

Stress Assessment for Augmented Reality Applications based on Head Movement Features

Anna Ferrarotti, *Student Member, IEEE*, Sara Baldoni, *Member, IEEE*, Marco Carli, *Senior Member, IEEE*, and Federica Battisti, *Senior Member, IEEE*

Abstract—Augmented reality is one of the enabling technologies of the upcoming future. Its usage in working and learning scenarios may lead to a better quality of work and training by helping the operators during the most crucial stages of processes. Therefore, the automatic detection of stress during augmented reality experiences can be a valuable support to prevent consequences on people's health and foster the spreading of this technology. In this work, we present the design of a non-invasive stress assessment approach. The proposed system is based on the analysis of the head movements of people wearing a Head Mounted Display while performing stress-inducing tasks. First, we designed a subjective experiment consisting of two stress-related tests for data acquisition. Then, a statistical analysis of head movements has been performed to determine which features are representative of the presence of stress. Finally, a stress classifier based on a combination of Support Vector Machines has been designed and trained. The proposed approach achieved promising performances thus paving the way for further studies in this research direction.

Index Terms—Augmented Reality, Stress detection, Machine Learning Classifier

I. INTRODUCTION

In recent years, the success of immersive technologies has steadily increased [1] and nowadays Augmented Reality (AR) and Virtual Reality (VR) are used in a wide range of applications. While VR is often employed for entertainment [2] (e.g., gaming or streaming applications), AR systems are expected to become a key element in future work and learning scenarios (e.g., professional training, medical studies) [3], [4]. Under these circumstances, the onset of stress while using an AR application may impair the completion of the required tasks. Therefore, the detection of stress conditions while using Head Mounted Displays (HMDs) is a relevant aspect to be investigated to fully exploit the potentiality of these new technologies. In fact, physiological and psychological consequences of stress can interact with each other causing dangerous changes in the components of the physiological systems of the human body and, at the same time, causing changes in a person's behavior, thus leading to an unhealthy status [5]. In addition, being a technology still under development, the HMD itself can contribute to the occurrence of a stress condition (e.g.,

heaviness, blurred vision, uncomfortable fit), thus precluding the adoption of HMDs.

The onset of a stress condition can be identified and measured as the human body reacts to stress with a defense mechanism, represented by physiological variations [6]. These reactions cause changes in different biosignals, such as the Electroencephalogram (EEG) or the Electrocardiogram (ECG) signals [7]. Although in the literature several studies have been proposed for the automatic detection of stress or workload levels [8], the design of non-invasive systems is still an open problem. Therefore, in this work, we present a stress assessment system in an AR scenario, focusing on an AR static task (e.g., reading), which exploits the information on head movement data recorded by the HMD. In this way, the sensors embedded into the HMD can be directly employed for stress analysis thus offering considerable benefits with respect to physiological measurements. Differently from the approaches based on physiological parameters, the proposed solution allows to employ the same device used during working/learning activities for detecting the onset of stress without the need for other hardware than the AR HMD.

The main contributions of this research are the following:

- The design and implementation of a subjective experiment involving two stress-inducing tests in an AR scenario. The acquired data are available to the research community and can be found at <https://muse.uniroma3.it/headdataset>.
- The demonstration of the relation between head movements and the presence of stress through the recording and analysis of head movements acquired during the subjective tests.
- The design and test of a stress classifier based on the acquired data. The usage of two stress-inducing tests allowed to verify the generalization capabilities of the proposed classifier.

The paper is structured as follows. In Section II, an overview of the related works is presented. In Section III, the proposed method for stress assessment is described. More specifically, the stress-inducing tests, the experimental protocol, and the head movement features extracted through the HMD are presented. Moreover, the architecture of the proposed stress classifier is detailed. In Section IV, the results obtained from the statistical analysis and the classification performances are presented. Finally, in Section V the limitations of the current study and the future research directions are sketched, and in Section VI the conclusions are drawn.

A. Ferrarotti and M. Carli are with the Department of Industrial, Electronic, and Mechanical Engineering, Roma Tre University, Rome Italy.
E-mail: {name.lastname}@uniroma3.it

S. Baldoni and F. Battisti are with the Department of Information Engineering, University of Padova, Padua, Italy
E-mail: name.lastname@unipd.it

Manuscript received XXXX XX, XXXX; revised YYYYY YY, YYYY.

II. RELATED WORKS

In this section, a brief summary of the most widely adopted stress-inducing tasks is provided and the stress detection approaches presented in the literature are described.

A. Stress Inducing Tasks

Due to the complex mechanisms behind the onset of stress, it is necessary to evaluate the body response to a stressor in controlled environments before testing the designed systems in real-life scenarios. In the literature several methods have been proposed to induce stress in a laboratory environment. Among them, one of the most used is the Stroop Color Word Test (SCWT) [9] since its capability to induce stress is well-established [7], [10]. In [11] heart rate, frequency of skin conductance responses, and self-reported anxiety have been recorded while performing a SCWT for demonstrating its correlation with stress. The performed analysis proved that the SCWT may act as an efficient laboratory stressor. Similarly, in [12], the stress induced by the SCWT was measured using electrocardiographic and heart rate variability signals. The results of this study highlight the presence of a significant variation in these signals between the stressed and the normal conditions. The SCWT presents to the user different color names in two conditions. In the first condition, i.e. the *congruous* condition, the words are all written in black and the user is asked to read the words aloud. In the second condition, i.e. the *incongruous* condition, the color of the words and their meaning are not coherent (e.g., the word yellow may be written in red). In this case, the user is asked to pronounce the color in which the word is represented and not the word's meaning. This creates a conflict between the automatic reading process and the task of naming the color in which the word is written [13]. Variations of the SCWT have also been proposed, either combining the visual stimulus with an auditive stimulus [14] or introducing time limitations [11]. In addition, some authors have added a second congruous condition in which the word colors match their meaning, thus realizing an intermediate step between the original congruous and incongruous phases [10], [15].

Another class of stress-inducing tests, which also requires a high level of cognitive engagement from the users, is represented by Mental Arithmetic (MA) tasks. The task of mental arithmetic involves performing calculations and solving mathematical problems only through mental processes, without the use of external aids or tools. It requires the ability to manipulate numbers, apply mathematical operations, and accurately derive results in one's mind. Different implementations of the MA test have been proposed in the literature [16]–[18]. The Pased Auditory Serial Addition Test (PASAT) presents to the user a series of single-digit numbers. Each time a new number is presented, the user is asked to sum it to the immediately preceding one [19]. This test is also used in the Mannheim Multi Component Stress Test (MMST), which incorporates five distinct stressors presented in combination to the users. The first stressor is represented by the PASAT, during which 44 pictures depicting negative feelings are presented in the background, serving as the second stressor. Moreover, acoustic

(i.e. white and random explosion noise) and motivational stressors are included. Finally, considering the category of stress-inducing tasks related to MA, the Montreal Imaging Stress Task (MIST) requires the user to solve mathematical problems in a limited amount of time, which varies depending on the cognitive capability of the user himself [20].

