34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Vital Analysis: Annotating sensed physiological signals with the stress levels of first responders in action*

P. Gomes[1], M. Kaiseler[2], C. Queirós[3], M. Oliveira[3], B. Lopes[1] and M. Coimbra[1], *Member, IEEE*

***Abstract*— First responders such as firefighters are exposed to extreme stress and fatigue situations during their work routines. It is thus desirable to monitor their health using wearable sensing but this is a complex and still unsolved research challenge that requires large amounts of properly annotated physiological signals data. In this paper we show that the information gathered by our Vital Analysis Framework can support the annotation of these vital signals with the stress levels perceived by the target user, confirmed by the analysis of more than 4600 hours of data collected from real firefighters in action, including 717 answers to event questionnaires from a total of 454 different events.**

I. INTRODUCTION

A first responder is a person trained to intervene in emergency situations in order to help the general population. More specifically, in the case of a firefighter (FF), he is trained to perform under dangerous and high-pressure situations that are both physically and psychologically demanding, such as: forest fires; vehicular accidents; or rescue missions. Recent studies have shown that these professionals have the highest occupational fatality rates in the U.S., mainly due to cardiovascular problems [1]. These problems can result from the long time exposures to several traumatic, non-traumatic, and organizational stressors over time [2], which are known to be associated with psychological and physical illness [3]. This motivates the need for new systems and technologies capable of monitoring, in real time, the physiological signals and the behaviour patterns of these professionals, in order to assess their stress levels.

With the goal of providing a first response monitoring system on critical emergency scenarios, the "Vital Responder" project (http://www.vitalresponder.pt) was created. It is based on a wearable shirt (Vital Jacket® in Fig. 1) that is capable to continuously collect electrocardiogram, accelerometers and GPS signals, that are sent to a base station, in real time, in order to be analysed. However, signal processing and machine learning research, aiming to combine these signals into an estimation of an individual's stress levels, requires the data to be complemented by adequate annotation that can contextualize it. Did an individual's heart rate rise because he was stressed? Or did he simply start running? Was he doing a training exercise or in a real dangerous life-threatening situation? For this purpose, we have created the Vital Analysis Framework (VA), which is a smartphone based solution capable of annotating physiological signals of FF in action with both context (details about the event the FF was in) and perceived psychological stress levels (retrieved from the analysis of psychological questionnaires). In this paper, we will focus on the later, in an attempt to understand if these questionnaires are gathering data with enough quality to achieve this objective, based on more than 4600 hours of data collected by FF during real life events. Furthermore, we will also show a simple usability analysis on the framework, to understand if our solution is used correctly by the FF.

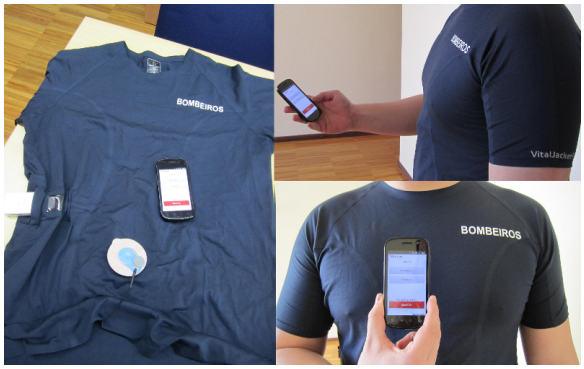

Fig. 1. Images of a version of the Vital Jacket®, specially made for the firefighters, and the Vital Analysis framework running on a smartphone.

In this paper, in Section II, we will explain how psychological stress levels are usually annotated. Section III will describe the Vital Analysis Framework workflow, focusing on the stress annotation questionnaires. In Section IV we will present some results followed by conclusions in Section V.

