wearable biometric and sociometric sensors/devices are open source, as is the analytic engine.

The Analytic Engine was tested and improved using TPOT, a Python Automated Machine Learning Tool that optimizes machine learning pipelines using genetic programming. Essentially, TPOT makes a large number of models with different variations of algorithms and data processing, and only the "strongest" ones survive each generation. This works similar to natural selection where only the strongest species (pipeline) survives and in the end (in theory) the best possible evolution has occurred.

TPOT is initialized to train with five-fold cross validation based on 80% of the data, and then validates the trained model on 20% of the data TPOT was initialized with a population size of 100 and the number of generations was set to 400. Essentially, this means that TPOT will train 100 models for each generation it iterates through.

Testing will be performed in the Wide Area Virtual Environment (WAVE) (Liu 2018, Liu 2020). The WAVE is an 8,000 sq. ft. immersive virtual reality theatre designed for medical team instruction. It comprised of 24 screens forming two circular pods connected by a corridor. Each pod is approximately 25 ft. in diameter. The corridor is 20 ft long. The corridor tapers from 12 ft. at each end to 9 ft. in the middle. Within the WAVE, learners wear lightweight passive stereo glasses. They perceive an immersive and immersive 3D virtual environment while moving freely within the space. During a training exercise, the WAVE incorporates standardized patients, part task trainers, and light-weight props to simulate actual environments where a medical team can be deployed. E.g., combat or civilian mass-casualty, natural disaster, and chemical/biological/nuclear scenarios. Access to the WAVE is via an entrance in each pod. Each entrance is made up of two screens pivoted to swing outward. During use, they are closed to present a seamless environment. Fig. 2 illustrates.

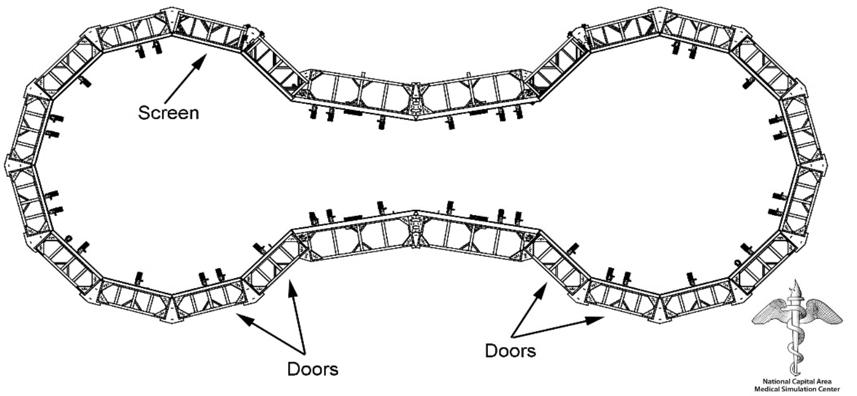

Figure 2: WAVE layout.

The WAVE has been used to simulate intense combat environments in support of student learners at USU. Varying scenarios are supported. Figs 3 – 5 illustrate. As both physical and virtual aspects of the scenario are controlled by the WAVE, highly consistent learning environments can be generated. This capability facilitates the investigation of team learning and performance.

RESULTS AND DISCUSSION

We ran TPOT multiple times, and in the nature of a genetic library each run resulted in slightly different scores and often with different algorithms chosen in the pipeline, but the scores were consistent around F1 0.90. A large study sample (e.g., 5000 or more participants) will be required to improve results and make claims at a sufficiently high level of confidence.

Note that working memory capacity and personality traits might perform better as a predictor (mediator) of stress, rather than as a moderator of stress. Field testing of a large sample will be needed to make these determinations.

NEXT STEPS

To date, all testing during product research and development has been with the research and development team as participants and with TPOT data. Despite this constraint, we achieved the project goal of delivering a product ready for large scale field testing using the Uniformed Services University's (USU) Val G. Hemming Simulation Center's Wide Area Virtual Environment (WAVE) and combat medic student participants.

Figure 3: Conducting a Focused Assessment with Sonography for Trauma (FAST) examination in a simulated forward surgical base environment.

