Virtual Reality Adaptive Stimulation in Stress Resistance Training

Krešimir Ćosić[1], Siniša Popović, Marko Horvat, Davor Kukolja, Branimir Dropuljić
University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

Ivica Kostović, Miloš Judaš, Marko Radoš, Milan Radoš,
Lana Vasung, Branka Bartolić Spajić
University of Zagreb School of Medicine, Croatian Institute for Brain Research, Šalata 12, 10000 Zagreb, Croatia

Svjetlana Doričić, Dalibor Mesić
Ministry of Defense, Republic of Croatia

ABSTRACT

Serious mental health problems in the current large-scale NATO operations underscore the importance of predeployment mental stress resistance program. Therefore, the development of new effective training tools and coping strategies for the minimization of operational stress disorders is extremely important. The concept of closed-loop virtual reality adaptive stimulation (VRAS) proposed in this paper may strengthen cognitive capacities and cognitive strategies in mission-threatening situations through repetitive delivery of stressful stimuli and simultaneous practice of relevant stress-coping skills. Stimulation training strategy is based on the gradual exposure of trainees to real-life mission-oriented video clips characterized by different stressful context, semantics and emotional properties. These audio-visual stimuli activate a cascade of events in the brain, which evokes various emotional and cognitive reactions in trainees. Their emotional stress response assessment is based on comprehensive physiological measurements and artificial neural network emotional state estimation. The neurobiological objective of such training program is focused on strengthening the inhibition of the amygdala circuitry network response by the prefrontal cortex. The orbitofrontal/ventromedial prefrontal cortex (OFC/VMPFC) context-dependent regulation and attenuation of amygdala reactivity might be enhanced by learning new cognitive appraisal of perceived threats, which is supported by the VRAS training system. Better regulation of the amygdala by the OFC/VMPFC may reduce operational failures and may even decrease severe operational casualties. A stronger inhibition of the amygdala leads to a decreased activation of the hypothalamus and the brainstem nuclei. Therefore, we may expect to observe lowering of the physiological reactivity to stressful stimuli with improvement in the stress-coping skills. The context, semantics and emotional richness of stress exposure stimuli, as well as individual cognitive appraisal, have a dominant impact on the intensity of perceived threats during VRAS training exercises. This is critically important for the entire VRAS training psychological concept. Learning new cognitive appraisals of perceived threats is extremely important for maintaining soldiers' emotional and behavioral control, mission success and minimization of the negative impact of stress on their mental health. The better amygdala regulation facilitated by the VRAS stress-resistance training program as a kind of new learning may provide contextual and cognitive regulation of stress response. Changes in the physiological reactivity of trainees during delivery of various stressful stimuli can be longitudinally monitored by the VRAS closed-loop system. Comparative analysis based on physiological measurements and functional magnetic resonance imaging (fMRI) in a pilot study of a small group of novice trainees versus ISAF deployment-ready Croatian soldiers is illustrated. Obtained pilot data indicate the importance of the parietal cortex and the temporo-parietal junction in coping with potentially stressful stimuli.

---

[1] Corresponding author: kresimir.cosic@esa.fer.hr

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **APR 2011**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Virtual Reality Adaptive Stimulation in Stress Resistance Training**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Zagreb Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**		
13. SUPPLEMENTARY NOTES **See also ADA578905. Mental Health and Well-Being across the Military Spectrum (Bien-être et santé mentale dans le milieu militaire). RTO-MP-HFM-205**		

14. ABSTRACT

Serious mental health problems in the current large-scale NATO operations underscore the importance of predeployment mental stress resistance program. Therefore, the development of new effective training tools and coping strategies for the minimization of operational stress disorders is extremely important. The concept of closed-loop virtual reality adaptive stimulation (VRAS) proposed in this paper may strengthen cognitive capacities and cognitive strategies in mission-threatening situations through repetitive delivery of stressful stimuli and simultaneous practice of relevant stress-coping skills. Stimulation training strategy is based on the gradual exposure of trainees to real-life mission-oriented video clips characterized by different stressful context, semantics and emotional properties. These audio-visual stimuli activate a cascade of events in the brain, which evokes various emotional and cognitive reactions in trainees. Their emotional stress response assessment is based on comprehensive physiological measurements and artificial neural network emotional state estimation. The neurobiological objective of such training program is focused on strengthening the inhibition of the amygdala circuitry network response by the prefrontal cortex. The orbitofrontal/ventromedial prefrontal cortex (OFC/VMPFC) context-dependent regulation and attenuation of amygdala reactivity might be enhanced by learning new cognitive appraisal of perceived threats, which is supported by the VRAS training system. Better regulation of the amygdala by the OFC/VMPFC may reduce operational failures and may even decrease severe operational casualties. A stronger inhibition of the amygdala leads to a decreased activation of the hypothalamus and the brainstem nuclei. Therefore, we may expect to observe lowering of the physiological reactivity to stressful stimuli with improvement in the stress-coping skills. The context, semantics and emotional richness of stress exposure stimuli, as well as individual cognitive appraisal, have a dominant impact on the intensity of perceived threats during VRAS training exercises. This is critically important for the entire VRAS training psychological concept. Learning new cognitive appraisals of perceived threats is extremely important for maintaining soldiers emotional and behavioral control, mission success and minimization of the negative impact of stress on their mental health. The better amygdala regulation facilitated by the VRAS stress-resistance training program as a kind of new learning may provide contextual and cognitive regulation of stress response. Changes in the physiological reactivity of trainees during delivery of various stressful stimuli can be longitudinally monitored by the VRAS closed-loop system. Comparative analysis based on physiological measurements and functional magnetic resonance imaging (fMRI) in a pilot study of a small group of novice trainees versus ISAF deployment-ready Croatian soldiers is illustrated. Obtained pilot data indicate the importance of the parietal cortex and the temporo-parietal junction in coping with potentially stressful stimuli.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**18**	