II. ANNOTATING STRESS LEVELS

When investigating stress, it seems important to acknowledge that over the last decades the term has increased popularity across different areas of study such as behaviour and health sciences. As a result, it remains difficult to define the concept, at least in simple terms [4], and therefore the necessity to use standardized measures of assessment. For the purpose of this paper, stress conceptualization will follow the transactional model of stress, defined as "a situation that taxes or exceeds one's personal resources or threatens the person well-being has the potential to cause stress" [5] (p.19). Thus, the emotions experienced and physiological responses

*This work was supported by FCT project CMU-PT/CPS/0046/2008

[1]P. Gomes, B. Lopes and M. Coimbra are with the Instituto de Telecomunicações in Faculdade de Ciências da Universidade do Porto, Portugal (email:{ptmgomes, brunolopes, mcoimbra}@dcc.fc.up.pt)

[2]M. Kaiseler is with Instituto de Telecomunicações in Faculdade de Engenharia da Universidade do Porto and the Laboratório de Reabilitação Psicossocial at Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto, Portugal (email:mkaiseler@fpce.up.pt)

[3]C. Queirós and M. Oliveira are with the Laboratório de Reabilitação Psicossocial from FPCEUP/ESTSIPP, Portugal (email:cqueiros@fpce.up.pt, monicaqoliveira@gmail.com)

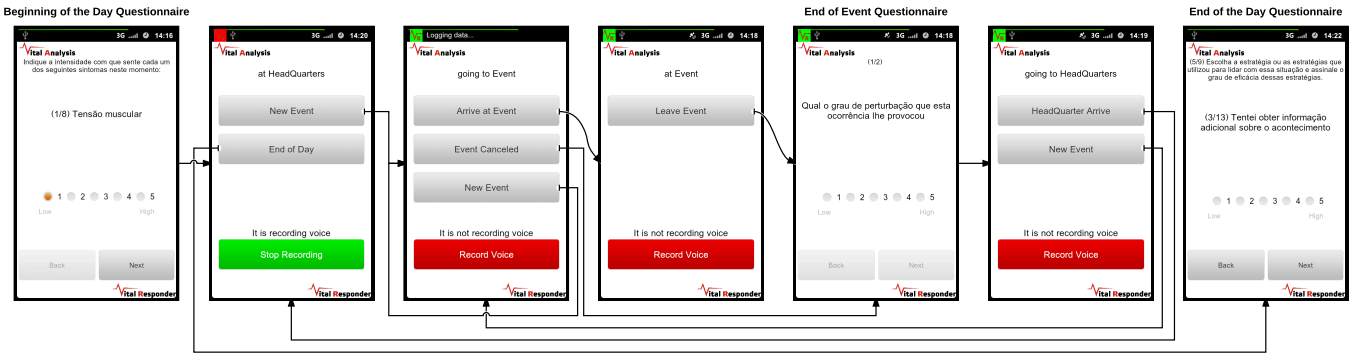

Fig. 2. System Image of the Vital Analysis application with real screenshots.

are initiated due to the individual interpretation of the event and its perceived meaning to their well-being [6].

Following these theoretical conceptualization of stress, its assessment should combine physiologic and psychological measures [7], including a longitudinal research designs, and decrease time delay between real world experience of event and stress ratings [5] since previous research methodologies rely mainly on self-report measures of the concept (e.g., questionnaires) and are fulfilled in paper several hours after the event [8]. To overcome issues related with recall errors, the current study will use the experience sampling method (ESM) developed for *in situ* recording [9]. ESM in the current study will ask firefighters to rate their stress levels at predetermined times (e.g. beginning, and end of the day and following an event). While ESM originally relied on paper surveys, for the purpose of the current study, ESM will be conducted using a smartphone based framework as successfully used in previous research [10].

III. VITAL ANALYSIS FRAMEWORK

Given that our target is firefighters in action during real events, we designed a simple to use smartphone framework that can adapt itself to their daily routines. Three annotation methodologies were designed that together provide context to an event and to the collected signals (screens depicted in Fig. 2). These are the *Stress Annotation Methodology* (stress levels annotation), the *Event-driven Annotation Methodology* (event context), and the *Voice Annotation Methodology* (a ”break the glass” mechanism).

