

# Human-Machine Interaction Assessment by Neurophysiological Measures: a Study on Professional Air Traffic Controllers

Pietro Aricò, Maxime Reynal, Jean-Paul Imbert, Christophe Hurter, Gianluca Borghini, Gianluca Di Flumeri, *Student Member, IEEE*, Nicolina Sciaraffa, *Student Member, IEEE*, Antonio Di Florio, Michela Terenzi, Ana Ferreira, Simone Pozzi, Viviana Betti, Matteo Marucci, Enea Pavone, Alexandru C. Telea, Fabio Babiloni, *Member, IEEE*

**Abstract**— This study aims at investigating the possibility to employ neurophysiological measures to assess the human-machine interaction effectiveness. Such a measure can be used to compare new technologies or solutions, with the final purpose to enhance operator’s experience and increase safety. In the present work, two different interaction modalities (Normal and Augmented) related to Air Traffic Management field have been compared, by involving 10 professional air traffic controllers in a control tower simulated environment. Experimental task consisted in locating aircrafts in different airspace positions by using the sense of hearing. In one modality (i.e. “Normal”), all the sound sources (aircrafts) had the same amplification factor. In the “Augmented” modality, the amplification factor of the sound sources located along the participant head sagittal axis was increased, while the intensity of sound sources located outside this axis decreased. In other words, when the user oriented his head toward the aircraft position, the related sound was amplified. Performance data, subjective questionnaires (i.e. NASA-TLX) and neurophysiological measures (i.e. EEG-based) related to the experienced workload have been collected. Results showed higher significant performance achieved by the users during the “Augmented” modality with respect to the “Normal” one, supported by a significant decreasing in experienced workload, evaluated by using EEG-based index. In addition, Performance and EEG-based workload index showed a significant negative correlation. On the contrary, subjective workload analysis did not show any significant trend. This result is a demonstration of the higher effectiveness of neurophysiological measures with respect to subjective ones for Human-Computer Interaction assessment.

P. Aricò, G. Borghini, G. Di Flumeri and A. Di Florio are with Dept. Molecular Medicine, University of Rome “Sapienza”, Italy and IRCCS Fondazione Santa Lucia, Rome, Italy (e-mail: [pietro.arico; gianluca.borghini;gianluca.diflumeri]@uniroma1.it; a.diflorio@hsantalucia.it).

N. Sciaraffa is with Dept. Anatomical, Histological, Forensic & Orthopedic Sciences, Sapienza University of Rome, Italy (e-mail: nicolina.sciaraffa@uniroma1.it).

M. Reynal, C. Hurter and J-P. Imbert and with ENAC, Toulouse, France (e-mail: [maxime.reynal; christophe.hurter; jean-paul.imbert]@enac.fr).

M. Terenzi, A. Ferreira and S. Pozzi are with Deep Blue Research and Consulting, Rome, Italy (e-mails: [michela.terenzi; ana.ferreira; simone.pozzi]@dblue.it).

V. Betti is with Dept. Psychology, Rome, Italy (email: viviana.betti@uniroma1.it).

E. Pavone and M. Marucci are with BrainTrends srl, Rome, Italy (emails: [efpavone; m.marucci91]@gmail.com).

A. Telea is with Dep. of Mathematics and Computing Science, University of Groningen, Netherlands (e-mail: a.c.telea@rug.nl)

F. Babiloni, is with the Dept. Molecular Medicine, University of Rome “Sapienza”, Italy (e-mail: fabio.babiloni@uniroma1.it), BrainSigns srl and Hangzhou Dianzi University.