1.0 VRAS BASED STRESS RESISTANCE TRAINING

The main task and objective of NATO forces in the ISAF mission is to maintain maximum efficiency in the highly uncertain and stressful Afghanistan environment. However, intensive mission stressors may have huge impact on degradation of soldiers' cognitive performance to maintain course of actions and to correctly and effectively execute critical operational tasks. Stressful emotions and thoughts may reduce cognitive performance and decision-making, which can strongly affect soldiers' behavior and may cause tactical failures and even severe casualties reducing NATO's operational effectiveness. Therefore, assessment of emotional and behavioral control of ISAF forces and their mental health is an important indicator of NATO's readiness. Personal hardiness, stress reactions, stress resistance, as well as threat appraisal and coping strategies are important predictors of NATO soldiers' potential in the highly unpredictable and ambiguous Afghanistan environment. Monitoring and predicting individual cognitive performance under stress is therefore an important research topic. Lack of psychological training and skills may significantly increase operational risks due to uncertain stress-related reactions of individuals. The development of stress coping skills and capabilities is particularly important for novice soldiers who may lack coping resources. Their limited coping potential, including high physiological arousal, gives them less capacity to cope with combat risks and dangers. Therefore, the development of their coping adaptive strategies to get better control of their arousal levels, and more cognitive resources for mission effectiveness in life-threatening situations, is vitally important. Improving their stress resistance capabilities through different innovations in military psychological training programs may have important role in the modernization of NATO's military training. Virtual reality adaptive stimulation (VRAS) predeployment training tools presented in the paper are essentially a kind of stress exposure training, which may enhance soldiers' cognitive, emotional and behavioral characteristics and coping strategies. VRAS stressful stimulations in combination with psychological and tactical support by the trainers' team during the training session may strengthen the cognitive and behavioral potential of each individual to successfully cope with stressful situations. Strengthening the soldier's appraisal and cognitions by learning principles of mental readiness training [1] using VRAS development tools may improve predeployment training quality. The VRAS training program can be designed to gradually increase stressor intensity from an initial baseline to highly stressful stimulations, developing trainee self-confidence and mental toughness. The diversity of VRAS real-life video clips stimulation makes stress situations less unfamiliar to the novice trainees and may lead them to more positive expectations and self-efficacy. The VRAS stimuli generation flexibility increases the stressful stimuli repertoire in teaching and developing soldiers' relevant stress-coping skills, which complements stress-coping skills developed during the existing military and psychological training. Multimodal feedback is a crucial component of the VRAS development system since it ensures learning and developing new cognitive and emotional capacities. The VRAS multi-sensor feedback fusion is aimed to improve self-regulation of cognitive potential and coping skills of each individual. Emotional self-regulation using positive coping thoughts may minimize the deteriorating adverse physiological effects of acute stress and its long-term chronic consequences. Controllability and matching of new stimuli intensity with the emotional response of trainees is a main characteristic of VRAS adaptive mechanisms and its built-in state machine adaptive algorithms. Semantically and emotionally adaptive training tools are important for learning new coping skills that are transferable and applicable in uncertain stressful real-life operational environments. Customization of tactical operational contexts in which training exercises may take place is the main responsibility of the training team for developing the stimulation scenarios. Different contexts, multiple stressors, different missions and different landscapes based on real-life video clips enable comprehensive and well-crafted training scenarios. The role of military trainers, instructors, commanders and psychologists in the VRAS training exercises is extremely important in the context of the complexity and dynamics of the military training and coping skills development of trainees. Adverse stress impact on cognitive capacities can increase the likelihood of errors in soldiers' judgment and functioning in the field, which may contribute to failures in responding to the life-threatening combat situations. Cognitive-behavioral training based on comprehensive VRAS stress exposure training tools may enhance existing military psychological training and improve stress-coping potential of the military personnel, based on personalized stimuli, emotional state estimation and adaptive matching of new stimuli to the physiological reactions of trainees. Basic building blocks of VRAS development system and its main functional units are described in the following section.

2.0 CONCEPT OF VRAS DEVELOPMENT SYSTEM

The focus of the VRAS development system is to support personalized delivery of audio-visual stimuli to the trainees based on changes in their emotional responses. Stimuli may generally include static pictures, sounds, real-life video clips and synthetic virtual stimuli. Emotional response is related to the physiology, voice, speech, facial expressions etc. In this sense, VRAS encompasses time-synchronized stimuli delivery, acquisition of emotional response, estimation of the trainee’s subjective emotional state and closed-loop control that selects the stimuli of appropriate semantics and emotional properties based on the estimated emotional state.

The VRAS development system (Figure 1) has a diverse set of tools and an integrated workflow for configuration of different VRAS training sessions. The appropriate stimuli database (STIMDB) contains semantic and emotional annotations of stimuli, while training scenarios are contained in a scenario database (SCNDB). Processes related to the generation of the required stimuli relevant for the impending mission are described in section 2.1. These activities may include appropriate narratives read by the commanders in a stressful style if this is necessary to enhance the stimuli reality. VRAS trainees provide emotional annotations of delivered stimuli immediately after the session ends, which are stored in the trainee’s aggregated knowledge database (TAKDB) of each trainee, as well as in the STIMDB.

Another aspect of the VRAS configuration is related to the estimation of the trainee’s subjective emotional state by the Emotional State Estimator (*EmoEst*). *EmoEst* learning is based on previously collected information regarding the trainee’s subjective emotional experience, physiological, vocal and facial emotional manifestations, semantic and emotional properties of all previously delivered stimuli. Such information, collected from VRAS sessions, has been stored in the TAKDB and is available for use in learning the *EmoEst*. The learned *EmoEst* is stored in the estimator database (ESTDB) and ready for use in the next VRAS session.

The supervisors of the VRAS training team, which include a commander, a psychologist and a software engineer, may wish to analyze the results obtained from all previous sessions or current VRAS session in order to update broader knowledge regarding the trainee’s comprehensive reactions and to adjust the new stimulation strategy used in the next VRAS session. The stimulation strategy of the VRAS session may be defined in various ways according to the supervisors’ preferences and stored in the strategy database (STRATDB), which is described in a later section of this paper.

Any information regarding the trainee history and profile potentially useful for VRAS training should be defined in the appropriate machine-readable form. Examples of such structured information may include keywords that describe the earlier traumatic combat-related events that the trainee has encountered, the intensity of the trainee’s subjective emotional reaction to those events etc. In this way, VRAS has useful information regarding the semantics of the potentially highly stressful stimuli, and this information can be used in adjusting the stimulation strategy. Such information is also placed in the TAKDB of an individual trainee, and thereby in the reference knowledge database (RKDB), which includes TAKDBs of all individuals who participate in VRAS sessions.

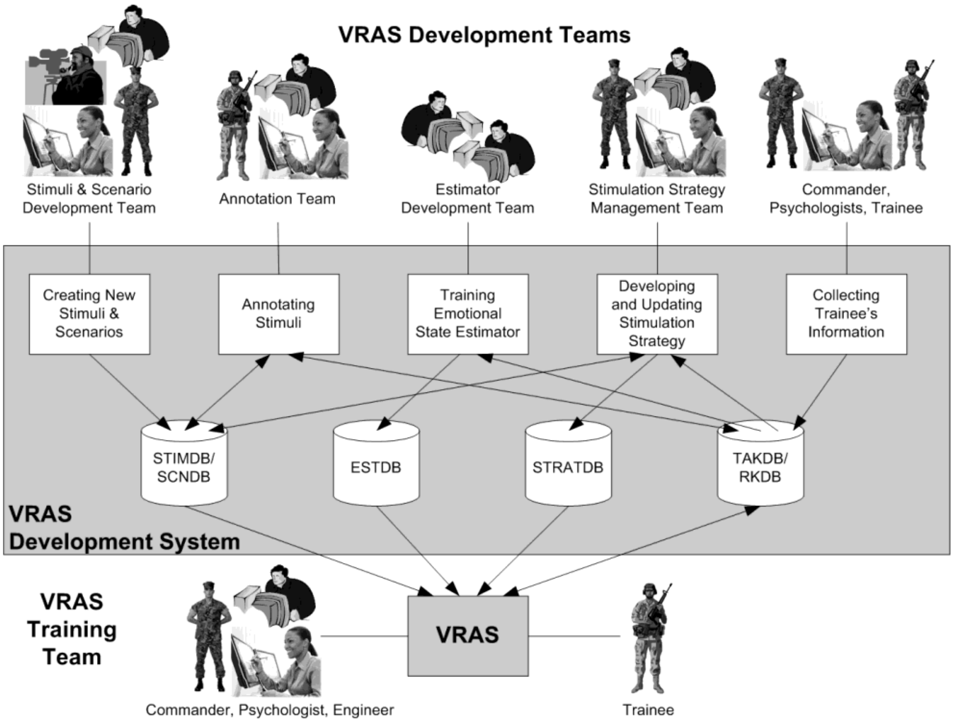

Figure 1: VRAS development system organizational structure.

Figure 2 sketches the four major logical subsystems [1] that make up the VRAS system, and their interactions during a VRAS session. The Stimuli Generator (*StimGen*) is a subsystem that generates audio-visual stimuli to the trainee. The *StimGen* receives control signals $\boldsymbol{u}(t)$ from the Adaptive Subsystem (*AdaptSub*) describing the desired media form, semantics and emotional properties of the stimuli to be generated. It then finds the most closely matching stimuli and exposes them to the trainee. The media in which stimuli may exist in the STIMDB may include static pictures, sounds, real-life video clips, or synthetic virtual stimuli.