A. Stress Annotation Methodology

Due to the complexity of the definition of the stress concept, several self-report measures exist. Despite differences in types of questions, those vary according to the context and population under study. Most self-report measures aiming to access stress levels include questions related with physical and cognitive symptoms of stress. Thus, following this principle, our measures of stress included 4 questions related with physical and 4 questions related with cognitive aspects, used previously in validated stress questionnaires [11]. An example of a physical symptom question is: ”I have a stiff neck”; an example of a cognitive symptom question is ”I lack concentration”. Participants were asked to rate each item on a free scale ranging from ”0” to ”4” kinds of ratings, where a rating of ”0” was used to represent not felt at all, and a rating of ”4” extremely felt. These questions were fulfilled at the beginning and end of the day, aiming to evaluate whether there were alterations in stress symptoms experienced, from beginning to end of the day. Furthermore, end of the day and beginning of the day stress symptoms mean scores will be subtracted in order to accomplish an overall mean score, symbolizing accumulated stress symptoms over the day. Internal consistency of the 8 questions was calculated using Cronbach’s alphas. This value provides a coefficient of reliability, and it is used as a measure of internal consistency for participants’ answers. As recommended by [12] these values should be above 0.80. In the current study, the Cronbach’s alphas found for the 8 questions was 0.93. Additionally, another question was fulfilled after each event, indicating stress appraisal of the event. Participants were asked to rate how they appraised each stressful event, on a free scale ranging from ”0” to ”4” kinds of ratings, where a rating of ”0” was not at all stressful, and a rating of ”4” was extremely stressful.

B. Event-driven Annotation Methodology

The *Event-driven Annotation* gives us the possibility to detail an event by dividing it into several predefined stages, allowing us to evaluate and quantify the collected physiological signals differently for each one. These predefined stages are the basic stages for every single event, and are usually consecutive. A normal event starts with as emergency call, followed by the trip to the event, the event itself, and finally the return trip to the headquarters. Nevertheless special occasions can occur such as: a high priority call during any period of other event; or the cancellation of an event. This workflow is represented in Fig. 2, where all the options available in the framework can be seen.

C. Voice Annotation Methodology

Motivated by the unpredictability of a firefighter’s job we have designed the Voice Annotation Methodology. This methodology will be our ”break the glass” mechanism, allowing that at least one of the firefighters can report unexpected activities that happen during an event, or add valuable psychological information, allowing for rich and

TABLE I

RESULTS FROM THE USABILITY EVALUATION, COMPILED FROM THE DATASET COLLECTED FROM THE FIREFIGHTERS (FF). WE HAVE ANALYSED ALL EVENTS (TOTAL), AS WELL AS WHEN ONLY A SPECIFIC AMOUNT OF FF WHERE PRESENT (COLUMNS 1FF..5FF), AND WHEN MORE THAN A SPECIFIC AMOUNT WAS PRESENT (COLUMNS >1FF..>3FF)

	Total	1FF	2FF	3FF	4FF	5FF	>1FF	>2FF	>3FF
Nr. of Events	454	259	151	29	6	6	192	41	12
% of Events with Annotation	53,5%	38,2%	69,5%	82,8%	83,3%	100,0%	72,9%	85,4%	91,7%
% of Events with good Annotation	64,2%	57,6%	68,6%	29,2%	40,0%	83,3%	61,4%	40,0%	63,6%
% of Annotated Events with Audio	14,8%	11,1%	10,5%	41,7%	20,0%	50,0%	17,9%	40,0%	36,4%
% of Questionnaires with Audio	20,2%	19,2%	13,3%	33,3%	40,0%	50,0%	19,3%	37,1%	45,5%

expressive contributions. This methodology is also used to allow the user to add annotation outside events and to enrich the data gathered using the previous methodologies.

IV. RESULTS

A. The Dataset

The Dataset compiled for this study was collected from 12 firefighters, with a mean age of 37.8 and a standard deviation of 5.3, between July, 2011 and January, 2012. During this time we have collected more than 4600 hours of Vital Analysis data from which we have retrieved a total of 717 answers to the event questionnaires from 454 different events. We have also retrieved 319 stress evaluations from the differences between the beginning of day and end of day questionnaires.

B. Usability Evaluation

To evaluate the usability of our framework we have collected the official information about the events, already gathered using today's firefighter protocol, and we compared it with our annotations. The measures chosen were: the percentage of real events that were annotated; the percentage of annotations that were done correctly, in which a correct (good) annotation is one that has all the stages implemented in the Event-driven Methodology (in subsection III-B), with a time difference between them above 1 minute; the percentage of annotated events with audio annotation; and the percentage of event questionnaires with audio annotation.