Exploiting sensory cues to elicit stress has been proposed in other stress-inducing tests. The unpleasant pictures in the International Affective Picture System (IAPS) database can be used to induce a stress response [21]. In [22], [23], the onset of stress is caused by a horn sound played randomly during the test. In the Cold Pressor Test [24], participants are asked to immerse a hand or foot in cold water.

Another category of stress-inducing tests exploits social pressure. In the Trier Social Stress Test (TSST) [25], participants are asked to prepare in a limited amount of time a presentation to be performed in front of an audience that does not give any positive feedback. In addition, participants are asked to perform a MA test without previous notice. Moreover, the Maastricht Acute Stress Test (MAST) combines the TSST, the cold pressor, and the MA tests [26].

Other stress-inducing tests are based on a gamification approach [27], [28]. As an example, during the Wisconsin Card Sorting Test [29], users are asked to match the cards of a deck without knowing the sorting rule, which must be discovered during the test itself. Lastly, other stress-inducing tasks are related to specific target scenarios, e.g. stress induced during driving tasks [30]–[32] or during university examinations [33].

In this work, the SCWT and the MA have been chosen as stress-inducing tasks. In fact, these tests are widely accepted as highly reliable for the onset of stress in participants as highlighted in [7], [34]. They are also more similar to the target scenario of this study, i.e. a system for detecting stress in workers/students performing static activities.

B. Stress Detection Techniques

In the literature several authors have analyzed the effects of stress on people, focusing on automatic stress detection [35], [36]. One of the most accurate methods exploits the processing of biological signals, whose variation can be related to the onset of stress [7].

Among the biosignals, EEG is widely used for stress assessment [37], [38]. Although it can be analyzed in both time and frequency, different levels of cognitive engagement are often accurately differentiated through the analysis of the EEG power spectral density [39]. The ECG, which represents the heart electrical activity during contraction and relaxation, is also commonly used [40], [41]. Although different features of this signal can be exploited in order to detect the presence of stress, Heart Rate (HR), Heart Rate Variability (HRV), and blood pressure, are the most widely adopted [42]. A possible substitute of HRV is the Pulse Rate Variability (PRV), that has been employed for stress detection in [43]. In addition, the Galvanic Skin Response (GSR), which measures human skin conductivity, can be employed for stress detection [44]. For instance, in [45], a multimodal system that exploits both the ECG and the GSR for stress assessment is presented.

Another multimodal approach involving ECG, GSR and an accelerometer has been presented in [46]. Other biosignals that can be exploited to detect the presence of stress are strictly related to the muscular activity of the human body. In particular, eye gaze, eye blink, and pupil dilation are often tracked or measured for stress assessment applications. In [47], these features have been measured through a specific device for eye activity tracking. Moreover, electrodermal activity has been analyzed for comparison. The results highlighted that the variation of the pupil diameter is more discriminative than electrodermal activity features.

Depending on the target application, some biosignals may be more suitable than others. The key factors for performing this choice are the type of sensor adopted for signal acquisition and the requirements in terms of detection accuracy. It is useful to notice that most of the described signals require devices that are typically not available to the consumer market, such as the electroencephalograph and the devices to measure the GSR. In addition, these devices can be difficult to operate for non-professional users. The EEG signal, for example, requires a complex setup for its acquisition. Moreover, although some consumer-grade devices have been proposed [48], they require the user to wear additional equipment for signal acquisition purposes, thus representing a burden in the performance of specific tasks.

The analysis of signals like voice, whose acquisition requires only a microphone, partially addresses this problem. In [49], for instance, a VR application aimed at detecting the presence of stress from the user's voice has been designed. Voice recording was implemented directly in the VR application. However, the voice might not be actively used during certain working and learning activities, in particular those heavily reliant on written communication or where reading is the predominant task to perform. Also, the use of voice in very noisy environments can lead to unreliable results. Therefore, other cues must be exploited to gauge the emotional state of an employee/student.

In the past years, several works have demonstrated that posture and head movements are related to the presence of stress in non-AR scenarios. In [50], the time variation of head motion, in the presence and absence of stress, has been studied. In addition, the influence of speech on head motion and its variation due to the presence of stress have been considered. Head motion has been analyzed through video recordings of different user activities. From the videos, the head position has been estimated and analyzed through its (x, y, z) components as well as the rotation angles around the three axes. In addition to these features, head velocity has been evaluated. The obtained results highlighted an overall head mobility increase, which is not related to the act of speaking. Thus, although language cognitive processes are strongly linked to head movements, the mobility increase can be attributed to a stress factor. Other studies have shown that body posture and postural sway vary with the presence of stress. In [10] and [15], these variations have been analyzed by placing pressure sensors on a commercial office chair so that a person could not sense their presence while sitting. The proposed experimental setup considered a standard 2D

monitor, on which a SCWT was presented to induce stress in the participants. In [10], the authors showed that when a higher cognitive engagement level was required, participants tended to change their posture, moving closer or further away from the screen. In [15], the posture variation has been further analyzed by considering the variation of postural sway. One of the most interesting results concerns the speed of movement, which decreased when the cognitive engagement was larger, due to an increased difficulty in maintaining the balance.

Finally, in [51], stress assessment for AR or VR applications has been performed. The study was based on the analysis of the ballistic forces generated by the human heart (also referred to as ballistocardiography). More specifically, since the heart activity causes involuntary movements, these can be analyzed in order to extract heart rate information, which in turn leads to stress assessment.

In this work we investigate the applicability of head movement-based stress detection to the AR scenario. More specifically, the users' head movement is recorded to detect the presence of stress during a static task such as reading, using the data extracted from the HMD. The proposed method has the following advantages with respect to the approaches based on biosignals:

- it is easy to use - as no specific setup is required;
- it is non-invasive - as the AR HMD is employed for monitoring the user's behavior while the user is performing a task, thus avoiding skin and/or bulky sensors;
- it does not require additional equipment - the HMD is employed by the user for other AR applications and stress detection is performed without adding burdens.

To the best of our knowledge, this work represents the first attempt to directly link head movements and stress in AR applications. In particular, we analyze whether the relation between head movements and stress, which other works already investigated in different scenarios, still holds when using an AR application. Moreover, we designed a classifier to detect stress automatically from the analysis of the head movements, thus realizing a non-invasive and easy-to-use stress assessment approach.

III. PROPOSED METHOD

This work focuses on analyzing the relationship between stress and head movements while using an AR application. Due to the foreseen spreading of AR solutions for working [52] and learning activities [53], we considered a static task (e.g., reading, informative training, tutorial watching, data analysis [54]) and analyzed several head movement features to determine whether they show a significant difference during a stress situation. The most significant features have been used to define a Machine Learning (ML) classifier able to detect the presence of stress. To this end, as will be detailed in Section III-A, two stress-inducing tests have been used: the SCWT and the Mental Arithmetic (MA) test.

For each test, a Microsoft Hololens 2 application consisting in a virtual screen displaying the test content has been designed. An example is shown in Figure 1. The dimensions of the virtual screen were chosen by taking into account the

size of 40 inches commercial monitors and its distance was set according to the ITU-R BT.500-14 recommendation for HD 16:9 resolutions [55]. During the tests, the position of the AR HMD was recorded in order to analyze the users' head movements.



