Physiological Characterization of Need for Assistance in Rescue Missions with Drones

(Invited Paper)

Fabio Dell'Agnola, Leila Cammoun, David Atienza
Embedded Systems Laboratory (ESL)
Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
Email: {fabio.dellagnola, leila.cammoun, david.atienza}@epfl.ch

***Abstract*—The use of drones is recently gaining particular interest in the field of search and rescue. However, particular skills are still required to actively operate in a mission without crashing the drone. This limits their effective and efficient employment in real missions. Thus, to assist the rescuers operating in stressful conditions, there is a need to detect an increase of workload that could compromise the outcome of the mission. In this work a simulator is designed and used to induce different levels of cognitive workload related to search and rescue missions. Physiological signals are recorded and features are extracted from them to estimate cognitive workloads. The NASA Task Load Index is used as subjective self-report workload reference. Then, performance is recorded to objectively evaluate the execution of the tasks. Finally, the analysis of variance (ANOVA) is used to verify intra- and inter-subject variability. Results show statistical decrease of the mean normal-to-normal (NN) interval with an increase of cognitive workload. Moreover, it is observed a decrease of performance while an increase of cognitive workload exists. This information can be used to detect the need for assistance.**

I. Introduction

The use of Unmanned Aerial Vehicles (UAVs), or drones, was originally limited to the military sphere. Subsequently, thanks to their commercial accessibility and their large degree of versatility, the utilization has exploded in different other fields (e.g., aerial mapping, search and rescue, transportation and delivery, or even for private leisure [1]). Moreover, enhancements in technology allow the drones to be used to collect, process, store and relay data on their own and in real time, thus transforming them in some kind of evolved smart multi-sensor networks [2–4].

Nowadays, drones are extremely interesting devices for search and rescue applications, where collecting information is of primary importance. Once a disaster occurs, rescuers have to gather information to immediately evaluate each emergency situation upon arrival. This first task is crucial to take the proper decisions and effectively master the emergency. Gathering information from a chaotic place is hard and time consuming, because accesses are often difficult and resources relatively limited. A drone, or a network of drones, can easily facilitate and accelerate this task, especially providing information that is not available from a ground perspective. Moreover, drones can be used to establish a communication with victims, or to provide them first assistance (e.g., with water, oxygen, or moral support).

Even though drones find many applications in the field of search and rescue, limitations exist in their effective and efficient utilization in real missions. The reason is that employing drones in stressful conditions remains a challenge, even if driving a drone in controlled environments has become quite straightforward. Operating in extreme conditions, dealing with the scarcity of human resources, and having the feeling of urgency in finding victims, demands an important cognitive effort. In such conditions, simple tasks as driving a drone become more complicated. Therefore, the risk of not being able to perform the mission or to crash the drone is extremely high.

To address this problem, researchers started to protect the drones with cages that prevent crashes [5], developed shared controllers that modulate the level of assistance [6], and focused on embodied interaction that facilitate their control [7]. However, there is a need to ensure efficient interaction between rescuers and drones. Thus, the online monitoring of the cognitive workload of the rescuers has to be considered to dynamically adapt the level of assistance, in order to ensure that missions are efficiently performed, and to avoid crashes of drones.

The monitoring of cognitive workload has been widely studied in the past. Indeed it has been proven that changes in cognitive level of workload are visible in physiological signals [8]. Therefore, characterizing the cognitive workload level from physiological signals is the key to develop a minimally intrusive wearable interface in order to estimate a possible need of assistance.