The results, as seen in Table I, show that our framework was used in 53.5% of all events and that 64.2% of them were correctly annotated. We can generically consider this as a good result, given the harsh environments that these FF have faced. Nevertheless, some results require a special attention. Low percentages of annotation are obtained when a single FF is sent to an event with VA, which was an expected result. Our solution requires that the FF by himself both remembers and has the time to use the VA which does not exploit the redundancy of the team. Another interesting result is the low percentage of good annotations in events where we have 3 or 4 firefighters with VA present. We would expect to see an increase in this percentage but an explanation might be that in situations when many men are deployed, the situation tends to be more serious and chaotic, making them less prone to use the system. Interestingly, when we have low percentages of good annotations, we have higher percentages of audio annotations, making us speculate that the firefighters were aware that they were not able to perform proper annotations, compensating this by giving us extra information after the event using audio annotation.

Globally, results support that our framework works well in these environments, either using the conventional methodology or the provided "break the glass" alternative.

C. Stress Annotation Validation

Table II provides the mean and standard deviation for the stress appraisal of various events. To analyse whether means for each stress category differed, One-Way Anova analysis was conducted. As expected, we found significant differences between the stress appraisal categorization of the various events. As can be observed by F value (F= 2.518) which is found by dividing the between group variance by the within group variance. When testing the statistical significance, the p-value found was 0.01. Since statistical significance should be equal or lower than 0.05 [12], this value confirms that the null hypothesis is rejected. It is important to highlight, that higher ratings of stress were provided for certain events such as fire, accidents and pre-hospital assistance. These findings are similar to those found in the literature using detailed psychological methods to assess stress among firefighters [13]. This fact is likely to support the accuracy of the measure to assess stress appraisal of FF across different events.

TABLE II

MEAN AND STANDARD DEVIATION (SD) VALUES OF THE STRESS APPRAISAL FOR EACH EVENT CATEGORY WHERE 0 IS MINIMUM AND 4 THE MAXIMUM.

Event categories	Mean	SD	Min	Max
Fire	1,23	1,002	0	4
Accident	1,40	0,857	0	3
Infrastructure/communications	0,50	0,798	0	2
Pre-hospital assistance	1,11	0,813	0	4
Legal conflict	0,80	0,837	0	2
Technological/Industrial	0,50	0,548	0	1
Services	1,01	1,033	0	3
Activities	0,77	0,725	0	2
Total	1,10	0,918	0	4

Table III shows the end of the day stress symptoms minus the beginning of the day stress symptoms mean scores and standard deviations, repeated measures T-Test on the

beginning of the day and end of the day cognitive and physical stress symptoms, and correlations between stress appraisal measure and each question of physical and cognitive stress symptom. As expected, mean values found, show a high mean value of stress symptoms across all physical and cognitive stress symptoms questions. These findings, suggest that physical and cognitive stress symptoms have accumulated for FF over the working day, probably due to experience of stress appraised over events.

In order to statistically test whether, there was a difference in mean scores from beginning to end of the day physical and cognitive stress symptoms a Repeated measures T-Test was conducted. Results show that physical and cognitive symptoms of stress increased significantly from the beginning until the end of the day ratings. Again, these results are similar to the literature in the area, suggesting that firefighter's experience an increased in symptoms of physical and cognitive stress by the end of the day [14], probably due to stress experienced over events during the day. These findings also support the accuracy of the questions used, to assess physical and cognitive symptoms of stress among firefighters.

TABLE III

END OF THE DAY STRESS SYMPTOMS MINUS THE BEGINNING OF THE DAY STRESS SYMPTOMS MEAN SCORES (MEAN) AND STANDARD DEVIATIONS (SD), REPEATED MEASURES T-TEST ON THE BEGINNING OF THE DAY AND END OF THE DAY COGNITIVE AND PHYSICAL STRESS SYMPTOMS (T-TEST), AND CORRELATIONS BETWEEN STRESS APPRAISAL MEASURE AND EACH QUESTION OF PHYSICAL AND COGNITIVE STRESS SYMPTOM (R. PEARSON)