## I. INTRODUCTION

In many operational environments (e.g. aircraft piloting, air-traffic control, industrial process control, robot-assisted surgery) operators have to face with complex systems and machines to accomplish operational activity. Improvements in such technology or even new solutions are often proposed, with the aim to enhance the human-machine interaction (HMI) and consequently increase operator’s performance and consequently overall safety. In this context, the most studied user’s mental state is the *Mental Workload*, due to its strong relationship with the user’s performance variations [1]. The mental workload can be assessed by using different approaches: i) *Performance* assessment e.g. by using a secondary task, provides an objective but indirect measure of the workload; ii) *subjective questionnaires* (e.g. NASA-TLX) provide a direct but subjective measure of the perceived workload; iii) *neurophysiological measure* provide both a direct and objective measure of the experienced workload [2]. With respect to the former two techniques, the latter has the advantages to not impact on the main task, since it does not require any input by the user’s side, and to be available even online, i.e. during the execution of the task. A lot of works demonstrated as electroencephalography (EEG) – based workload measures outperform the other kind of techniques (e.g. ECG, fNIRs) [3]–[5]. In particular, most of the studies showed that the brain electrical activities mainly involved in the mental workload analysis are the theta and alpha brain rhythms respectively gathered from the *Pre-Frontal Cortex* (PFC) and the *Posterior Parietal Cortex* (PPC) regions. Previous studies demonstrated as the EEG theta rhythm over the PFC presents a positive correlation with the mental workload [6], [5]. Moreover, published literature stressed the inverse correlation between the EEG power in the alpha frequency band over the PPC and the mental workload [7]. Depending on such evidences, theta EEG rhythms over frontal sites and alpha EEG rhythms over parietal sites have been used to define an EEG-based workload index, by using the ratio between frontal theta and parietal alpha rhythms. Nowadays, there are not many studies performed in real settings demonstrating the practical advantages of neurophysiological measures for HMI assessment. For example, Borghini et al., [2] performed a preliminary study on few helicopter pilots in which it has been investigated a neurophysiological workload measure to compare avionic technologies. The present work targets tower ATCOs working in a small or medium size towers. In this context, sounds emitted by airplanes play a very important role to accurately locate aircrafts even without

looking at them. Sound is also crucial with low visibility conditions where it provides important input information. In this context, ATCOs should be able to easily identify within the airfield any kind of stimuli (i.e. aircrafts engine sounds), but nowadays one of the encountered difficulties is to locate such stimuli precisely. In this regard, it has been developed at ENAC an interaction modality (the “*Augmented*” solution) by which operators could more easily retrieve information from the airfield. This solution has been compared with a “*Normal*” condition (i.e. without any augmentation) by using performances, subjective and neurophysiological measures.

## II. MATERIAL AND METHODS

### A. Experimental Subjects

Ten French professional ATCOs (6 males, Mean Age:  $\sim 41 \pm 4.6$  years) coming from different airports and formations took part to the experiment. All the ATCOs were normal hearing. Their mean experience in hours was 7330 hours (SD = 5349.98). Their mean number of years in Control Tower was 12.8 (SD = 7.24). Informed consent was obtained from each participant after explanation of the study, which conformed to the revised Declaration of Helsinki and was approved by the local institutional ethics committee.

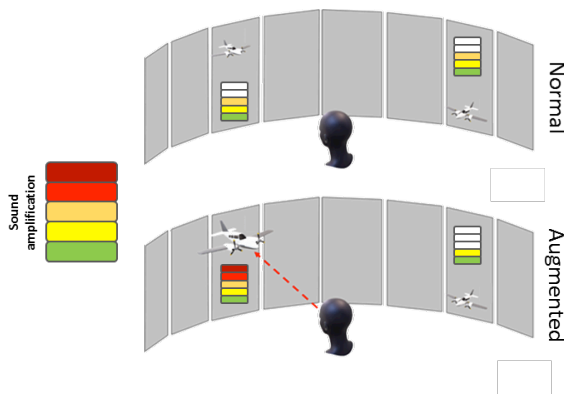


Figure 1. Schematic representation of the two compared interaction modalities (i.e. Normal vs Augmented). In particular, in the *Normal* one all the sound sources (aircraft) have the same amplification factor. In the *Augmented* solution, the amplification factor of the sound sources located along the head axis is increased, while the gain of sound sources located outside this axis is decreased.

### B. Augmented solution

The augmented solution has been implemented by retrieving the participant head orientation via a Microsoft HoloLens mixed-reality headset (we use here only its inertial measurement unit feature). Hence, we can detect in which location the participant’s head is pointing at. This head axis (sagittal) will select sound sources (engine sounds) aligned with its position. All these sound sources are related to aircraft which are spatially placed in the airport vicinity, i.e. in the tower competency. In this way, the sounds that are heard by the participants are spatialized. Figure 1 shows the two experimental conditions. In particular, in the *Normal* modality all the sound sources have the same amplification factor. In the *Augmented* solution, the amplification factor of the sound sources located along the head axis is immediately increased, while the gain of sound sources located outside this axis is decreased.