In this context, different works show how to characterize cognitive workload from physiological signals. For instance, multi-tasks activities, such as, auditory, mental arithmetic, and memory tasks can be combined to induce different levels of cognitive workload that are visible in physiological signals [9], [10]. Based on the similar principle, the Multi-Attribute Task Battery (MATB-II) has been developed to evaluate operator performance and workload [11]. Artificially combining multi-tasks activities is the first step to properly characterize cognitive workload. This allows a design of an experiment in a controlled environment, reducing the amount of undesired external factors. Then, with a better knowledge of the problem, it is possible to extend the study in the applied field. As an example, researchers studied fighter pilots during simulations [12–14], and real flights [15–17]. Other investigated the roots of the problem and validated their hypothesis by studying car drivers during real-world driving tasks [18], and people in everyday-life office-work scenario [19], [20]. Recent studies moved to physical human robot interaction related to rehabilitation activities [21]. However, to our knowledge there is no study that explore physiological changes due to cognitive workload induced while performing search and rescue missions employing drones.

Although the subject is widely covered in the literature, characterizing cognitive workloads from physiological signals is not properly addressed for drone pilots. Physiological signals exhibit high intra- and inter-subject variability as a result of age, gender, time of day and other factors [22]. Therefore, it is important to characterize the physiological effects related to the specific case study of rescue missions with drones. Thus, this work contribute by:

- designing a drone simulator to induce cognitive workload related to search and rescue missions;
- selecting physiological features that give a dynamic indication of rescuers' internal state in the context of search and rescue missions employing drones;
- showing a decrease of performance while increasing the level of cognitive workload, information that can be translated into a possible need for assistance.

The rest of the paper is organized as follows. In Section II we introduce the method used to access cognitive workload from physiological signals. In Section III we describe the experimental setup. Then, in Section IV we present the experimental results that validate the approach. Finally, in Section V we draw the main conclusions of this work.

II. METHOD FOR COGNITIVE WORKLOAD CHARACTERIZATION

A block diagram of our proposed method is shown in Fig. 1. The simulator (Sec. III-A) contains several tasks (Sec. III-B), with a particular level of cognitive workload, that the recruited subjects (Sec. III-E) perform. The execution of the tasks is objectively evaluated by monitoring the performance (Sec. III-C) of the subjects. The level of cognitive workload is self evaluated by the subjects after each task, based on the National Aeronautics and Space Administration Task Load Index (NASA-TLX) [23]. The workload affects the physiological signals that are constantly recorded. Then, features (Sec. II-A) are extracted from the physiological signals and compared with both performance and NASA-TLX. Finally, a complete data analysis (i.e., Analysis of Variance or ANOVA) is used to verify intra- and inter-subject variability [24].

A. Signal Acquisition and Feature Extraction

We record Electrocardiogram (ECG), in particular the lead II. From the ECG, the so called normal-to-normal (NN) intervals are extracted. Then, the Heart Rate Variability (HRV), the variations of the NN intervals, are analyzed in both time and frequency domains as described in [25].

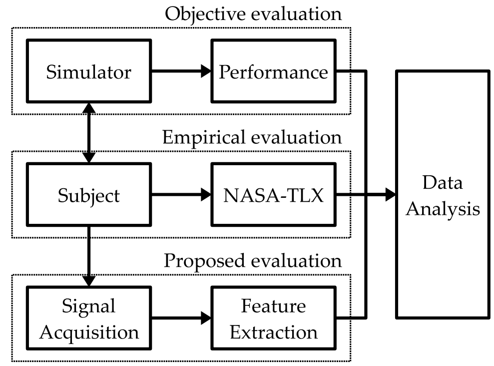

Fig. 1. Block diagram showing the method used to characterize cognitive workload from physiological signals.

B. Time-domain features

For the time-domain analysis, we considered both statistical and geometrical method.

Regarding statistical methods, the mean NN interval and the square root of the mean squared differences of successive NN intervals (RMSSD) are calculated. The number of interval differences of successive NN intervals greater than 50 ms (NN50), and the proportion derived by dividing NN50 by the total number of NN intervals (pNN50), are not considered because of their inferior statistical properties compared to RMSSD [25]. The standard deviation of NN intervals (SDNN), the standard deviation of the average NN interval calculated over short periods (SDANN), and the SDNN index are not considered as well, being all of them appropriate for long-term recordings only.