The *EmoEst* receives multimodal response signals $\boldsymbol{MMR}(t)$ of the trainee and estimates the trainee's emotional state in a fixed frame time rate. Estimated emotional state may be represented in a multitude of ways. Typical representations include discrete emotions or dimensional model of emotions (e.g. see [3]), but any other representation may be appropriate as well, such as the level of stress, negative-neutral-positive etc. The multimodal response, computed features and the estimated subjective emotional state are forwarded as $\boldsymbol{MME}(t)$ to the *AdaptSub*, which may use any part of this information in decision-making.

The *AdaptSub* may contain an application-specific closed-loop control strategy that personalizes the stimuli delivery to the trainee, based on the trainee's emotional state. In order to deliver stimuli appropriately, the *AdaptSub* needs all available information that may be potentially relevant for making the right decision. The VRAS development system concept further allows that the *AdaptSub* by design has its decision making subordinate to the supervisors' authority. Via appropriate message exchange, various modes of collaboration between the supervisors and the *AdaptSub* can be supported.

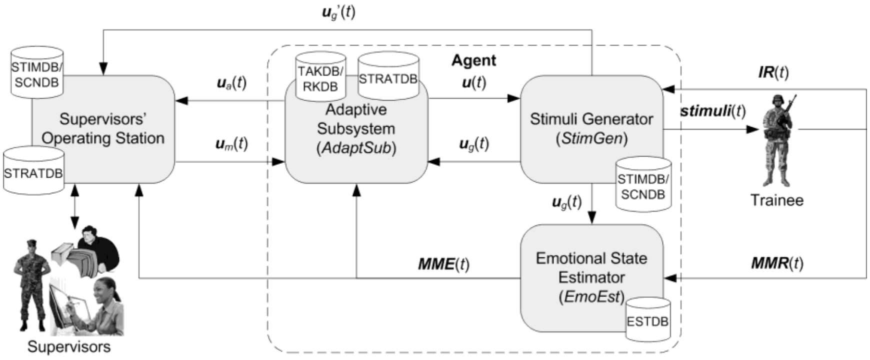

Figure 2: Model of VRAS in closed-loop self-adaptive mode.

The Supervisors' Operating Station collects for the supervisors all information that may help improve situational awareness and facilitate decision-making. To this end, the Supervisors' Operating Station generally needs to receive the same data from the *StimGen* and the *EmoEst* as the *AdaptSub*. Removing the *AdaptSub* from the VRAS model implies that the supervisors manually deliver the stimuli by means of the *StimGen*, while using the *EmoEst* to obtain extensive information regarding the trainee's emotional state. Furthermore, the *StimGen* and the *EmoEst* may be functionally usable for the supervisors even in a standalone fashion.

2.1 VRAS Tools for Stimuli Generation

The VRAS *StimGen* enables generation of multimedia and virtual stimuli, which are the best match with respect to semantic content, emotional properties, and general context of the soldier's future tasks and missions. VRAS *StimGen* supports several different methods of formal knowledge presentation and knowledge discovery. The first step in the intelligent generation of stimuli is the creation of a mission-specific annotated database with multimedia and VR scenes. The Stimuli & Scenario Development Team in Figure 1 includes several specialists like squad leader, psychologist, software engineer etc. The *StimGen* searches VRAS stimuli libraries based on input parameters given by the team leader or the *AdaptSub*. The software engineer uses custom command language syntax for stimuli database access to generate the most relevant multimedia stimuli for specific exposure strategy. The command language is designed to be user-friendly, enabling efficient man-machine communication to the trainer. In the VRAS closed-loop mode, the *StimGen* is controlled by the *AdaptSub* with control messages according to the appropriate format. These control messages integrate the *AdaptSub* and the *StimGen* into a closed-loop system mode for self-adaptive stimuli session. The control messages are semantically and emotionally streamlined, and also individually adapted to trainee's current physiological response. Command language and control messages allow complete administration of the *StimGen* actions in open- and closed-loop operational modes.

Management of multimedia and VR stimuli databases

The *StimGen* uses emotionally, semantically and contextually annotated stimuli stored in its own local multimedia library. The International Affective Picture System (IAPS) [4] and the International Affective Digitized Sounds (IADS) [5] are examples of publicly available databases of static pictures and sounds. These databases have poorly expressed semantics and no context descriptors whatsoever. In addition, their stimuli have generalized content which is not customized to military settings of current NATO ISAF peacekeeping mission. It is necessary to construct domain-specific VRAS stimuli databases in order to reach a better semantic alignment between stimuli content, sights and sounds of Afghanistan. Desirable stimuli have to be created from various multimedia documents: images, video and VR scenes. The images, video and VR scenes are retrieved from any available multimedia resource such as an Internet image repository, television and news agency archives, actual helmet-cam video footage, etc. Video clips are more favorable emotion elicitation media, because they have very high attentional capture and intensity [6]. The retrieved documents are thoroughly looked over and examined by the VRAS Stimuli & Scenario Development Team and Annotation Team. They are accepted for VRAS training sessions. VRAS-accepted stimuli affective, semantic and contextual content has to be approved, annotated and stored in a machine-readable format in the local VRAS database. This task is challenging and time-consuming due to its high requirements and sheer volume concerning the emotional, semantic and contextual annotation of multimedia documents. Examples of command language messages issued by the training team to the Stimuli Generation are illustrated below:

```
get    content(sem(keywords(Mazar al sharif,ambush,village)) AND
       emo(rectangle(2.1, 6.5, 3.1, 9)) AND context(year=2006));

get    content(sem(keywords(Helmand,country,Taliban,poppy fields)) OR
       context(year=2010,month=1-5)),stimuli-count(20);

get    content(sem(keywords(Kandahar,ambush,village), metrics(Wu-Palmer)));
```

Keywords illustration for Afghanistan-specific stimuli are: Afghanistan, attack, ambush, IED, improvised explosive device, RPG, AK-47, bomb, mine, anti-tank mine, mortar attack, explosives, insurgency, suicide bomber, soldier, police, Taliban, civilians, locals, dead people, dead children, injured people, injured children, hospital, school building, bombed school, massacre, village, road, bridge, mountain, mountain pass, cave, poppy field, opium, irrigation, military base, safe zone, vehicle, transport, convoy, vehicle, Humvee, LAV, patrol, village clearing, assault, rescue, close support...

Knowledge representation

The semantic content of VRAS stimuli and its affective annotation predominantly influences perceived threats on trainees and is of the greatest importance for the success of the mental readiness training. Since IAPS and IADS databases are information-poor, semantically scattered and with patchy taxonomy, we implemented Princeton’s WordNet lexical database of the English language [7] in the VRAS stimuli database. WordNet is a widely accepted formal knowledge representation mechanism containing over 200,000 word-sense pairs organized into synonym rings (synsets) that can be used to expand the vocabulary for description of VRAS stimuli content. All synsets are arranged by their meaning into a single graph which allows assessment of semantic distance between any two words in the WordNet’s vocabulary. This important ability enables detection of stimuli annotated with keywords that are similar and not necessarily identical to the keywords given as input in the command language and control messages. Similarity, relatedness and distance between hierarchically ordered keywords can be automatically calculated using several metrics which are implemented in the VRAS. The most important metrics family is based around the edge-counting principle where the semantic distance between any two nodes in the semantic graph is directly proportional to their shortest distance, i.e. to the least number of nodes between them.