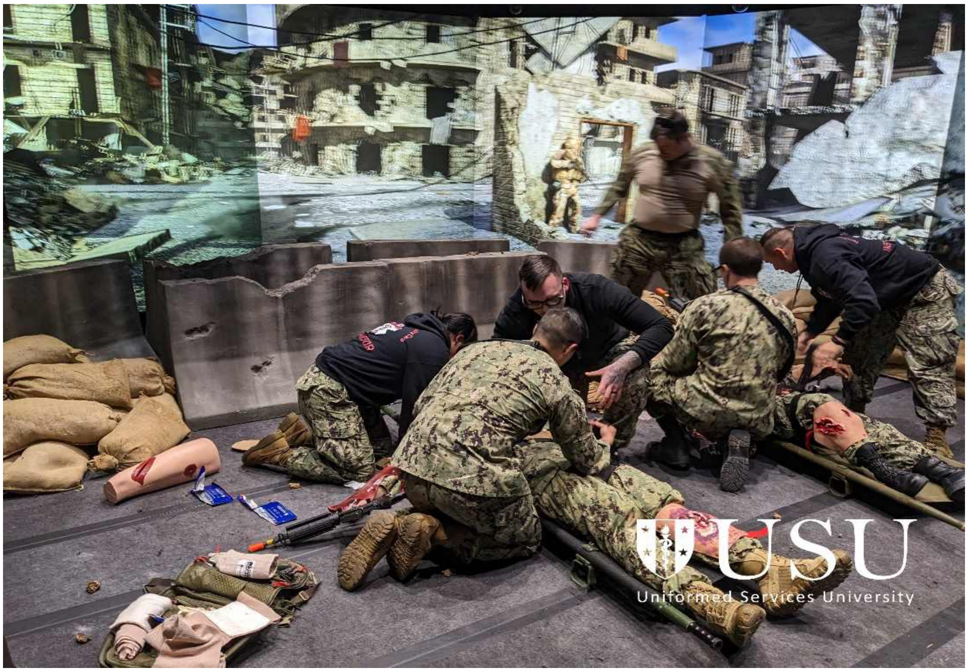

Figure 4: Simulating a point of injury scenario in a combat environment.

Figure 5: Additional effects, such as smoke, can be dynamically added to vary sensory load based on training requirements.

REFERENCES

Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the mini-IPIP in a large, nationally representative sample of young adults. Journal of personality assessment, 95(1), 74–84. https://doi.org/10.1080/00223891.2012.700466

F. Alamudun, J. Choi, R. Gutierrez-Osuna, H. Khan and B. Ahmed, “Removal of subject-dependent and activity-dependent variation in physiological measures of stress,” 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, San Diego, CA, USA, 2012, pp. 115–122, doi: 10.4108/icst.pervasivehealth.2012.248722.

Hale, H. B., Ellis, J. P. Jr., & Kratochvil, C. H. Jrnl of App. Physio, 14, 629–631. (2013).

Healey, J. Wearable and automotive systems for affect recognition from physiology. PhD thesis, MIT, Cambridge, MA, (2000).

Kirschbaum, Clemens, Karl-Martin Pirke, and Dirk H. Hellhammer. “The ‘Trier Social Stress Test’–a tool for investigating psychobiological stress responses in a laboratory setting.” Neuropsychobiology 28.1–2 (1993): 76–81.

Liu, A., Acosta E., Cope J., Henry V., Reyes F., Bradascio J., Meek W. (2018). The Wide Area Virtual Environment: A New Paradigm for Medical Team Training. In: Schmorrow, D., Fidopiastis, C. (eds) Augmented Cognition: Users and Contexts. AC 2018. Lecture Notes in Computer Science (), vol. 10916. Springer, Cham. https://doi.org/10.1007/978-3-319-91467-1_23

Liu, A., Acosta E., Cope J., Henry V., Reyes F., Bradascio J., Meek W. (2020). The Wide Area Virtual Environment - A Novel Immersive Environment for Medical Team Training. In: Ahram, T., Taiar, R., Gremeaux-Bader, V., Aminian, K. (eds)

Human Interaction, Emerging Technologies and Future Applications II. IHIET 2020. Advances in Intelligent Systems and Computing, vol. 1152. Springer, Cham. https://doi.org/10.1007/978-3-030-44267-5_61

Schoofs, D., Wolf, O. T. and Smeets, T., 2009. Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. Behavioral neuroscience, 123(5), p. 1066.