Fig. 1. Virtual screen for providing the stress-inducing test to the participants.

A. Design of the stress-inducing tests

1) *Stroop Color Word Test*: The first test we considered is the SCWT [9]. For its design we selected the SCWT version used in [10], [15]. The test consists of presenting a set of slides containing different color names to the user and is composed of three phases, with an increasing level of difficulty. During the first phase, the words are all colored in black, while in the second phase, each word color matches its meaning (*congruous* conditions). In these two phases, the user is asked to read the words as quickly as possible. During the third phase, the color of the words and their meaning are not coherent (*incongruous* condition) and the user must pronounce the color in which the word is written. The third phase induces the stress condition. During the test, the number of words presented in each slide and the duration of each slide varied. These two parameters have been set in accordance with [10], [15], to allow a fair comparison of the achieved results. In the third phase, the difficulty of the task is increased both through the mismatch between the word color and its meaning and by reducing the amount of time provided to complete the task. The overall test has a duration of approximately 2 minutes. The details of the test are represented in Table I, which shows the number of words per slide and the amount of time scheduled to complete the task. In Figure 2, an example of a slide is provided. In Figures 2(a)-2(c), a sample slide for each phase is presented, while Figure 2(d) shows the maximum number of words presented to the users during the third phase.

During the test, only head movements have been recorded, disregarding the correctness of the answers provided by the users. In fact, this work aims to analyze the relation between the presence of stress and head movement features, thus making the correctness of the answers not relevant to our purposes.

2) *Mental Arithmetic test*: The second stress-inducing test is based on MA. The designed test comprises two phases. During the first phase, which is 1 minute long, the participants are asked to close their eyes and relax. During the second one, which lasts for approximately 3 minutes, the participants are asked to solve 19 mathematical operations. The complexity

TABLE I
SCWT STRUCTURE.

Phase	No. of words	No. of slides	Provided time [s]
Phase 1	3	2	2
	4	2	2.5
	5	1	3
	6	1	3.5
Phase 2	3	2	2
	4	2	2.5
	5	1	3
	6	1	3.5
Phase 3	3	2	2.5
	4	2	3
	5	2	3.5
	6	2	4
	9	3	4
	12	1	6

of each calculation is mixed so that a person would not be discouraged by the fact that he or she could not solve one of the previous problems [56]. Each operation is presented to the participant for a different time interval depending on the difficulty of the operation: 5s for easy tasks (e.g., compute 5×2), 10s for medium tasks (e.g., compute $15 + 28$), and 15s for complex tasks (e.g., compute 32×17). To increase the level of stress perceived by the user, the time left for solving each operation was displayed on the screen through a filling bar and audio feedback was provided to inform the participant whether the given answer was correct or not.

B. Experimental protocol

The designed experimental protocol consists of three sessions involving three independent groups of participants as detailed in Section IV-A:

- Session #1: participants performed the SCWT; the acquired data were used for analyzing the head movements and selecting the features. The selected features have been used as the training set for the proposed stress classifier.
- Session #2: participants performed the SCWT; the acquired data have been used as validation set for the proposed stress classifier.
- Session #3: the participants performed both the SCWT and the MA test; the acquired data have been used as the test set for the proposed stress classifier and to verify its generalization capabilities.

The stress-inducing tests were performed in a controlled environment. The virtual screen was superimposed on a white background in a quiet room so that external stimuli would not interfere with the execution of the different tasks.

For Session #1 and Session #2 the experimental protocol consisted of three steps: i) participant screening procedure, ii) AR test, and iii) final questionnaire. The screening comprises a Snellen Test [57] and a test for color blindness [58]. The participants were asked to wear any vision-correcting devices (glasses or contacts) that they normally wear. Before the test, the eye gaze calibration procedure for the headset was performed. During Sessions #1 and #2, after performing the

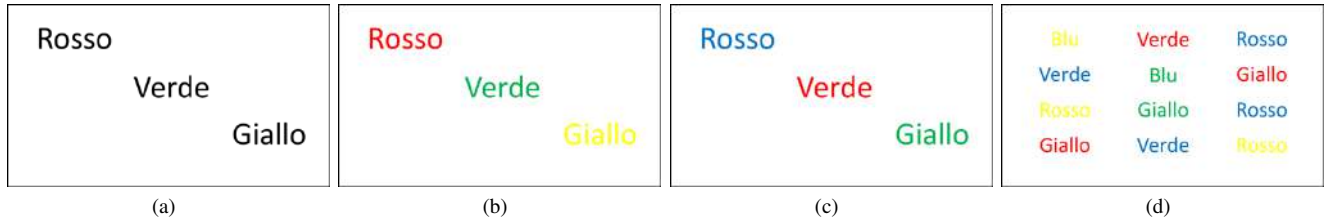


Fig. 2. Example of words presented to the users during the SCWT: (a) first phase; (b) second phase; (c) third phase; (d) third phase, with the maximum number of displayed words.

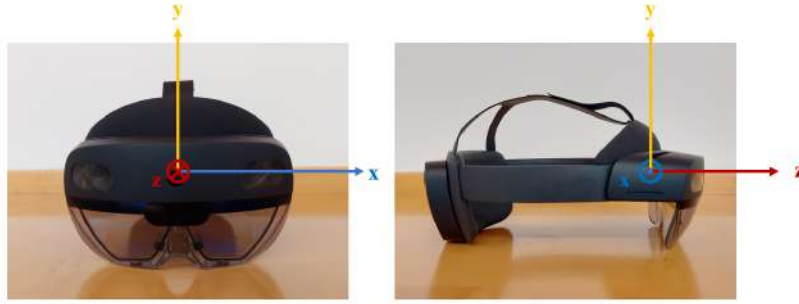


Fig. 3. Reference system used by Microsoft Hololens 2.

SCWT, the participants were asked to answer the Simulator Sickness Questionnaire (SSQ) [59] to evaluate the presence of cybersickness due to the use of the HMD. The SSQ allows the participant to rate 16 cybersickness-related symptoms on a four-level scale: none, slight, moderate, and severe. Session #3 had a different structure. More specifically it was organized as follows:

- participant screening procedure;
- SCWT;
- NASA-TLX questionnaire;
- 5 minutes break;
- MA test;
- NASA-TLX questionnaire;
- SSQ.

The screening procedure and the SCWT were performed in the same way as during the previous sessions. The NASA-TLX (National Aeronautics and Space Administration Task Load Index) questionnaire [60], was introduced in the 80's to measure the perceived workload and mental effort. It consists of a series of questions that aim to capture six dimensions of task load (i.e., mental demand, physical demand, temporal demand, performance, effort, and frustration). Since the SCWT is a well-known and standardized stress induction procedure, while the MA tests have not reached the same level of maturity, the NASA-TLX questionnaire has been employed to verify the similarity of stress levels induced by the two tests.

C. Feature extraction and statistical analysis

The HMD used for the experiments, Microsoft Hololens 2, allows to track the head movements and to extract the variation in time of the x, y, z coordinates, according to the coordinate space presented in Figure 3.