Knowledge about the objects incorporated in the stimuli, events and their respective interworking is stored separately from multimedia STIMDB database in the STIMKDB knowledge database. Stimuli content is identified during the VRAS stimuli database construction process and represented with formal keywords from a supervisor’s vocabulary. Every change in this type of vocabulary has to be approved and new keywords cannot be added or changed without retaining semantic relationships with other keywords in the WordNet graph. Consequently, the semantic description of stimuli remains organized and undisturbed. Each stimulus can be described with any number of keywords. There are no restraints in detail or complexity of the semantic annotations apart from those that are imposed by the content of the WordNet graph. The necessary level of detail in the semantic description of VRAS stimuli is agreed upon by the Annotation Team (Figure 1). The annotated objects are described with nouns, whereas the actions they take are described with verbs. The several tens of thousands different English words available in the WordNet are usually enough to achieve adequate description level, but if even richer description is necessary it is possible to introduce specialized WordNet domains. We also use the emotionally-annotated databases Affective Norms for English Words (ANEW) [8], Affective Norms for English Text (ANET) [9], WordNet domains for expression of emotions and public opinion - WordNet Affect [10] and SentiWordNet [11]. Data mash-up from all these sources ensures richer semantic expressivity to the experts and enables them to use a broader vocabulary in VRAS stimulation tools. To enrich these semantics further, we integrated IEEE's Suggested Upper Merged Ontology (SUMO) [12], which has built-in WordNet sense mappings and allows reasoning about the commonsense knowledge that is implied in the description of multimedia files.

Stimuli meta-data and context

Apart from the semantics and emotions, each multimedia document has a specific context, i.e. information about itself and its content. This meta-data cannot be overlooked since it may be helpful in querying the multimedia database and stimuli retrieval. Images, sounds, video-clips and VR scenes are a part of emotionally annotated databases, however their properties such as title, creator, date and time when they were created, location name, location description, GPS coordinates of the location, multimedia type and format must be explicitly specified. Also, VRAS uses links to DBPedia pages in order to achieve a common understanding in data structure and allow expansion of context description. For enrichment of knowledge about the multimedia content and description of stimuli context, the *StimGen* uses the ISO-standardized Dublin Core metadata element set [13]. Since this set is a *de facto* standard for cross-domain information resource description it can ensure a simple and standardized set of conventions for description of resources in emotional annotated databases such as IAPS and IADS and simplifies collaborative and cross-domain data exchange.

Automatic stimuli selection

Design of an appropriate exposure sequence takes into account psychologically relevant parameters of each individual trainee and their military operational challenges. Intelligent generation of multimedia and VR relieves members of the training team from repetitive and labor-intensive tasks enabling optimal adaptation of stimuli sequences during mental readiness training sessions. *StimGen* supports data retrieval by searching and browsing by means of implemented procedures of knowledge discovery in affectively annotated multimedia documents. In the searching paradigm, *StimGen* selects the concrete stimuli by distance minimization along three data axes: semantics, emotions and context. Firstly, stimuli are selected semantically by pattern matching of the keywords from the stimuli database versus the keywords specified by the control messages. It is important to note that these keywords in fact represent knowledge representation formalism, since they have a well-defined syntax and semantics that can be used to generate new knowledge from the existing one. Secondly, stimuli are searched according to the affective valence/arousal values specified by the machine control messages or the VRAS command language and the emotional annotations of stimuli in the multimedia database. Finally, the third component in selection of concrete stimuli is context, meta-data or data about the stimulus itself such as time, date and location where the multimedia document was created, who is the author and the owner, what are the technical properties of the documents (multimedia format, width, height, resolution, color depth, etc.). Emotional properties, semantics and context in the search query can be specified separately or together, and the *StimGen* will choose the most appropriate stimuli according to cumulative distance minimization. After the most appropriate stimuli have been retrieved from the stimuli database they have to be sorted according to their rank so the best-matched documents appear before others. In this stage of the automatic stimuli selection process, the ranking algorithm will calculate compound semantic and emotional distance between the input terms in the supervisor's command language or control messages, and the descriptive keywords in the retrieved multimedia documents. Further improvements of the selection are possible through application of data mining, intelligent algorithms and artificial reasoning to automatically identify content in new multimedia documents.

The following example shows a list of messages that will generate a sequence of video-clips from Afghanistan, from relatively peaceful but anxiety-packed scenes of driving through Kabul, interacting with local people, with intensive and violent footages of close combat taken with helmet-cams:

$\boldsymbol{u}(t_1)$ = (1; VIDEO; EMO; ε; Afghanistan, Kabul, driving, road, trucks, people, city; ε; ε; 3.8 6.3; ε)

$\boldsymbol{u}(t_2)$ = (2; VIDEO; EMO; ε; Kandahar, countryside, ambush, RPGs, AK-47, shooting, taking cover, returning fire, calling support; ε; ε; 3.5 6.1; ε)

These Afghanistan-related video sequences, illustrated[2] in the following figure, are obtained by the selected control messages of 30-second duration and will be used in experimental testing based on physiological VRAS measurements and neural activity functional magnetic resonance imaging (fMRI).

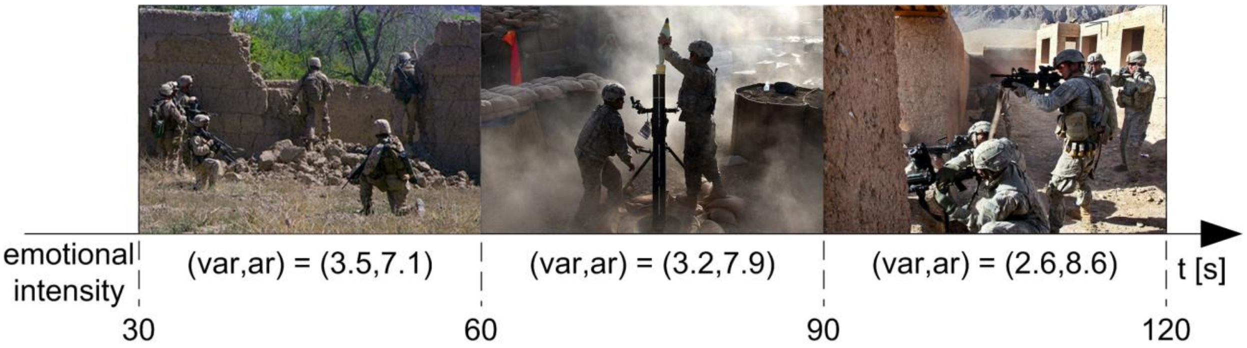

Figure 3: Example of low valence/high arousal video sequence with Afghan-related semantics used for VRAS training.

2.2 VRAS Tools for Emotional State Estimation

The *EmoEst* provides comprehensive multimodal information regarding the emotional state response of each individual in the form of physiological (heart rate, skin conductance, respiration, etc.) and audiovisual (facial expression, voice, gesture, posture, etc.) response of the trainees. The emotion relevant signal patterns may widely differ from person to person and from situation to situation; therefore, *EmoEst* needs to be gradually learned for each trainee. This section describes the concept of progressive emotional state learning and estimation (Figure 4), based on physiology signals like skin conductance (SC), electrocardiogram (ECG), heart rate (HR), respiration rate etc.