Concerning geometrical methods, we extract different features from the Poincaré (or Lorenz) plot indicating vagal and sympathetic function [26], such as, the length of the transverse axis (T), which is vertical to the line $\mathrm{NN}_k = \mathrm{NN}_{k+1}$; the length of the longitudinal axis (L), which is parallel with the line $\mathrm{NN}_k = \mathrm{NN}_{k+1}$; the ratio L/T, called Cardiac Sympathetic Index (CSI); the modified CSI (L^2/T); and the $log_{10}(L \cdot T)$, called Cardiac Vagal Index (CVI) [25], [26]. On the other hand, the HRV triangular index and the triangular interpolation of NN interval histogram (TINN) are not considered, as they are inappropriate to access short-term changes in HRV [25].

C. Frequency-domain features

For the frequency-domain analysis of HRV, we first estimate the Lomb-Scargle Power Spectral Density (PSD) of the NN intervals [27], and then we compute the power in the low-frequency band (LF) between 0.04 and 0.15 Hz, in the high-frequency band (HF) between 0.15 and 0.4 Hz, and the ratio of the two, referred to as the LF/HF ratio. LF and HF are normalized by dividing the values by the total power minus the very-low-frequency (VLF) component ($\leq$ 0.04 Hz) [25].

III. Experimental validation

A. Rescue simulator

For this study we designed and implemented a flight simulator with Unity3D [28]. The simulator aims to reproduce a rescue scenario where the pilot of a drone has to deal with two different activities, namely, flying and mapping.

The flying activity consists in flying a drone following a precise path. The flight path is showed by 90 waypoints (white clouds) distributed over a village every 20 m along a randomly generated trajectory.

The mapping activity consists in mapping a damage situation of a disaster area. In the simulator, the damage situation is represented by cubes of 4 different colors randomly distributed over the flying trajectory. The colors are: yellow to indicate rescue situations, red for fire, blue for water damages, and green for accidents. Colors are chosen according to the regulation of the Swiss Firefighters [29].

With a modulation of both flying and mapping activities, it is possible to induce different levels of workload. The same principle is applied for MATB-II, where different tasks are combined to induce different levels of workload [11]. The combination used in this study is explained next.

B. Tasks inducing cognitive workload

To induce different levels of workload, both flying and mapping activities are combined yielding in five different tasks: Baseline (B), Training (T), Flying (F), Flying and Mapping 1 object (F1M), and Flying and Mapping 3 objects (F3M). The tasks F, F1M, and F3M are designed to have low, medium, and high cognitive workload, respectively. The tasks are described as follows:

1) Baseline (B): As baseline we consider a flying sequence controlled by an auto-pilot. The speed of the drone is constant at 6m/s (same for all the other tasks). No mapping activity is required during this sequence. For this task, the subjects are instructed to relax and to simply watch the sequence. This sequence puts the subjects in a framework that is the same for the entire experiment, avoiding as much as possible changes of uncontrollable variables.

2) Training (T): To get confident with the simulator, a training sequence is presented. The sequence is characterized by a mix of both flying and mapping activities. For this task, the number of objects to be mapped are 60 per session, i.e., 15 per color. During this sequence, the subjects are asked to fly as close as possible through the center of the waypoints, and to press the button of the controller relative to the color of the objects that randomly appear.

3) Flying (F): The low level of workload is the flying activity alone, without mapping. During this task, the subjects are asked to control the drone and to fly as close as possible through the center of the waypoints.

4) Flying and Mapping 1 object (F1M): The medium level of workload is the same as training, including the same combination of both flying and mapping activities. As in the training task, the subjects are asked to fly as close as possible through the center of the waypoints, as well as pressing the button of the controller relative to the color of the objects that randomly come into sight.

5) Flying and Mapping 3 objects (F3M): The high level of workload is again a combination of both flying and mapping activities but with a more demanding mapping activity. For this task, the objects displayed at the same time on the screen are three, and not only one as in the F1M or T. The total number of objects to be mapped are 240 per session, i.e., 60 per color. Again, the subjects are asked to fly as close as possible through the center of the waypoints, and to press the button of the controller relative to the color of the objects that become visible.