From the acquired data, different head movement features were computed and used as observations for the statistical analysis. First of all, the movements along the three axes were extracted to analyze the users' movement trends. Then, the absolute value of the difference between adjacent temporal samples was evaluated. This feature, referred to as *entity of displacement*, allowed us to verify if there was a privileged axis for the head movement. In addition, the sign of the difference between adjacent temporal samples, referred to as *sign of displacement*, was computed. This feature has been considered to verify if there was, along a specific axis, a privileged direction of motion. Moreover, the total displacement has been evaluated to analyze the overall head movement. This feature has been computed as the norm of the vector with components x, y , and z . As further analysis, the head velocity along the three axes and the total velocity, evaluated as the norm of the velocity vector, were computed.

Since further insights could be gained by combining the information coming from different axes, we included the time variation of the angles between two axes ($(x, y), (x, z), (y, z)$) and their respective angular velocities. Also in this case, the entity and sign of the angular displacement were evaluated.

Finally, since an oscillatory movement trend was noticed, we performed a Short-Time Fourier Transform (STFT) analysis to investigate it. Both the module and phase of the STFT coefficients were analyzed in order to evaluate their frequency content. To deal with the inherent discontinuity of phase coefficients, a phase unwrapping operation was performed. Moreover, in order to obtain a frequency representation of the signals independent from the length of the time sequence, a cumulative sum of the STFT coefficients belonging to the same frequency bin was performed.

In order to determine whether the described features were

significantly different among the three SCWT phases, two statistical tests were considered. An Analysis of Variance (ANOVA) test [61] was performed using the three phases of the test as levels of the independent variable. Before performing the ANOVA test, the normality of the population of interest and the homoscedasticity between the samples related to different levels of the independent variable have been verified. Normality was checked for each feature through a Kolmogorov-Smirnov test [62]. Homoscedasticity was assumed to be respected if the highest variance was not larger than twice the smallest one. The null hypothesis was defined as the absence of a significant change in the analyzed feature in the three phases. If the results of the ANOVA test on a feature allowed to reject the null hypothesis, a Tukey Honestly Significant Difference (HSD) [61] test was also performed, to evaluate between which phases such significant difference held.

When the assumptions for the ANOVA were not respected by the feature under analysis, a Welch's ANOVA test [63] was performed. This test is the corresponding non-parametric test of the ANOVA for the within-groups design. Similarly to the previous case, if the results of Welch's ANOVA test on a feature allowed to reject the null hypothesis, a Games-Howell [64] post-hoc test was performed.

D. Proposed stress classifier

The features that showed a significant difference between the third phase and at least one of the other two phases have been selected as possible input features for the proposed stress classifier. In fact, these features were assumed to be more effective in discriminating between the presence and absence of stress. Since all time variations along the three axes showed a significant difference among the three phases, as it will be shown in Section IV-C, the total displacement was used as a comprehensive feature. The selected features are total displacement, total speed, and the three angular displacements. All these features represent a combination of the information collected from different axes.

For all test sessions, the input data were divided into two classes: presence and absence of stress. The labeling of the input data was carried out considering the different phases of the stress-inducing tests. In particular, for the SCWT, the first two phases were labeled as non-stress, while the third one was labeled as stress. For the MA, the first phase was labeled as non-stress, while the second one was labeled as stress. In fact, the third phase of the SCWT and the second phase of the MA test require a higher level of cognitive engagement and are designed to increase the perceived level of stress.

The proposed architecture for the stress classifier is presented in Figure 4. The computed features undergo the STFT analysis described in the previous section. This choice is motivated by a variation of the frequency content of the signal during the three phases, as will be detailed in Section IV-C. More specifically, the unwrapped phase of the STFT coefficients is chosen as the input feature for the stress classifier. In fact, the STFT phase carries more information than its module and can be more representative of the head movement

signal [65]. The cumulative sum of the phase coefficients allows a representation which is independent from the length of the analyzed sequence. This is fundamental since ML algorithms, such as Support Vector Machines (SVMs), do not take into account the temporal variations of the input data but rely on the variations of a limited number of predictors [66].

To account for the different characteristics of the selected features, a SVM classifier is trained for each of them. In order to obtain a more robust design, we propose to combine the outputs of the classifiers at two different stages. First, the classification decisions of the three angular displacement classifiers are combined through a majority voter. Then, the decisions concerning the total displacement, d_1 , the total speed, d_2 , and the angular displacement, d_3 , are combined through a weighted sum. This approach allows on one hand to fit each classifier to the peculiarities of each feature, and on the other to overcome the flaws of the single SVM by relying on multiple features.

The outputs of the individual classifier, d_i ($i = 1, 2, 3$), are set to -1 if the input is classified as a non-stress observation, and to 1 otherwise. The weights, w_1 , w_2 , and w_3 , have been defined based on the accuracy of the different classifiers on the validation set, composed of the data acquired during Session #2. Indicating the accuracy values as a_1, a_2, a_3 , and setting

$$a_m = \min \{a_1, a_2, a_3\}, \quad (1)$$

the weights are computed as follows:

$$w_i = \frac{a_i/a_m}{\sum_{i=1}^3 a_i/a_m}, \quad i = 1, \dots, 3. \quad (2)$$

The weighted sum is obtained as:

$$d = \sum_{i=1}^3 w_i d_i. \quad (3)$$

Since the sum of the weights is equal to 1, the value of the weighted sum can vary in the interval $[-1, 1]$.

The final decision is taken by comparing the weighted sum with a threshold. In Section IV, this threshold is set to 0, to evaluate the accuracy of the proposed classifier.

IV. RESULTS

A. Test sample

During the three acquisition sessions, a total of 100 users participated in the AR tests. All participants were Italian speakers, and the words presented during the SCWT were in Italian. Before performing the tests, all participants were asked to sign a privacy agreement for data collection.

During Session #1, the SCWT was performed by a group of 60 subjects, 32 men and 28 women, whose ages varied between 19 and 47 years (25.6 ± 4.5). Session #2 comprised a group of 20 participants, 10 men and 10 women, whose age varied between 22 and 47 years (28.8 ± 6.2). Also in this case the SCWT has been submitted to the participants. Finally, during Session #3, the SCWT and MA test were performed by a group of 20 subjects, 12 men and 8 women, whose age was in the range between 21 and 31 years (24.4 ± 3.3).

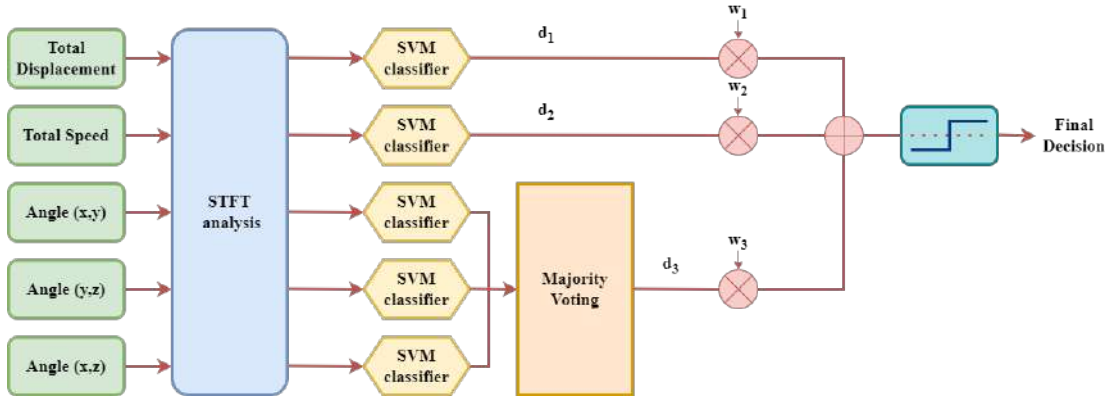


Fig. 4. Proposed stress classifier architecture.