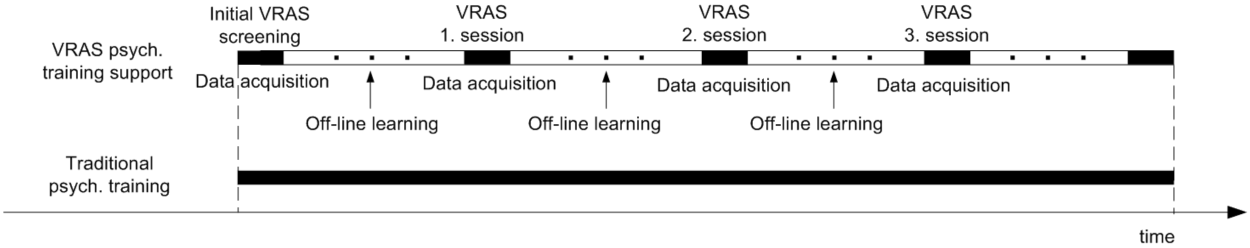

Figure 4: *EmoEst* learning process.

---

[2] Photos used in this paper are public domain (source: Wikimedia Commons) and serve merely to demonstrate the nature of the actual video content.

Initial VRAS screening based on IAPS/IADS stimuli

The initial screening sequence begins with corresponding physiological measurements of a neutral low-arousing “baseline” stimulus and pairs of IAPS pictures and IADS sounds with congruent valence, arousal and semantics. The emotional patterns of initial screening sequence uniformly cover 2D valence/arousal space. After the initial screening sequence, the trainee provides subjective ratings of emotions elicited by each presented stimulus in terms of valence and arousal. Based on their subjective ratings, an emotional map is created that serves as a reference starting point. The data collected during this session represent an initial snapshot of the trainee’s reactions on general IAPS/IADS stimuli.

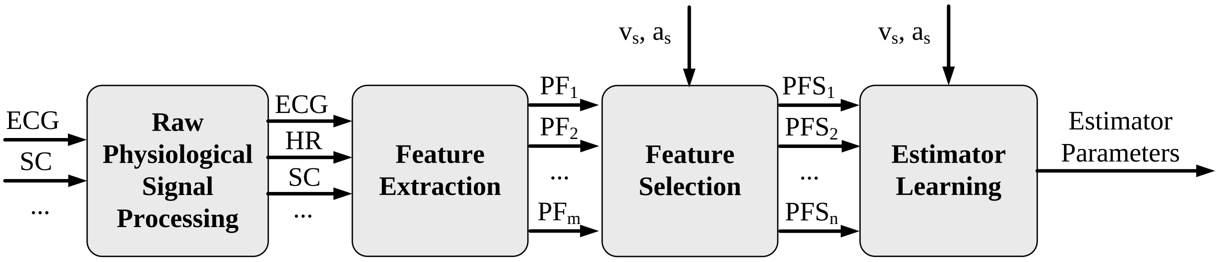

Figure 5: Off-line learning steps.

Based on the IAPS/IADS delivered stimuli, recorded physiology and subjective ratings, the first version of the *EmoEst* is learned. *EmoEst* separately estimates unquantized values of valence and arousal. The off-line learning procedure has the following steps:

1. **Raw physiological signal processing.** Includes preprocessing, waveform detection (like QRS complex detection, SCRs detection, etc.) and additional physiological signal extraction (like heart rate from ECG, respiration rate from respiration sensor, etc.)

2. **Physiological features extraction.** Based on the preprocessed signals, a set of physiological features that might indicate the trainee’s emotional state is calculated. Typical physiological features are statistical measures such as mean, standard deviation, minimum, maximum or mean of first difference of each signal, heart rate variability (HRV) features, amplitudes and frequency of skin conductance responses (SCRs) etc.

3. **Feature selection.** Using the Sequential Floating Forward Selection (SFFS) [14] algorithm, a selection of physiological features is conducted for valence and arousal estimators. As a criterion function of SFFS that needs to be maximized in order to determine the best feature set, an accuracy of estimation obtained using leave-one-out cross-validation [15] has been used. Selection of the physiological features speeds up *EmoEst* learning and contributes to the accuracy of estimation.

4. ***EmoEst* supervised learning.** Using data obtained by pairing the selected physiological features (PFS_1,…PFS_n) with valence/arousal subjective ratings (v_s, a_s) supervised learning is used to set the estimators’ parameters (see Figure 6).

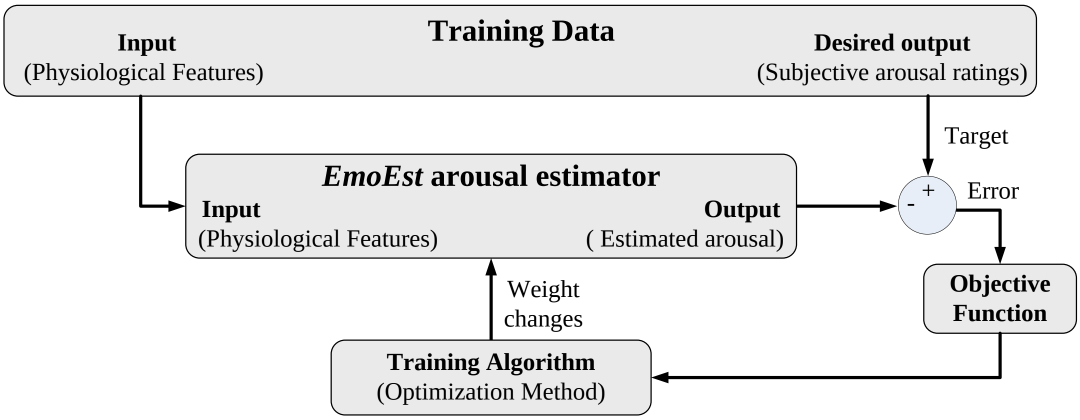

Figure 6: Supervised learning scheme for *EmoEst* arousal estimator.

The output of the first learning phase is the first version of the learned *EmoEst*: the relevant physiological features have been selected and functions that compute the emotional state of the trainee based on the selected features have been learned. Such functions can be obtained by using, for example, an artificial neural network, support vector machine, REPTree, etc. Learned *EmoEst* can be used later during VRAS sessions to estimate, on the basis of measured physiology, the emotional state of the trainee. In Figure 7 an example of the learned *EmoEst* arousal estimator based on an artificial neural network is shown. Certainly, the *EmoEst* also has a similar valence estimator, which may use a different set of physiological features compared to the arousal estimator.

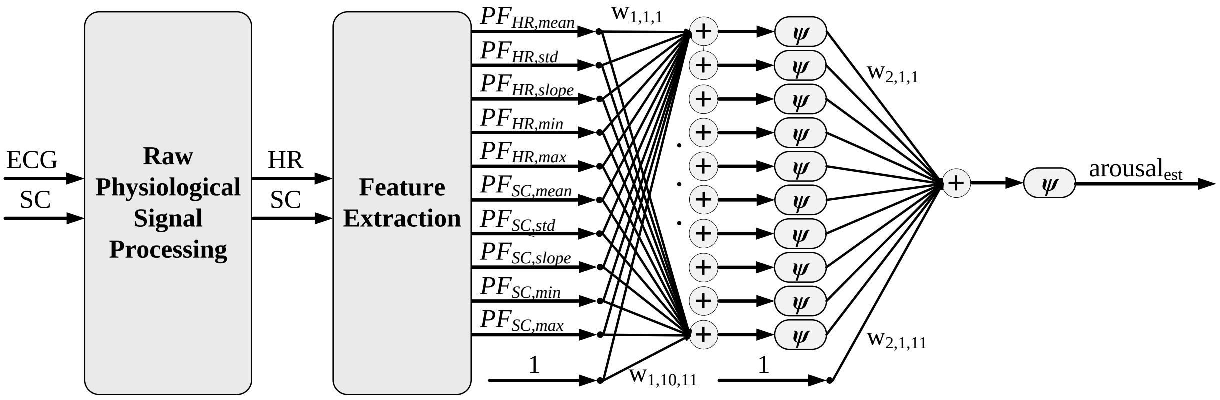

Figure 7: An example of the learned *EmoEst* arousal estimator based on artificial neural network.