C. Performance metrics

The performance (ρ) of the pilot is evaluated based on both flying and mapping activity, as shown in Eq. 1:

$$\rho = 1 - \frac{1}{4}(t_{\mathrm{d1}} + t_{\mathrm{d2}} + e_{\mathrm{m}} + t_{\mathrm{r}}). \tag{1}$$

t_{d1} is the root-mean-square of the time delay, which is the difference between the time actually needed to fly through two subsequent waypoints and the optimal time. Being the waypoints equidistant (20 m) and the speed of the drone constant (6 m/s), the optimal time is a constant as well, which was computed by dividing the distance of two waypoints with the speed of the drone. t_{d2} is the root-mean-square of the distance between the drone and the center of the waypoint divided by the speed of the drone. This distance lays in the plane perpendicular to two subsequent waypoints and tells how close to the waypoints the subject was able to fly. e_{m} is the error detection rate for the mapping activity and t_{r} is the reaction time in detecting the objects.

Finally, before combining them, all the metrics are normalized to have values between 0 and 1.

D. Study protocol

During the experiment, the subjects seated in a room in front of a screen and played the simulator with a Gamepad from Logitech [30]. They were asked to not talk and to avoid as much as possible any kind of unnecessary movements during the tasks, but they were free to rest and move otherwise. After each task, they reported a subjective cognitive workload estimation based on the NASA-TLX procedure.

The lead II of the ECG was recorded in a noninvasive way through the Biopac MP160 data acquisition system [31]. The signal was acquired with Ag/AgCl snap electrodes. From the ECG, sampled at 2 kHz, the NN intervals were determined with AcqKnowledge 5.0 [32]. The HRV analyses was performed using Matlab R2016a (The MathWorks Inc., Natick, Massachusetts). An overview of the experimental setup is shown in Fig. 2.

A first subject participated in a preliminary experiment over three different days. The aim of the study was to evaluate physiological intra-variability caused by external factors. During one experimental session, the subject repeated eight times a task with a specific workload level interleaved with a baseline

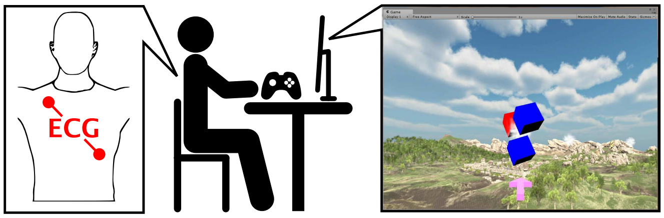

Fig. 2. Setup of the experiment. The subject seats in front of a screen and interacts with the simulator through a gamepad. The electrodes placement on the chest for ECG monitoring is depicted on the left. On the right, a screenshot of the F3M task, namely, flying and mapping 3 objects is shown. The arrow on the image indicates the direction of the next waypoint (white cloud), and the cubes are the objects to be detected by pressing a button of the controller.

B	F	···	B	F

Fig. 3. Protocol to evaluate physiological intra-variability caused by external factors. A baseline (B) sequence is interleaved with one level of workload (i.e., F, F1M, or F3M) and repeated eight times. Here an example with the flying (F) task.

B	T	T	F	F	F1M	F1M	F3M	F3M	B

Fig. 4. Protocol of the full experiment to assess inter-variability. The subjects start with a baseline (B), continue with two training (T) sessions, move to the flying and mapping tasks (F, F1M, and F3M), and end with a baseline again.

sequence (see Fig. 3). Each task was 5 minutes long. The level of workload was constant for one session, but differed between days. The subject was confident with the simulator, then no training sessions was proposed. For the simulator, the speed of the drone was kept constant at 6m/s for the entire experiment.

Afterwards, we extended the study to different subjects to assess the inter-variability. All the recruited subjects participated in the full experiment twice, on two different days. The subjects performed all the tasks (5 min. each), as shown in Fig. 4. Apart from the baseline that was recorded once at the beginning and once at the end of the experiment, all the other tasks were repeated twice in a row.