Categorization of questions	Mean[a]	SD	T-Test[a]	R. Pearson
Stiff neck	0,433	0,66	12,085[1]	0,132[2]
Tiredness in the eyes or heavy head	0,453	0,75	11,256[1]	0,132[2]
Uncomfortable abdominal pain or stomach ache	0,178	0,55	5,986[1]	0,241[3]
Difficulty to keep the body straight	0,281	0,6	8,842[1]	0,239[3]
Lack of Concentration	0,269	0,59	8,544[1]	0,252[3]
Difficulty to think, and make decisions	0,243	0,54	8,314[1]	0,296[3]
Anxiety	0,192	0,62	5,777[1]	0,081
Difficulty in controlling reactions	0,221	0,54	7,613[1]	0,242[3]

a. (end day - beginning day)
1. p = 0; 2. p $<$ 0,05; 3. p $<$ 0.01

Finally, correlations between mean score for symptoms of physical and cognitive stress questions, shows strong positive correlations with the stress appraisal of events, suggesting that a firefighter's experience of stress symptoms at the end of the day is strongly associated with an increased stress appraisal of events experienced during the day. Thus, the associations found between these measures, are not only in line with previous findings in the area [3], but also gives appropriateness of measures used to access the stress concept [5].

V. CONCLUSION

Overall, the Vital Analysis framework was well accepted and highly used by the firefighters in their daily routines. Results show that the questions used were reliable, and accurate enough to assess physical, cognitive symptoms of stress, and stress appraisal over events. This confirms the success of the proposed framework's ability to support the annotation of physiological signals with the stress levels of the user. We also believe, that these findings culminate previous research limitations observed in the past when accessing stress, and the observed successful measurement of the stress concept found in this study, encourages researchers to integrate a combination of physiologic and psychological measures when investigating the concept of stress among first respondents under real world conditions. Thus, findings are likely to impact future research in this area, and can also be used to design not only more efficient practical interventions, but also larger epidemiological studies.

ACKNOWLEDGMENT

We are thankful to Cmdt. Jorge Rocha, from Bombeiros Voluntários de Amarante, as well as to Cmdt. Eng. Salvador Almeida, from Companhia de Bombeiros Sapadores de Gaia, and their teams, for their collaboration.

REFERENCES

[1] A. Fabio, M. Ta, S. Strotmeyer, W. Li, , and E. Schmidt, "Incident-level risk factors for firefighter injuries at structural fires," *J Occup Environ Med.*, vol. 44, pp. 1059–63, Nov 2002.

[2] D. A. Alexander and S. Klein, "First responders after disasters: a review of stress reactions, at-risk, vulnerability, and resilience factors," *Prehospital and disaster medicine*, vol. 24, no. 2, pp. 87–94.

[3] C. A. Reynolds and S. L. Wagner, "Stress and first responders: The need for a multidimensional approach to stress management," *International Journal of Disability Management*, 2008.

[4] S. Levine, *Handbook of Stress and the Brain*, ch. Stress: An historical perspective, pp. 3–23. Elsevier, 2005.

[5] R. S. Lazarus and S. Folkman, *Stress, Appraisal, and Coping*. Springer, 1984.

[6] W. R. Lovallo, *Stress & health: Biological and psychological interactions*. Sage Publications, 2005.

[7] N. Semmer, S. Grebner, and A. Elfering, *Research in Occupational Stress and Well-being*, vol. 3, ch. Beyond Self-report: using observational, physiologival, and situation-based measures in research on ocuppational stress, pp. 205–263. Emerald Group Publishing, 2003.

[8] S. Cohen, R. Kessler, and L. Gordon, *Measuring stress: a guide for health and social scientist*. Oxford University Press, 1997.

[9] J. Hektner and M. Csikszentmihalyi, *Handbook of environmental psychology*, ch. The experience sampling method: Measuring the context and content of lives. John Wiley Sons, Inc., 2002.

[10] P. Klasnja et all, "Using wearable sensors and real time inference to understand human recall of routine activities," in *UBIComp'08*, 2008.

[11] S. Cohen and G. Williamson, *The Social Psychology of Health*, ch. Perceived Stress in a Probability Sample of the United States, pp. 31–67. Sage Publications, 1988.

[12] B. Tabachnick, *SAS for windows workbook for Tabachnick and Fidell using multivariate statistics*. Allyn and Bacon, 2001.

[13] P. Barnes, "The experience of traumatic stress among urban firefighters," *Australian Journal of Emergency Management*, vol. 15, no. 4, pp. 50–64, 2000.

[14] M. Chamberlin and H. Green, "Stress and coping strategies among firefighters and recruits," *Journal of Loss and Trauma*, vol. 15, pp. 548–560, 2010.