VRAS training sessions based on real-life video clips

VRAS sessions use real-life video clips with content which approximates the situations that might be potentially encountered in the impending mission. The video clips can be shown to the trainee in a predetermined fashion, or they can be determined by *AdaptSub* based on the trainee's physiological, vocal, facial and/or subjective manifestations of emotion. Video clips are presented to the trainee jointly with congruent narratives, in which the instructions and comments of the commander and potentially other unit members can be heard as a part of emotionally based communications and cognitive and emotional learning.

After each VRAS session, the trainee provides subjective ratings of emotion elicited by each presented stimulus in terms of valence and arousal. Similar to the initial screening, after each sequence of VRAS training, *EmoEst* off-line learning is being preformed (see Figure 4), which contains raw physiological signal processing, physiological features extraction, feature selection and *EmoEst* supervised learning (Figure 5).

2.3 VRAS Tools for Adaptation of Stimulation Strategy

All VRAS tools are introduced to provide new value to the existing psychological preparations for highly stressful missions. VRAS supervisors use these available tools jointly with existing training approaches to shape the trainees' cognitive appraisals of highly stressful situations in the most adaptive way. To this end, during multiple VRAS sessions, delivery of audio-visual stimuli evolves with the progress of the trainee, as well as with the growing supervisors' understanding what the best stimulation strategies might look like.

The first session in VRAS psychological training support is the initial VRAS screening session, which according to section 2.2 includes delivery of IAPS/IADS congruent pairs of static pictures and sounds. The supervisors may update stimulation strategy for all later sessions that use real-life video clips, after analyzing the data from TAKDB collected during earlier sessions. For example, the supervisors may find strong physiological reactions across VRAS sessions to stimuli of certain semantics, and decide to create a specific predetermined scenario that focuses on this semantics and context. At a particular point during the subsequent VRAS session, such scenario can be selected by the supervisors and delivered to the trainee. Alternatively, the supervisors may specify the keywords denoting the desired semantics for the next session, and type the keywords during the VRAS session for interactive closed-loop delivery of desired stimuli. If additional automation is desired to adjust stimuli duration to the fluctuations in the trainee's emotional reactions, the automatic closed-loop stimulation strategy focusing on specific semantics and tactical scenario may be defined by the commanders and psychologists. In any case, the stimulation strategy can be automatically updated or predefined in advance and stored for use in a subsequent VRAS session. Besides referring to the stimuli semantics when defining the stimulation strategy for the next VRAS session, emotional valence/arousal properties of stimuli can also be used explicitly, because all stimuli have emotional annotations assigned in the STIMDB.

Updates to the stimulation strategy between sessions by the supervisors, as well as delivery of individual stimuli within a particular session, may be represented as a state machine that also includes different fuzzy logic algorithms. In this sense, depending on the judgment of the VRAS supervisors, parts of the strategy may be implemented in the *AdaptSub*. However, structure of this state machine depends on specific stimulation training goals and objectives. A sketchy illustration of a potential state machine is given in Figure 8, which is constructed from the pseudocode of the stimuli delivery algorithm [1]. The figure shows that at the beginning of any session other than the first one, the *AdaptSub* retrieves its internal state and required data from prior sessions. Next, the session includes several minutes of baseline measurements, e.g. without delivering specific stimuli. This is necessary to control for the effects of day-to-day variations in the trainee's emotional responses, like those in autonomic physiology. Following the baseline measurements, the session consists of multiple cycles of exploratory and repetitive stimuli delivery stages. In the exploratory stage, the trainee is presented with various stimuli from the stimuli database, according to the search algorithm in the stimuli space determined by the VRAS supervisors. The trainee's estimated emotional state may be used for ranking the stimuli based on their potential to elicit observable emotional responses. The stimuli rank list is updated upon every delivery of a new stimulus in the exploratory stage. The supervisors may choose the stimuli in real time according to their judgment, or use the search algorithm implemented as the computer program (e.g. which applies fuzzy rules to the trainee's estimated emotional states). When a sufficient number of stimuli has been found to elicit considerable emotional manifestations, the repetitive stage may repeatedly deliver them for application and strengthening of the trainee's stress-coping skills under the supervisors' guidance. Exploratory and repetitive stimuli delivery stages may alternate multiple times before the session ends. Finally, the *AdaptSub* stores its internal state for retrieval at the beginning of the subsequent session. All the time during the session, time-synchronized information regarding the delivered stimuli, emotional responses, and estimated emotional state is stored in the TAKDB.

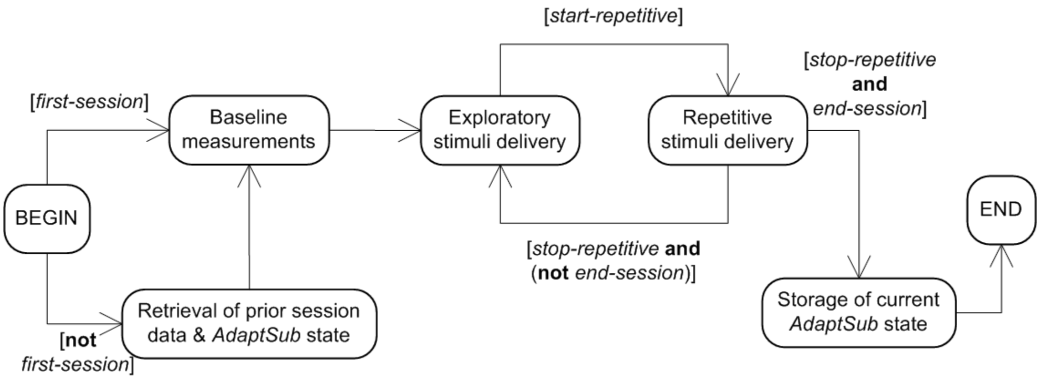

Figure 8. An illustration of a high-level state machine for stimuli delivery during VRAS session.

3.0 NEUROBIOLOGICAL CONSIDERATIONS

Upon delivery of an audio-visual stimulus, the information from audio and video sensors travels via the sensory thalamic nuclei to the primary auditory and visual cortices, and later on to the secondary associative cortices for further processing. The multimodal integration of these audio-visual mental maps with long-term memories is performed in the parietal-temporal-occipital association cortex (PTOAssoc) [16]. These brain circuits have bidirectional connections with the prefrontal association cortex (PFC), which is likewise bidirectionally connected with the amygdala. The “top-down” regulation of amygdala by the orbitofrontal/ventromedial prefrontal cortex (OFC/VMPFC) [17] is particularly important, since it involves an integration of cognitive appraisal information, by bidirectional connections between parietal-temporal-occipital region and dorsolateral prefrontal cortex (DLPFC). This regulation of emotions based on strengthening cognitive appraisal potential and coping skills provided by the new contextual stress adaptive learning through PTOAssoc and DLPFC interconnections is extremely important. Therefore, an approach to improving the stress resistance, which involves “familiarization” with stressful situations via repetitive stimulation, may strengthen soldiers’ cognitive appraisal abilities regarding potential threats. These training processes make amygdala regulation and inhibition by the OFC/VMPFC activity during the execution of stressful tasks more efficient on synaptic level. A stronger inhibition of the amygdala leads to a decreased activation of the hypothalamus and the brainstem nuclei and consequent downregulation of the visceral system. Therefore, we may expect lower physiological reactivity to stressful stimuli with improvement in the stress-coping skills. Changes in the physiological reactivity of an individual during delivery of various stressful stimuli can be also longitudinally monitored by VRAS system. The context, semantics and emotional richness of stress exposure stimuli and cognitive appraisal of these stimuli as real threats [18] predominantly determine amygdala response and its couplings with other brain structures. This suggests the importance of threat appraisal cognitive brain structures and mechanisms, like PTOAssoc. This is critically important for the entire VRAS training psychological concept. Threat perception and corresponding response can be evaluated by VRAS physiological measurements, as well as fMRI.