E. Subjects

Seven young subjects (4 males and 3 females) aged between 25 and 37 years old (27.7 ± 4.2), volunteered to participate in the study and provided informed consent before participating. The subjects were healthy, free of any cardiac abnormalities and were receiving no medical treatment.

IV. Experimental Results

The results are obtained from two separate analysis. The first one, the intra-subject analysis shows the effect of the day on the physiological measurements on the same subject. The second analysis shows the effect of induced cognitive workload on different subjects.

A. Intra-subject analysis confirms the need for normalization

The results of the intra-subject analysis are based on a task with one level of workload interleaved with a baseline and repeated 8 times. Our analyzed data in Fig. 5 comes from one subject only, who repeated the experiment three times in three different days, each day with a different level of workload. The one-way ANOVA showed that the effect of the day on physiological measurements is significant ($p<0.05$) for most of the features extracted from ECG: RMSSD, CSI, CVI, LF, HF, and LF/HF. This indicates a significant difference on the baseline, thus a need exists for normalization. This finding is in line with the normalization requirement proposed in [21].

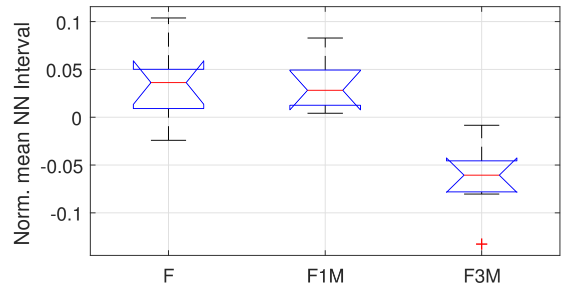

Fig. 5. Intra-subject analysis showing the trend of the normalized mean NN interval for the three induced cognitive workloads F ,F1M, and F3M. A one-way ANOVA shows that the effect of the induced level of workload is significant on the mean NN interval F(2,21) = 21.6, Prob>F = 8.0247e-06.

After normalization of the physiological measurements by the baseline, the one-way ANOVA shows that the effect of the induced level of workload is significant ($p<0.05$) on the mean NN interval, CSI, LF, HF, and LF/HF. Although the study focuses in a different context, similar results are reported in [33]. Fig. 5 shows a decreasing trend of the normalized mean NN interval while increasing the cognitive workload. Therefore, changes in cognitive workload related to search and rescue missions are visible in physiological signals.

B. Mean NN interval characterizes workloads

The results of the inter-subject analysis are based on the data collected from all the seven subjects. All the measurements are normalized by the baseline to avoid the effect of the day. A two-way ANOVA is conducted on the influence of two independent variables (cognitive workload level and subject) on the features extracted from the physiological signals to estimate the level of cognitive workload. Cognitive workload includes three levels (i.e., F, F1M, and F3M).

The main effect for cognitive workload yields a F ratio of $F(2, 63) = 3.7$, Prob>F = 0.03 on the mean NN interval, indicating a significant difference between the tasks F1M and F3M (p-value = 0.0227). Therefore, it is possible to use the mean NN interval to detect changes when the mapping activity becomes more demanding. The mean NN interval was used in other studies showing similar changes [33], [34]. However, it is difficult to distinguish the task F from F1M (p-value = 0.3560), or F from F3M (p-value = 0.3810). F1M is characterized by a relative simple mapping task, which probably does not seem to be perceived as having different complexity from F. This could explain the fact that no significant differences are visible on the mean NN interval. Moreover, a lack of training could also explain the physiological response. Thus, our analysis

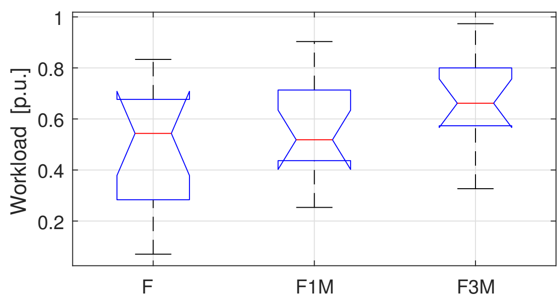

Fig. 6. Influence of induced cognitive workload level on the reported cognitive workload level. The NASA-TLX shows a consistent, but non-significant difference of the reported overall workload between the tasks F and F3M, F(2,39) = 2.98, Prob>F = 0.0622.