TABLE II
AVERAGE NUMBER OF SAMPLES PER USER. THE DASH SYMBOL INDICATES NON APPLICABLE FIELDS.

Test phase	Training set		Validation set		Test set			
	SCWT		SCWT		SCWT		MA test	
	Before pre-processing	After pre-processing	Before pre-processing	After pre-processing	Before pre-processing	After pre-processing	Before pre-processing	After pre-processing
Phase 1	1376.8 ± 4.8	275.8 ± 1.1	1374.6 ± 5.6	275.3 ± 1.2	1373.3 ± 4.7	275.6 ± 1.0	4489.1 ± 7.4	898.35 ± 1.5
Phase 2	2218.2 ± 8.3	444.0 ± 1.7	2215.9 ± 2.7	443.6 ± 4.3	2217.4 ± 13.6	443.9 ± 2.7	9684.9 ± 19.2	1937.4 ± 3.9
Phase 3	2903.5 ± 13.9	581.1 ± 2.8	2906.8 ± 21.4	581.7 ± 4.3	2905.4 ± 9.3	581.5 ± 1.8	-	-

B. Pre-processing

The collected data were first divided based on the duration of each phase of the stress-inducing tests as reported in Section III-A. In order to filter out possible noise, we considered a windowed mean of the samples. To this aim, we employed a window with a fixed length instead of setting the window equal to a predefined time interval. In fact, since the sampling interval of Microsoft Hololens 2 is not constant, the same time interval may contain a slightly variable number of samples. A fixed window length allows to filter noise while preserving the original signal variance. This aspect is important for the performed statistical analysis, since a variation in the signal variance may prevent from meeting the assumptions required from the ANOVA test. The window length has been set to 5 samples based on the average number of samples included in a temporal window of 0.1 seconds. In fact, considering the limits for head speed during reading tasks [67], the head displacement during this time interval can be deemed not significant. In Table II, the average number of samples per user along with the corresponding standard deviation is reported for both SCWT and MA, before and after pre-processing.

C. Statistical analysis

In Table III, the ANOVA and HSD tests results for the different head movement features in the time domain are presented, while in Table IV the Welch's ANOVA and Games-Howell results for the frequency-domain representation of the head movement features are illustrated. The absence of changes in the features in presence and absence of stress was used as the null hypothesis. In particular, three outcomes are possible:

- the null hypothesis is not rejected (-);

TABLE III
RESULTS OF THE ANOVA TESTS FOR THE TIME-DOMAIN REPRESENTATION OF THE HEAD MOVEMENT FEATURES

Head movement feature	Test phase					
	1 - 2		2 - 3		1 - 3	
	Trend	p-value	Trend	p-value	Trend	p-value
X coordinate	↑	< 0.001	↓	< 0.001	↓	0.003
Y coordinate	↓	< 0.001	↓	< 0.001	↓	< 0.001
Z coordinate	↑	< 0.001	↑	< 0.001	↑	< 0.001
Total displacement	↓	0.004	↑	< 0.001	↑	< 0.001
X Entity of displacement	↓	< 0.001	↑	< 0.001	↑	< 0.001
Y Entity of displacement	-	0.124	↑	< 0.001	↑	< 0.001
Z Entity of displacement	-	0.102	↑	< 0.001	↑	< 0.001
X sign	↓	0.034	↑	< 0.001	-	0.754
Y sign	-	0.630	↓	< 0.001	↓	< 0.001
Z sign	↓	0.038	-	0.966	↓	0.049
Speed along x axis	-	0.865	-	0.929	-	0.973
Speed along y axis	-	0.954	-	0.814	-	0.686
Speed along z axis	-	0.984	-	0.995	-	0.995
Total speed	↓	0.009	↑	< 0.001	↑	< 0.001
Angle (x, y)	↑	< 0.001	↓	< 0.001	-	0.999
Angle (y, z)	↑	< 0.001	↑	< 0.001	↑	< 0.001
Angle (x, z)	-	0.610	↓	< 0.001	↓	< 0.001
Entity of displacement angle (x, y)	-	0.682	↑	< 0.001	↑	< 0.001
Entity of displacement angle (y, z)	-	0.205	↑	< 0.001	↑	0.047
Entity of displacement angle (x, z)	↓	0.020	↑	0.003	-	0.999
Sign of angle (x, y)	-	0.331	-	0.084	-	0.939
Sign of angle (y, z)	-	0.204	-	0.663	-	0.550
Sign of angle (x, z)	-	0.656	-	0.056	-	0.545
Angular speed (x, y)	-	0.987	-	0.990	-	0.959
Angular speed (y, z)	-	0.999	-	0.995	-	0.999
Angular speed (x, z)	-	0.991	-	0.999	-	0.983

- the null hypothesis is rejected and the analyzed feature increases during the stress-related phase (↑);
- the null hypothesis is rejected and the analyzed feature decreases during the stress-related phase (↓).

It is useful to highlight that in the SCWT Phase 1 and Phase 2 represent the *congruous* non-stress conditions, while Phase 3 represents the *incongruous* and stress-related condition. For

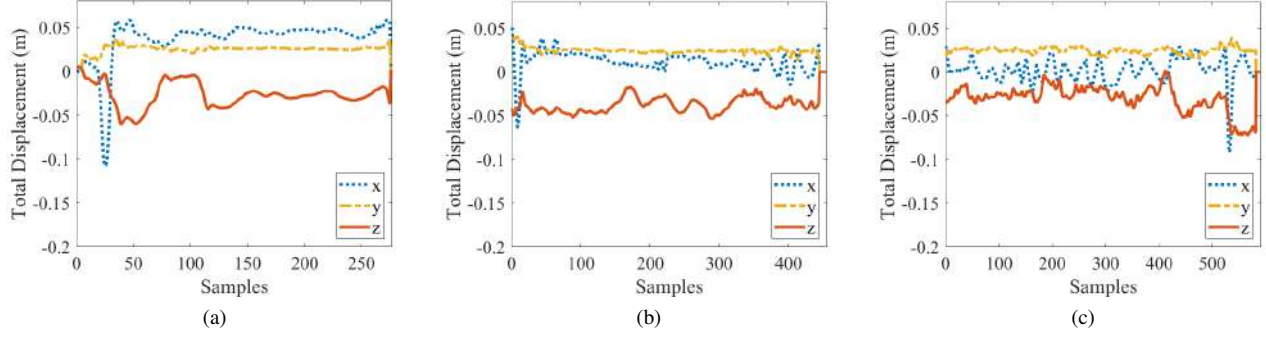


Fig. 5. Example of head movement along the three axes of the (a) first, (b) second, and (c) third phases of the SCWT for one participant.