### C. Experimental Protocol

The experiment was performed at École Nationale de l’Aviation Civile (ENAC, Toulouse, France). With respect to the aim of comparing the two different interaction modalities (i.e. Normal and Augmented); the experimental hypothesis was that the *Augmented* solution should be able to enhance ATCO’s performance and/or decrease the experienced workload with respect to the *Normal* solution. To investigate this hypothesis, it has been reproduced a synthetic but realistic Control Tower environment, in particular the *Muret* airport (France) by using 8 screens and Flight Gear (FG) open flight simulator (Figure 2).

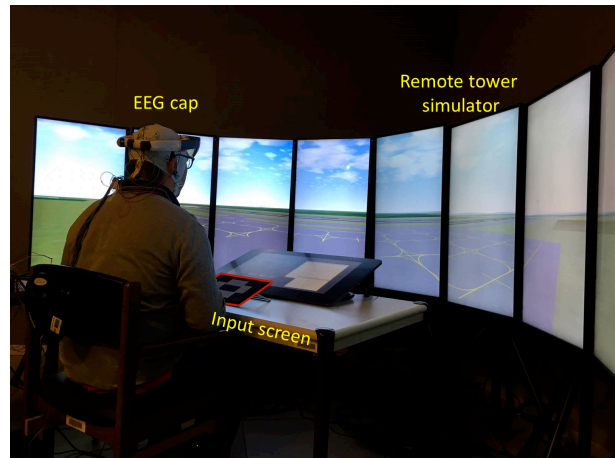


Figure 2. Overview of the experimental setup

The airport airfield has been divided into five distinct “Areas of Interest” (Figure 3). The ATCO has the point of view as the real Control Tower (located in front of the runway) and he had to discriminate the location of aircrafts in that airfield (each area could not contain more than one aircraft at a time) in the two different modalities (i.e. Normal and Augmented). In this regard, the ATCOs had an additional screen (named Input screen, Figure 2) in front of them by which to communicate the location of aircrafts within the five areas. In order to simulate different working scenarios, different trials have been realized as the combination of whether conditions (i.e. sunny or foggy) and difficulty levels (i.e. one or more aircraft at the same time), randomly changed along the experiment. Participants have been trained to use the simulator in both the experimental conditions before the experiment started. Each participant has been asked to perform the experimental task four times for each interaction modality (i.e. Normal and Augmented), in two blocks of two repetitions each (with a resting pause in the middle). The presentation of each modality has been randomized for each subject.

### D. Performed Analysis

This section describes the metrics (e.g. behavioural, subjective and neurophysiological indexes) that have been used to compare the effectiveness of the two modalities (Normal vs Augmented). In particular, to quantify *performance* achieved by the subjects, percentages of correct responses across experimental conditions have been computed and averaged over all the experimental trials. In order to evaluate the mental *workload perceived* by the

ATCOs during the different phases of the experimental protocol, users have been asked to fill the six (Mental demand, Physical demand, Temporal demand, Performance, Effort, Frustration) 100-points range subscales NASA-Task Load index (TLX, [8]). The global workload score from 0 to 100 was obtained for each modality (i.e. Normal vs Augmented) by averaging the individual dimension scale scores. Finally, a *neurophysiological* mental workload index has been computed from the EEG activity for each subject, as the ratio between frontal theta and parietal alpha frequency bands contributions ( $WL_{EEG}$ ). In particular, for each subject, scalp EEG signals have been recorded by the digital monitoring beMicro amplifier (EBNeuro system) with a sampling frequency of 256 (Hz) by 13 Ag/AgCl passive wet electrodes covering the frontal and parietal sites (Fpz, AFz, AF3, AF4, Fz, F3, F4, Pz, P3, POz, PO3, PO4) referenced to both the earlobes and grounded to the left mastoid, according to the 10-20 standard [9].

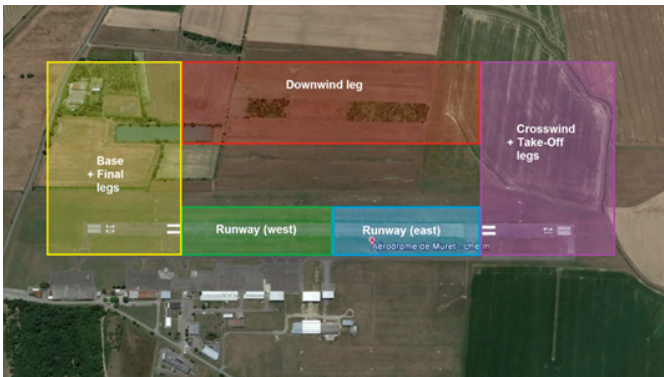


Figure 3. A satellite view of Muret airport, with the five “Areas of Interest” used for this experiment.