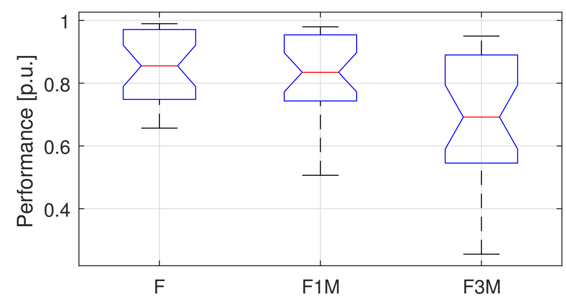

Fig. 7. Influence of cognitive workload level on global performance, which considers both flying and mapping performances. A one-way ANOVA shows a significant difference between F and F1M, and between F and F3M, F(2, 81) = 6.23, Prob>F = 0.0031.

indicates that clearly a bigger effort is required to manage flying related tasks that are not well known. This is the case of F, which is performed after only 10 minutes of training. Therefore, a longer training or a randomization of the order of the tasks is expected to reduce this effect.

The effect for subject yields a F ratio of $F(6, 63) = 29.62$, Prob>F = 0, indicating that the effect for subject is significant. The interaction effect between subjects and workload is not significant, $F(12, 63) = 0.69$, Prob>F = 0.75. Although there is a significant variation between the subjects, the mean NN interval can be used to characterize changes in cognitive workload. Moreover, no subject adaptation is required because of the absence of significant interaction between subjects and workload.

Although the intra-subject analysis shows a significant difference in many features, such as, CSI, LF, HF, and LF/HF, this is no longer the case when considering all the subjects of the study (inter-subject analysis).

These results indicate that a very limited number of features can be sufficient to perform workload characterization in rescue missions with drones for different subject. This conclusion is in contrast with other studies [34], [35], which cover physiological responses in other robotics-related fields, where a larger set of features are required to differentiate from subject to subject.

C. *Induced workload perceived by participants*

A one-way ANOVA is conducted on the influence of induced cognitive workload level on the reported cognitive workload level. The main effect for the performance yields a F ratio of $F(2, 39) = 2.98$, Prob>F = 0.0622, indicating a consistent, but non-significant difference between the tasks F and F3M. Thus, this trend, also shown in Fig. 6, indicates that participants perceived the induced cognitive workload levels as intended from the experiment design. However, the trend of the median lines showed in Fig. 6 is similar to the one of the mean NN interval (Sec. IV-B). This validates the results obtained with the inter-subject analysis, namely, physiological signals are affected by workload changes related to simulated rescue missions.

D. *Performance affected by increase of workload*

A one-way ANOVA is conducted on the influence of cognitive workload level on the flying and mapping performance. The main effect for the performance yields a F ratio of $F(2, 81) = 6.23$, Prob>F = 0.0031, indicating a significant difference between the tasks F1M and F3M (p-value = 0.0343), and between F and F3M (p-value = 0.0031), but not from F and F1M (p-value = 0.6723). A visual representation of the results is shown in Fig. 7.

Some other studies link highly aroused stress states with degraded performance [18]. Therefore, our analysis confirms that with increases of cognitive workload, there is always a significant decrease of performance. In accordance with the physiological measurements and the self-reported cognitive level of workload, this decreasing trend of performance validates the hypothesis that physiological signals can be used to detect a possible need for assistance in rescue missions with drones.