The threat appraisal objective includes stimulus evaluation according to the algorithm which takes into account the following criteria: threat/stimuli/event relevance, implication, coping potential and normative significance [19]. Relevance criteria are concerned with how relevant the threats/stimuli are for the trainee. For example, if the threats/stimuli are regarded as personally irrelevant, no emotion is elicited. Additionally, novel threats/stimuli are typically appraised as relevant, as they are unknown to the trainee and are worthy of mobilization of attention toward them. According to implication criteria, the threat/stimulus is evaluated with respect to how much it facilitates or obstructs the attainment of the trainee's goals. If the goals are obstructed, negative emotions are elicited; alternatively, positive emotions occur. The trainee also estimates likelihood of future consequences of the threat/stimulus, when assessing its implications. Assessment of coping potential is related to the ability of the trainee to deal with the consequences of the threat/stimulus event. Here, the trainee assesses whether it is in principle possible to influence the event or its outcomes and how much of the desired influence the trainee is capable of exerting. Furthermore, the degree to which one can adjust to the consequences of the event is evaluated, when the event is not controllable at all or there is a lack of means to exert the desired influence. Thus, the trainee seeks ways of restoring balance following the induced perturbation, e.g. by targeting the threat/stimulus that caused this perturbation or by accepting its consequences. Evaluation of the normative significance of the threat/stimulus event entails testing it against the criteria of compatibility or incompatibility with the trainee's self-concept and social norms and values. This assessment yields a degree of compatibility, can be applied both to the threat/stimulus itself and to the possible coping responses, and it requires the most sophisticated biological machinery of the PFC.

Therefore, strengthening the cognitive appraisal of real mission threats is extremely important for mission success and attenuation of the negative impact of stress on the mental health of NATO soldiers.

4.0 PILOT STUDY REPORT

Initial experimental fMRI research has been done with a few novice and ISAF deployment-ready soldiers who finished 6-month predeployment training and were ready for immediate deployment to Afghanistan. Some comparison of their emotional and cognitive potential versus novice candidates has been done using VRAS *EmoEst* and fMRI. Due to the budgetary restrictions, a small group of participants has been included in the fMRI study, with an equal number of novice and deployment-ready soldiers. However, this initial study is not the best choice for more evident evaluation of VRAS effectiveness in stress resistance training. Paradigm of Afghanistan video stressful situations that may potentially happen soon in the field to some of them has been designed by the Stimuli Generator expert team. fMRI paradigms with mission-relevant real life video clips were shown to the deployment-ready and novice soldiers. Each fMRI paradigm consisted of three static pictures of a black “+” sign on a white background, alternating with three video clips. Every static picture or video clip lasted 30 seconds, thus making a specific paradigm last 3 minutes.

Functional imaging data were preprocessed and analyzed using SPM2 (Wellcome Department of Imaging Neuroscience; see www.fil.ion.ucl.ac.uk/spm). Here, we have illustrated only the differences in their mental involvement related to the challenging Afghanistan mission environment. The most significant results obtained by the SPM analysis were the differences between deployment-ready and novice soldiers comparing their PTOAssoc and PFC activation on the combat paradigm (Figure 9). Deployment-ready soldiers showed greater activation in the temporo-parietal junction (areas that have been recently described as important for the negative affectivity and social inhibition [20]) and in the superior parietal lobe known for its roles in integrating sensory information from various parts of the body and visuospatial processing. Preliminary and limited testing shows greater involvement of the deployment-ready soldiers in the Afghanistan combat video clip; this is in line with appraisal theory regarding their perception of a threatening environment that is much closer to them than to the novice recruits. It suggests that Afghanistan stressful combat stimuli relevance, congruence and ego-involvement is more evident in the ISAF deployment-ready participants rather than novice soldiers.

Initial physiological measurements have also been done, including skin conductance, skin temperature, heart rate and respiration signals, together with soldiers’ subjective ratings. These very preliminary data in a small pilot study are insufficient for revealing statistical differences.