TABLE IV
RESULTS OF THE WELCH'S ANOVA TESTS FOR THE FREQUENCY-DOMAIN REPRESENTATION OF THE HEAD MOVEMENT FEATURES

Head movement feature	Module						Phase					
	Test phase 1 - 2		Test phase 2 - 3		Test phase 1 - 3		Test phase 1 - 2		Test phase 2 - 3		Test phase 1 - 3	
	Trend	p-value	Trend	p-value	Trend	p-value	Trend	p-value	Trend	p-value	Trend	p-value
Total displacement	↓	< 0.001	↑	< 0.001	↑	< 0.001	↑	< 0.001	↓	< 0.001	↓	< 0.001
Total speed	↓	< 0.001	↑	< 0.001	↑	< 0.001	↑	0.008	↓	< 0.001	↓	< 0.001
Angle (x, y)	↓	< 0.001	↑	< 0.001	↑	< 0.001	↑	< 0.001	↓	< 0.001	↓	< 0.001
Angle (x, z)	↓	< 0.001	↑	< 0.001	↑	< 0.001	↑	< 0.001	↓	< 0.001	↓	< 0.001
Angle (y, z)	↓	< 0.001	↑	< 0.001	↑	< 0.001	↑	< 0.001	↓	< 0.001	↓	< 0.001

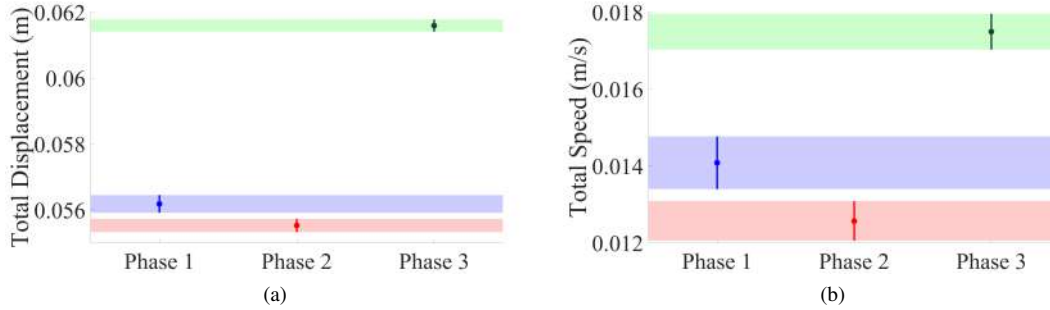


Fig. 6. Results of the HSD test obtained for (a) total displacement and (b) total speed. The filled dots represent the mean value for each phase, while the line and the associated shaded area represent the comparison interval corresponding to the 0.05 significance level. Different colors are used to indicate the phases which are significantly different.

this reason, we expect the stress-discriminant features to be significantly different when comparing Phases 1 and 3 and Phases 2 and 3.

In Figure 5, an example of the variation of the head movement of one participant along the three axes is presented. The temporal variation of the displacement along the axes and the total displacement have shown a significant difference among the three phases, demonstrating that the head movements vary when the complexity of the required task changes. In particular, except for the z -coordinate, the head movement decreased during the third phase. On the other hand, the total displacement has shown an increase during the third phase. Since the total displacement allows to aggregate the information provided by the movements along the single axes, it has been selected as input for the classifiers. The results of the Tukey HSD test are represented in Figure 6(a).

Concerning the entity of the displacement, although a signif-

icant difference between the three phases has been detected, the magnitude of the feature during the different phases is comparable, thus indicating that there is not a privileged axis of movement. A similar consideration can be made for the sign of the head movement. In this case, an oscillatory movement was observed, rather than a movement in a specific direction.

As for the statistical analysis performed on the head speed, the speed along a specific axis was not significantly different during the three phases. On the other hand, it is interesting to note that the total speed, which represents a combination of the other three speed features, is significantly higher during the third phase (Figure 6(b)).

Regarding the angular displacement, this feature is significantly different during the third phase. The only exception is represented by the (x, y) angle which has not shown a significant difference between the first and third phases. However, the decrease of the angular displacement between the second

and third phases indicates a variation of this feature with the task complexity. The results are represented in Figure 7. The outcomes for the entity of the angular displacement, and for the sign of the angular displacement have confirmed that the head movement was not characterized by a privileged direction and it was mainly oscillatory. The angular speed was not significantly different during the three phases.

A second set of statistical tests has been performed for the STFT coefficients of the selected features: total displacement, total speed, and the three angular displacements. These coefficients were evaluated considering a flat top window of 16 samples and 50% overlap. This operation resulted in 16×33 , 16×56 , and 16×73 complex coefficients for the first, second, and third phases of the SCWT, where the first number indicates the frequency bins whereas the second represents the temporal bins. After extracting the phase and module of the coefficients, and performing the phase unwrapping procedure, a cumulative sum along the temporal dimension has been computed thus obtaining 16 coefficients for each user and phase of the test. While the other features respected the assumptions for the ANOVA test, the module and phase of the STFT coefficients were not homoscedastic. Therefore, the Welch's ANOVA test was performed. The corresponding results are provided in Table IV. For both module and phase, a significant difference among the three phases was detected, thus demonstrating the oscillatory nature of the head movement. These results show that the frequency content of the oscillatory movement changes during the different test phases, thus motivating the choice of performing a STFT analysis of the signal before providing it as input to the classifier.

To summarize, the following insights can be gained from the performed analysis:

- the results indicate that the head motion is affected by stress;
- the results concerning the entity and sign of both the singular coordinates and the angular displacements show that there is neither a preferred axis nor a preferred direction of motion;
- the features that aggregate information concerning different axes (i.e., total displacement, total speed and angular displacement) are representative of the presence of stress;
- the results concerning the analysis in the frequency domain suggest that oscillatory head movements are triggered during a stressful event.

D. Classification results

The performances of the proposed classifier for validation and test sets are presented in Table V. More specifically, the results are reported in terms of classification accuracy:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \quad (4)$$

where TP (*True Positives*) indicates number of times a stress-related sample is correctly classified, FP (*False Positives*) represents the number of misclassified stress observations, TN (*True Negatives*) is number of times a non-stress sample is correctly classified, and FN (*False Negatives*) corresponds

to the number of non-stress observations classified as stress-related.

Since the MA test has a different time length compared to the SCWT, the acquired time series have been segmented to obtain sub-sequences whose length is comparable to that of the SCWT. More specifically, the data obtained during the first phase of the test were partitioned into two sub-sequences with an average sample count of 449.2 ± 1.3 that is comparable to the second phase of the SCWT. Similarly, the data collected in the second phase were segmented into three sub-sequences, with an average sample count of 645.8 ± 6.7 that is comparable to the samples in the third phase of the SCWT. For each participant, in order to obtain a single observation for the two classes, the mean of the features evaluated for each sub-sequence was performed. The obtained sequences underwent the same STFT analysis performed for the SCWT.