V. Conclusion

To assist the rescuers operating in stressful conditions of rescue missions with drones, there is a need to detect an increase of workload that could compromise the outcome of the mission. Thus, characterizing the cognitive workload level from physiological signals is the key to develop a minimally intrusive wearable interface for estimation of need for assistance. To address the problem, we first design a drone simulator that is used to induce cognitive workload related to search and rescue missions. Then, we select physiological features that give a dynamic indication of rescuers' internal state in the context of search and rescue missions employing drones.

Our results show a statistical decrease of the mean NN-intervals with an increasing of workload. Therefore, changes in cognitive workload are visible in physiological signals. The NASA-TLX confirms that the participants perceived the induced workload levels as intended from the experiment design. Moreover, we show a decreasing of performance with an increase of cognitive workload. Our results validate the hypothesis that physiological signals can be used to detect a possible need for assistance.

In conclusion, our experimental results yield to lay the foundations to design a wearable embedded system able to detect a need for assistance. This can prevent a possible drop of performance that could compromise the outcome of a search and rescue mission.

Being this experiment based on a simulated and controlled environment, the subjects were not exposed to the same stressful conditions as they would be in the case of real search and rescue missions. In particular, no real drone was employed, there was no fire, and no one was in real danger. Therefore, there is a need to further investigate unexpected physiological changes in the field during real-life rescue missions with drones, in order to estimate the possible benefits of person-specific thresholds indicating the need for assistance.

ACKNOWLEDGMENT

The authors would like to thank Alexandre Cherpillod from the Laboratory of Intelligent Systems (LIS) at EPFL for providing a first version of the drone simulator that was modified by us for the purpose of this experiment.

This work has been partially supported by the NCCR Robotics through the Symbiotic Drone project, and by the ONR-G through the Award Grant No. N62909-17-1-2006.

REFERENCES

[1] M. A. Goodrich and A. C. Schultz, "Human-Robot Interaction: A Survey," *Foundations and Trends® in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2007.

[2] K. Daniel, B. Dusza, A. Lewandowski, and C. Wietfeld, "Airshield: A system-of-systems muav remote sensing architecture for disaster response," *2009 IEEE International Systems Conference Proceedings*, pp. 196–200, 2009.

[3] M. Quaritsch, K. Kruggl, D. Wischounig-Strucl, S. Bhattacharya, M. Shah, and B. Rinner, "Networked UAVs as aerial sensor network for disaster management applications," *Elektrotechnik und Informationstechnik*, vol. 127, no. 3, pp. 56–63, 2010.

[4] L. Y. Sørensen, L. T. Jacobsen, and J. P. Hansen, "Low cost and flexible UAV deployment of sensors," *Sensors (Switzerland)*, vol. 17, no. 1, pp. 1–13, 2017.

[5] A. Briod, P. Kornatowski, J.-C. Zufferey, and D. Floreano, "A Collision-resilient Flying Robot," *Journal of Field Robotics*, vol. 31, no. 1556-4967, pp. 496—-509, 2014.

[6] T. Carlson, R. Leeb, R. Chavarriaga, and J. D. R. Millán, "Online modulation of the level of assistance in shared control systems," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, no. 2, pp. 3339–3344, 2012.

[7] A. Cherpillod, S. Mintchev, and D. Floreano, "Embodied Flight with a Drone," *CoRR*, vol. abs/1707.0, 2017.

[8] B. Cain, "A Review of the Mental Workload Literature. Toronto," *Defence Research and Development Canada*, no. 1998, 2007.

[9] A. F. Kramer, "Physiological metrics of mental workload: A review of recent progress," *Multiple-task performance*, no. June, pp. 279–328, 1990.

[10] A. Henelius, K. Hirvonen, A. Holm, J. Korpela, and K. Muller, "Mental workload classification using heart rate metrics," *Conference proceedings : IEEE Engineering in Medicine and Biology Society*, vol. 2009, pp. 1836–1839, 2009.

[11] Y. Santiago-Espada, R. R. Myer, K. A. Latorella, and J. R. Comstock, *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research : A User ' s Guide NASA/TM2011-217164*, 2011, no. July.