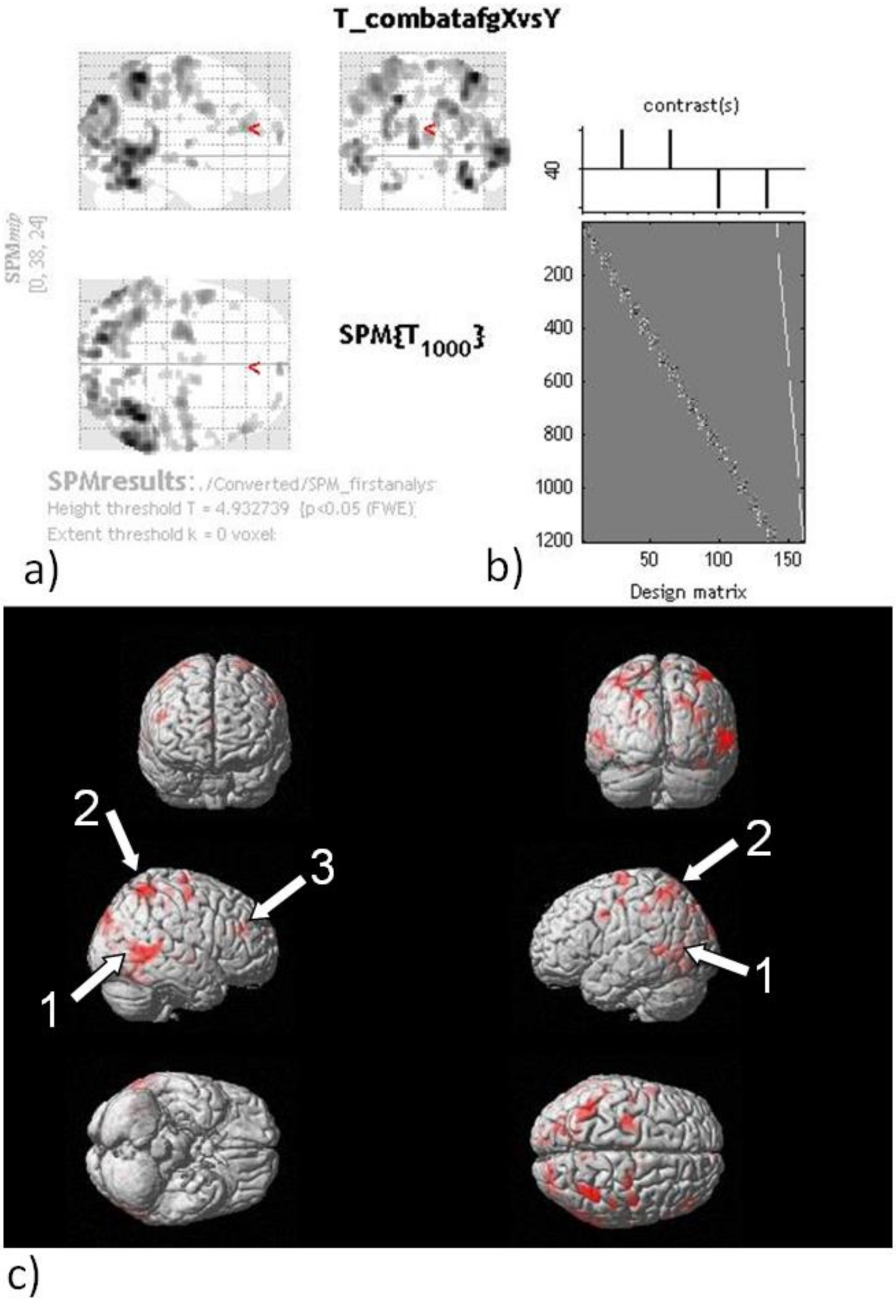

Figure 9: Activation patterns (p<0.05) in deployment-ready soldiers (group X) compared to novices (group Y) during the Afghanistan combat paradigm. In figure a, the overall pattern of activation is shown in the Talairach Tournoux Stereotactic space. Figure b shows the matrix of paradigms in whole study and in figure c the activation patterns are displayed on the surface render. Specific areas of activation: 1 – temporo-parietal junction, 2 – superior parietal lobule, 3 – gyrus frontalis medius. Measurements have been done at the Croatian Institute for Brain Research, with Siemens 3T MR scanner.

5.0 FUTURE WORK

In a joint cooperation with the Croatian Ministry of Defense, we are planning more statistically relevant validation and verification of VRAS tools in stress resistance training of Croatian Armed Forces contingents for NATO ISAF mission. Obtaining more evidence and empirical data that support the usefulness of VRAS as a mental health training tool requires much more time and resources than we had available in a pilot study academic research. Therefore, for example, comparative analysis between group of Operational Mentor and Liaison Team/Air Mentor Team (OMLT/AMT) soldiers trained with VRAS versus control group after their return from ISAF rotation would be much more important. This should be a challenging research topic of broader interest. But definitely, theoretical background of VRAS concept and its comprehensive state-of-the-art tools support expectations that such approach might contribute to increasing the individual soldier’s psychological readiness.

Regarding VRAS development system, we are also planning to enhance stimuli generation, emotional state estimation and closed-loop state machine methods and algorithms. To improve estimation accuracy, comparative analysis is planned between various emotional state estimation methods. Fusion of different estimator architectures including various feature sets will be analyzed. In order to emphasize human individuality, appropriate estimator customization to each individual session will also be considered. Since emotions are inherently multimodal, an integrated multimodal analysis may help to resolve ambiguities and compensate different errors. Accordingly, various acoustic features from the vocal cords oscillation period, energy and power of the voice, zero-crossing rate and the voice spectrum will also be analyzed in emotional state estimation using hidden Markov models.

6.0 ACKNOWLEDGEMENTS AND DISCLAIMER

This research has been partially supported by the Ministry of Science, Education and Sports of the Republic of Croatia. Kind support has also been provided by the Ministry of the Family, Veterans’ Affairs and Intergenerational Solidarity. We gratefully acknowledge organizational and coordination assistance by Brigadier Ivica Kinder and Colonel Mladen Prebežac from the Ministry of Defense of the Republic of Croatia. We are also indebted to Dr. Tanja Jovanovic of Emory University for kind assistance in improving readability of this paper. Our gratitude further goes to one of the technical evaluators for NATO HFM-205 symposium, for a suggestion related to the potential future research evaluating the usefulness of VRAS system in stress resistance training. The opinions, views and ideas expressed in this paper are those of the authors and do not necessarily reflect the position or policy of the sponsors or the institutions the authors are affiliated with.

7.0 REFERENCES

[1] M.M. Thompson, D.R. McCreary, “Enhancing mental readiness in military personnel”. In: Human Dimensions in Military Operations – Military Leaders’ Strategies for Addressing Stress and Psychological Support, Meeting Proceedings RTO-MP-HFM-134, RTO, Neuilly-sur-Seine, France, pp. 4-1–4-12, 2006.

[2] K. Ćosić, S. Popović, D. Kukolja, M. Horvat, B. Dropuljić, “Physiology-driven adaptive virtual reality stimulation for prevention and treatment of stress related disorders”. CyberPsychology, Behavior, and Social Networking. 13, 1; 73-78, 2010.

[3] E. Fox, “Emotion science: cognitive and neuroscientific approaches to understanding human emotions”. Hampshire, UK: Palgrave Macmillan, 2008.

[4] P.J. Lang, M. Bradley and B. Cuthbert, “International Affective Picture System (IAPS): Instruction Manual and Affective Ratings, Technical Report A-6“. The Center for Research in Psychophysiology, University of Florida, USA, 2005.

[5] M. Bradley and P.J. Lang, “International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2)“. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida, USA, 1999.

[6] J. Rottenberg, R.D. Ray, J.J. Gross, “Emotion elicitation using films”. In: J.A. Coan, J.J.B. Allen, editors. Handbook of emotion elicitation and assessment. New York, NY: Oxford University Press, 2007. pp. 9–28.

[7] G.A. Miller, R. Beckwith, C. Fellbaum, D. Grossand and K.J. Miller, “Introduction to WordNet: an on-line lexical database“. International Journal of Lexicography, vol. 3, no. 4, pp. 235–244, 1990.

[8] M. Bradley and P.J. Lang, “Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings”. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida, USA, 1999.

[9] M. Bradley and P.J. Lang, “Affective Norms for English Text (ANET): Affective ratings of text and instruction manual”. (Tech. Rep. No. D-1). University of Florida, Gainesville, USA, 2007.

[10] C. Strapparava and A. Valitutti, “WordNet-Affect: an Affective Extension of WordNet“. In: Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, pp. 1083–1086, 2004

[11] A. Esuli and F. Sebastiani, “SentiWordNet: A publicly available lexical resource for opinion mining”, In: Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation, pp. 417–422, Genova, Italy, 2006.

[12] A. Pease, I. Niles and J. Li, “The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications”. In: Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, 2002.

[13] D. Hillmann, “Using Dublin core. Recommendation”, Dublin Core Metadata Initiative, 2001.

[14] P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection”. Pattern Recognition Letters, vol. 15, pp. 1119–1125, 1994.

[15] R.R. Picard and R.D. Cook, “Cross-validation of regression models”. Journal of the American Statistical Association, vol. 79, no. 387, pp. 575–583, 1984.

[16] C.J. Forehand, “Integrative functions of the nervous system”. In: R.A. Rhoades, D.R. Bell, editors. Medical physiology: principles for clinical medicine, 3rd. ed., Lippincott Williams & Wilkins, 2008, pp. 122–139.

[17] A.F.T. Arnsten, “Stress signalling pathways that impair prefrontal cortex structure and function”. Nature Reviews Neuroscience, vol. 10, pp. 410–422, 2009.

[18] G.A. van Wingen, E. Geuze, E. Vermetten, and G. Fernandez, “Perceived threat predicts the neural sequelae of combat stress”. Molecular Psychiatry, published online 18 January 2011.

[19] K.R. Scherer, “Appraisal considered as a process of multilevel sequential checking”. In: K.R. Scherer, A. Schorr, T. Johnstone, editors. Appraisal processes in emotion: theory, methods, research. New York and Oxford: Oxford University Press, 2001, pp. 92–120.

[20] M.E. Kret, J. Denollet, J. Grezes, B. de Gelder, “The role of negative affectivity and social inhibition in perceiving social threat: an fMRI study”, Neuropsychologia, in press, uncorrected proof, available online 10 February 2011.