Table V reports the results obtained using the analyzed features independently and the performance achieved through their combination, as represented in Figure 4. The combination weights have been computed as described in Section III-D based on the classification performance on the data recorded during the validation session (Session #2). As can be noticed in the first column of Table V, the performances of the three classifiers are comparable, thus resulting in similar weights for the three classifiers. More specifically, the weights obtained for the SVM of the total displacement, total speed and angular displacement are $w_1 = 0.34$, $w_2 = 0.33$, and $w_3 = 0.33$, respectively.

Concerning the test session (Session #3), Table V shows that the Combined Classifier has not the highest accuracy for the SCWT. In fact, based on the computed weights, a wrong classification occurs when two out of three classifiers make a wrong decision on the input observation. Despite this, we believe it represents a more robust solution due to the aggregation of the information provided by different features. In fact, the lower accuracy for the SCWT is due to a mismatch in the performance of the different SVMs for the validation and the test sessions. This phenomenon could be overcome through a more extensive validation procedure.

TABLE V
CLASSIFICATION ACCURACIES.

Head movement feature	Accuracy (%)		
	SCWT Val.	SCWT test	MA test
Total displacement	100	98.33	100
Total speed	98.33	98.33	97.50
Angular displacement	98.33	91.67	100
Combined classifier	-	96.66	100

E. SSQ results and NASA-TLX

In Table VI, the SSQ answers obtained during all the acquisition sessions are summarized. In most cases, the participants did not present any discomfort related to the use of the HMD. In fact, for all the symptoms, most users reported no perception. The slightly more frequent symptoms were fatigue, eye strain, and difficulty in concentrating. This may be due

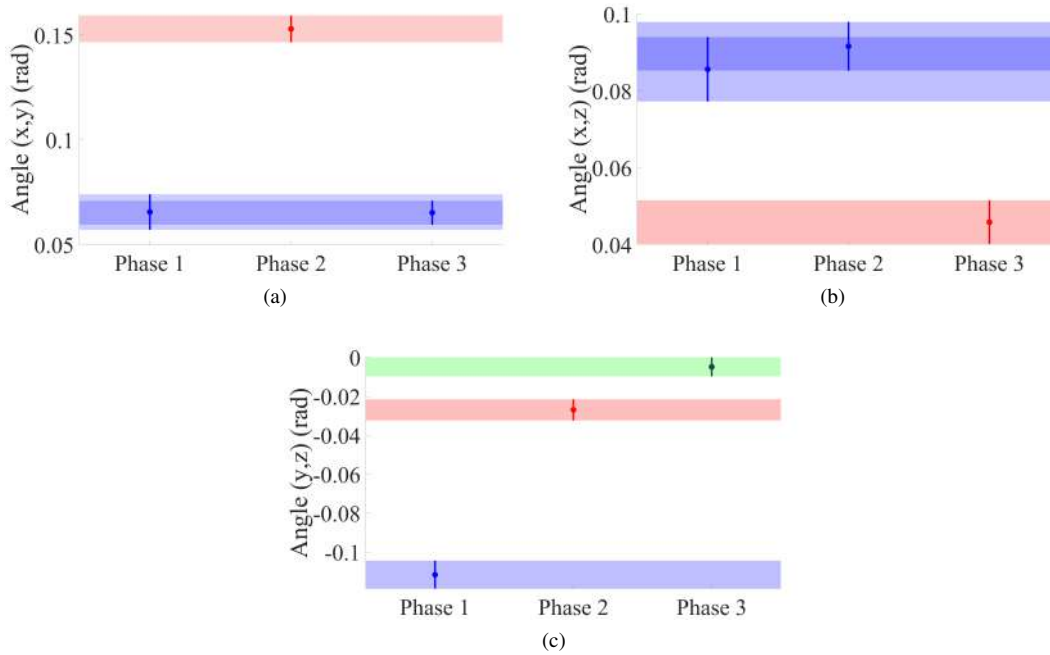


Fig. 7. Results of the HSD test concerning the angular displacement obtained for the (a) (x, y) , (b) (x, z) and (c) (y, z) angles. The filled dots represent the mean value for each phase, while the line and the associated shaded area represent the comparison interval corresponding to the 0.05 significance level. Different colors are used to indicate the phases which are significantly different.

to the cognitive engagement required to perform the different tasks. In general, it can be safely stated that the HMD and the stress-inducing tests did not cause in most cases unpleasant feelings to the users.

TABLE VI
SSQ ANSWERS REPORTING THE % OF PARTICIPANTS SHOWING EACH SYMPTOM.

SSQ symptom	Symptom score (%)			
	None	Slight	Moderate	Severe
General discomfort	85	14	1	0
Fatigue	44	41	15	0
Headache	86	12	2	0
Eye strain	53	40	5	2
Difficulty focusing	61	32	7	0
Increased salivation	88	10	2	0
Sweating	69	22	9	0
Nausea	98	2	0	0
Difficulty concentrating	56	31	10	3
Fullness of the head	62	34	4	0
Blurred vision	74	22	3	1
Dizzy (eyes open)	95	5	0	0
Dizzy (eyes closed)	90	10	0	0
Vertigo	99	1	0	0
Stomach awareness	97	3	0	0
Burping	96	4	0	0

In order to determine whether there was a significant difference between the NASA-TLX scores for the SCWT and MA test, both a two-tailed and a one-tailed paired samples t-tests [61] were performed. In the first case, the null hypothesis assumed the means of the scores obtained for the SCWT and

the MA test to be equal. The test rejected the null hypothesis at a 5% significance level, with a p-value of 0.017. The second test allowed to reject the null hypothesis, according to which the mean of the SCWT scores is larger than the mean of the MA scores, at a 5% significance level. These results show that the MA test was perceived as more stressful than the SCWT, thus achieving the desired goal. In Figure 8, the boxplots of the scores of the two tests are represented. It is possible to notice that the mean of the MA scores is slightly larger than the SCWT one. Although there is not a standardized thresholding system for the NASA-TLX score and the score interpretation is task-dependent [68], it is possible to gain interesting insights from Figure 8. In [68] the authors analyzed the use of NASA-TLX in the literature showing the variation of the scores for different tasks. Among these, the ones which more resemble the scenario considered in this paper are: cognitive tasks (scores between 13.8 and 64.90 with a median of 46) and computer activities (scores between 7.46 and 78 with a median of 54). Moreover, they reported that the smallest and highest scores are 6.21 and 88.5, respectively, and that daily activities range between 7.20 and 37.50 with a median of 18.30. Considering daily activities as baseline, Figure 8 shows that both SCWT and the MA test achieve scores that are remarkably higher. Moreover, for both tasks the median and maximum values are higher with respect to cognitive and computer activities. Therefore, we can conclude that both tasks were demanding in terms of workload and mental effort.

F. Comparison with state-of-the-art approaches

The state-of-the-art review presented in Section II highlighted a wide variety of methodologies and types of test for

TABLE VII
COMPARISON WITH STATE-OF-THE-ART APPROACHES FOR NON-INVASIVE STRESS DETECTION.