[12] G. F. Wilson, J. D. Lambert, and C. A. Russell, "Performance Enhancement With Real-Time," in *Human Factors and Ergonomics Society 46th Annu. Meeting*, vol. 3, Santa Monica, CA, 1999, pp. 61–64.

[13] J. A. Veltman and A. W. Gaillard, "Physiological workload reactions to increasing levels of task difficulty," *Ergonomics*, vol. 41, no. 5, pp. 656–669, 1998.

[14] M. De Rivecourt, M. N. Kuperus, W. J. Post, and L. J. Mulder, "Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight," *Ergonomics*, vol. 51, no. 9, pp. 1295–1319, 2008.

[15] G. F. Wilson, P. Fullenkamp, and I. Davis, "Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions," *Aviation Space and Environmental Medicine*, vol. 65, no. 2, pp. 100–105, 1994.

[16] G. F. Wilson, "An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures," *The International Journal of Aviation Psychology*, vol. 12, no. 1, pp. 3–18, 2002.

[17] M. J. Skinner and P. A. Simpson, "Workload Issues in Military Tactical Airlift," *The International Journal of Aviation Psychology*, vol. 12, no. 1, pp. 79–93, 2002.

[18] J. A. Healey and R. W. Picard, "Detecting Stress During Real-World Dring Tasks Using Physiological Sensors.pdf," vol. 6, no. 2, pp. 156–166, 2005.

[19] C. Kappeler-Setz, "Multimodal Emotion and Stress Recognition," Ph.D. dissertation, ETH, 2012.

[20] B. Cinaz, B. Arnrich, R. La Marca, and G. Tröster, "Monitoring of mental workload levels during an everyday life office-work scenario," *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 229–239, 2013.

[21] D. Novak, B. Beyeler, X. Omlin, and R. Riener, "Workload estimation in physical human-robot interaction using physiological measurements," *Interacting with Computers*, vol. 27, no. 6, 2015.

[22] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with Computers*, vol. 24, no. 3, pp. 154–172, 2012.

[23] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Ed. Amsterdam: North Holland Press., 1988, pp. 139–183.

[24] D. A. Freedman, "Statistical Models: Theory and Practice," *Cambridge University Press*, p. 442, 2009.

[25] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use." *European Heart Journal*, vol. 17, pp. 354–381, 1996.

[26] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku, "A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R-R interval." *Journal of the autonomic nervous system*, vol. 62, pp. 79–84, 1997.

[27] W. H. Press and G. B. Rybicki, "Fast Algorithm for Spectral Analysis of Unevenly Sampled Data," *Astrophysical Journal, Part 1*, vol. 338, pp. 277–280, 1989.

[28] "Unity3D." [Online]. Available: https://unity3d.com/

[29] D. Goepfert, R. Karlen, M. Hartmann, G. Stäheli, S. Enz, H. Cina, J. Signer, M. Thalmann, M. Knöri, H. Benz, and P. Zurkirchen, *Reglement Einsatzführung*. Bern: Feuerwehr Koordination Schweiz FKS, 2015.

[30] Logitec, "Gamepad F310." [Online]. Available: http://support.logitech.com/en/product/gamepad-f310

[31] Biopac, "MP160 Data Acquisition Systems." [Online]. Available: https://www.biopac.com/product/mp150-data-acquisition-systems/

[32] ——, "AcqKnowledge Software." [Online]. Available: https://www.biopac.com/product/acqknowledge-software

[33] B. Cinaz, R. La Marca, B. Arnrich, and G. Tröster, "Monitoring of mental workload levels," *Proceedings of the IADIS International Conference e-Health 2010, EH, Part of the IADIS Multi Conference on Computer Science and Information Systems 2010, MCCSIS 2010*, pp. 189–193, 2010.

[34] D. Novak, M. Mihelj, and M. Munih, "Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction," *Robotica*, vol. 29, no. 03, pp. 367–374, 2011.

[35] ——, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with Computers*, vol. 24, no. 3, pp. 154–172, 2012.