In order to compute the  $WL_{EEG}$  index the following steps have been performed. First, EEG signals have been band-pass filtered with a fifth-order Butterworth filter [1-30Hz] and segmented in 2-seconds long epochs, shifted of 125ms [1]. Artifacts contributions that could affect the morphology of theta and alpha bands (e.g. eyes blinks and saccades, muscular artifacts, amplifiers saturations) have been removed by following specific procedures available in the EEGLAB toolbox [10]. All the EEG epochs marked as artifact have been rejected from the EEG dataset with the aim to have an artifact-free EEG signal from which to estimate the brain variations along the different modalities (i.e. Normal vs Augmented). At this point, the *Power Spectral Density* (PSD) was calculated for each EEG epoch using a Hanning window [1] of the same length of the considered epoch (2 seconds). Then, the EEG frequency bands of interest have been defined for each participant by the estimation of the *Individual Alpha Frequency* (IAF) value [11]. In order to have a precise estimation of the alpha peak and, hence of the IAF, each ATCO has been asked to keep the eyes closed for a minute before starting with the experiment. Finally, the theta rhythm (IAF-6 ÷ IAF-2), over the EEG frontal channels (Fpz, AFz, AF3, AF4, Fz, F3 and F4), and the alpha rhythm (IAF-2 ÷ IAF+2), over the EEG parietal channels (Pz, P3, P4, POz, PO3 and PO4) have been divided to compute the  $WL_{EEG}$  index. In conclusion, the z-score method [12] has been employed to compute a normalization of  $WL_{EEG}$  index distribution.

### III. RESULTS

#### A. Behavioral data

A paired student’s t-test ( $\alpha=0.05$ ) has been performed between the two modalities in terms of achieved performance. Statistical analysis showed a significantly higher performance ( $p=0.001$ ) achieved during the Augmented modality with respect to the Normal condition (Figure 4).

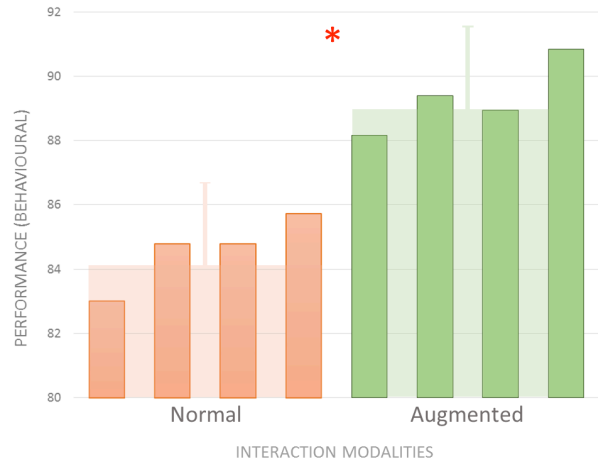


Figure 4. Performance and standard deviation values achieved by the users across the Normal and Augmented modalities for each repetition. The t-test highlighted a significant increment in performance during the Augmented modality.

#### B. Subjective data

A paired student’s t-test ( $\alpha=0.05$ ) has been performed between the NASA-TLX scores related to the two experimental conditions. NASA-TLX analysis showed a decreasing of perceived workload of ~5% during the Augmented modality, with respect to the Normal one. Anyhow, this difference was not significant ( $p=0.214$ ).

#### C. Neurophysiological data

A paired student’s t-test ( $\alpha=0.05$ ) has been performed to investigate differences between modalities. The statistical analysis performed among the experimental conditions (Figure 5), revealed a significant decreasing of the experienced workload during the Augmented solution with respect to the Normal one ( $p=0.003$ ).

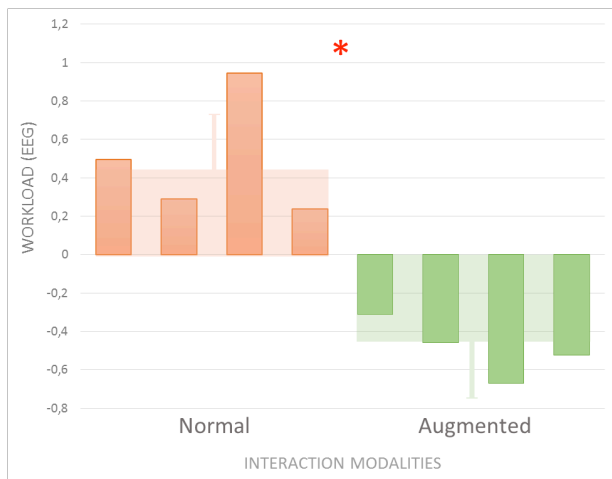


Figure 5.  $WL_{EEG}$  scores and standard deviation related to the workload experienced by the ATCOs, for each modality and repetition. The statistical analysis highlighted a significant decrement in  $WL_{EEG}$  scores during the Augmented modality.

#### D. Correlation analysis

A Pearson correlation analysis has been performed among performances and  $WL_{EEG}$  scores. Results showed a significant negative correlation between the two variables ( $R=-0.9$ ;  $p=0.0024$ ).

#### IV. DISCUSSIONS

Results highlighted a clear advantage of the Augmented interaction modality with respect to the Normal one. More precisely, both performance and neurophysiological workload index analysis have confirmed such behavior. Furthermore, such indexes showed a significant negative correlation, confirming that a higher performance is followed by a lower experienced workload and vice-versa. On the contrary, subjective analysis did not revealed any significant trend, underlining the poor resolution (in terms of sample size) of such technique in comparison to neurophysiological measures (i.e. same trend, but statistically different). In addition, neurophysiological measures allow to measure operator's mental states not only post-task (as for subjective analysis), but even during the task execution since the measure does not require any input from the user side and does not interfere with the task that he is performing. This feature could allow for a better tuning of the technology the operator is interacting with, in order to optimize the human-machine interaction enhancing the performance of the whole system.

#### V. CONCLUSION

The proposed study aimed at investigating the suitability of neurophysiological indexes to compare different interaction modalities by measuring the operator's experienced workload. Achieved results suggested that neurophysiological indicators could provide lower intra-subjects variability with respect to subjective methods (i.e. NASA-TLX), since the same trend resulted not significant for subjective measure, and significant for neurophysiological one.

This study represents an example of a new perspective in using neurophysiological measures, in which the objective is to assess the quality of human-machine interaction by the comparison of different technological solutions, with the final aim of enhancing user's performance and increasing safety.

#### ACKNOWLEDGMENT

This work is co-financed by the European Commission by Horizon2020 projects Sesar-06-2015 the embodied reMote Tower, "MOTO", GA n. 699379; Sesar-01-2015 Human Performance neurometricS Toolbox for highly automatEd Systems deSign, "STRESS", GA n. 699381; H2020-MG-2016 simulator of behavioural aspects for safer transport, "SimuSafe", GA n. 723386; Horizon2020 project HCO-06-2015, Multidisciplinary tools for improving the efficacy of public prevention measures against smoking, "SmokeFreeBrain", GA n. 681120.

#### REFERENCES

- [1] P. Aricò, G. Borghini, G. Di Flumeri, A. Colosimo, S. Pozzi, and F. Babiloni, "A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks," in *Progress in brain research*, vol. 228, 2016, pp. 295–328.
- [2] G. Borghini *et al.*, "Avionic technology testing by using a cognitive neurometric index: A study with professional helicopter pilots," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, vol. 2015, pp. 6182–6185.
- [3] P. Arico, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, and F. Babiloni, "Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1431–1436, Jul. 2017.
- [4] P. Aricò *et al.*, "Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment," *Front. Hum. Neurosci.*, vol. 10, p. 539, 2016.
- [5] G. Di Flumeri *et al.*, "On the Use of Cognitive Neurometric Indexes in Aeronautic and Air Traffic Management Environments," Springer, Cham, 2015, pp. 45–56.
- [6] M. E. Smith and A. Gevins, "Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator," 2005, vol. 5797, p. 116.
- [7] N. Jaušovec and K. Jaušovec, "Working memory training: Improving intelligence – Changing brain activity," *Brain Cogn.*, vol. 79, no. 2, pp. 96–106, Jul. 2012.
- [8] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Adv. Psychol.*, vol. 52, pp. 139–183, Jan. 1988.
- [9] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, Feb. 2007.
- [10] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [11] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999.
- [12] J.-H. Zhang, Chung, and Oldenburg, "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays," *J. Biomol. Screen.*, vol. 4, no. 2, pp. 67–73, Apr. 1999.