Method	[49]	[51] (Cognitive task)	[51] (Movie watching)	Ours (SCWT)	Ours (MA)
Accuracy (%)	82	68.5	74	96.66	100

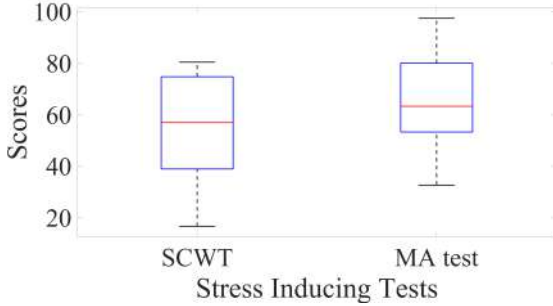


Fig. 8. Boxplot of the NASA-TLX scores obtained during the test acquisition session. The red line in the boxes indicates the median value.

stress detection. For this reason it is not possible to perform a direct comparison with our approach. However, we compare the statistical analysis results with [10] and [15], where the same experimental protocol for the SCWT has been adopted, and the performance of the proposed stress classifier with [49] and [51], where immersive technologies are employed. To further investigate the achievements of the proposed approach, we compare our results with the ones of the methods based on physiological data that selected the SCWT and/or MA test as stress inducers [37], [40], [41], [43], [46], [47].

In [10], [15], a variation in users' movements due to the increasing level of cognitive engagement has emerged. More specifically, an increase in the overall distance covered between the first and third phases was detected. These results are consistent with what has been discussed in Section IV-C, where an increase in the total displacement and in the entity of displacement along all three axes has been identified. In [15], a significant decrease in the mean speed was also highlighted. In this study, an opposite behavior was observed, since the statistical analysis highlighted an increase of the total speed. This difference can be attributed to the chosen experimental protocols. In [15], the participants performed the test sitting on an office chair, while in our case they were in a standing position. Moreover, the difference test modality (i.e., PC and AR screen) could have influenced the participants' behavior.

Moreover, we provide in Table VII the comparison with state-of-the-art approaches in terms of classification accuracy. We reported the best results presented in [49], [51]. As can be noticed, the proposed approach definitely outperforms previous research concerning non-invasive stress assessment in AR/VR.

As for the approaches using physiological measures, in [47] a stress detection method exploiting eye tracking data and electrodermal activity has been presented. The authors performed a binary classification between a "stress" and a "relaxation" phase using the SCWT and a three-level stress classification for the MA test. They achieved an accuracy of 88.43% for

the former scenario, and an accuracy of 91.10% for the latter. Another multimodal stress detection system involving ECG, GSR and an accelerometer has been proposed in [46]. They employed both the SCWT and the MA test in a single experiment. They trained a binary classifier achieving a classification accuracy of 92.4%. The ECG in combination with the SCWT has been employed also in [40]. Stress detection has been modeled as a binary classification problem achieving an accuracy of 96.41%. In [41] ECG has been used to detect stress based on a MA task to differentiate between "stress" and "rest" phases achieving an accuracy of 82.7%. Also in [43] a MA task and the SCWT have been employed to realize a stress classifier using photoplethysmogram signals. The MA test was designed to include up to five different levels of stress, while the SCWT has been employed following a three-level classification approach. The authors achieved 86% accuracy for the SCWT, and designed different detection models for the MA task corresponding to a number of classes from two to five. The presented method achieved an accuracy of 97.77% for two classes, 94.11% for three classes, 94.4% for four classes, and 94.33% for five classes. Finally, in [37] EEG has been employed to detect stress using the SCWT and MA test. The authors realized a binary classifier reaching an accuracy of 88% and 96% for SCWT and MA, respectively. Moreover, they presented a three-level classification approach considering three separate classes for rest, SCWT, and MA, under the assumption that SCWT induces milder stress than MA. In this case, they achieved an accuracy of 75%.

Based on these results, the proposed head motion-based method allows to achieve better or comparable performance with respect to the physiology-based counterparts.

V. LIMITATIONS AND FUTURE DIRECTIONS

In this section we report the limitations of the proposed approach which open new interesting research paths for the design of non-invasive AR stress detection systems.

One limitation of this study lies in the procedure adopted for calculating the weights assigned to the outputs of the individual classifiers. According to the performed analysis, the choice of weights based on the validation set did not allow the generalization to new test samples for the SCWT. The main mitigation for this problem is to increase the number of subjects in the validation phase so that the weights of the different classifiers can be more representative of their capabilities. Another aspect which could be improved is related to the usage of a hard threshold to differentiate between the stress and the non-stress conditions. Future developments could include the design of a more flexible system with a soft thresholding based on three-decision regions. In this case, two thresholds can be defined, one for negative and one for positive values of the weighted sum. In this scenario, the

external negative range represents the non-stress region, while the external positive range represents the stress region. The central region may represent a transition phase, which could be used to alert the user of an upcoming stress condition. The two threshold values should be based on the specific application scenario. In addition, the current system allows to perform only post-processing stress detection, while a relevant improvement could be the realization of a real-time stress detection system.

Finally, in this work, we have considered a static task (e.g., reading, informative training, tutorial watching). Future developments may include the extension of the proposed approach to dynamic and interactive activities. In this direction, improved classification results could be obtained by including additional information such as eye-tracking data. The usage of a multi-modal system could be beneficial for addressing data acquisition artifacts, and for isolating stress-related and task-related movements.

VI. CONCLUSIONS

The analysis of the effects of stress on the human body is a widely investigated research topic. The solutions proposed in the literature mainly rely on invasive systems and infrequently focus on AR/VR applications. To fill this gap, this work investigates whether the head movements of a user wearing an AR HMD vary due to the presence of a stress factor while performing static tasks. In order to induce stress, the SCWT has been used. From the statistical analysis, it has emerged that several head movement features vary when a stressful situation is presented to the user. More specifically, head displacement, head velocity, and angular displacement have shown to vary during all the different phases of the stress-inducing test. Therefore, they have been selected as stress-discriminating features. The phase of the STFT coefficients of these features has been used to define a stress classifier, based on a combination of SVMs optimized for each feature. In order to test the generalization capabilities of the classifier, both the SCWT and a MA test have been employed. Therefore, the main contributions of this work concern the thorough description of how head movements vary due to a stress state in an AR scenario and the definition and realization of a stress classifier, which has shown excellent performances on both tests. Moreover, the proposed system is completely non-invasive and easy to use.

ACKNOWLEDGMENT

This work was partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE0000001 - program “RESTART”).

REFERENCES

- [1] T. Andrade and D. Bastos, “Extended Reality in IoT scenarios: Concepts, Applications and Future Trends,” in *2019 5th Experiment International Conference (exp.at'19)*, 2019.
- [2] S.-W. Hung, C.-W. Chang, and Y.-C. Ma, “A new reality: Exploring continuance intention to use mobile augmented reality for entertainment purposes,” *Technology in Society*, vol. 67, pp. 101757, 2021.

- [3] A. D. Kaplan, J. Cruik, M. Endsley, S. M. Beers, B. D. Sawyer, and P. A. Hancock, “The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis,” *Human factors*, vol. 63, no. 4, pp. 706–726, 2021.
- [4] B. Marques, S. Silva, J. Alves, A. Rocha, P. Dias, and B. S. Santos, “Remote collaboration in maintenance contexts using augmented reality: Insights from a participatory process,” *International Journal on Interactive Design and Manufacturing (IJIDeM)*, pp. 1–20, 2022.
- [5] D. B. O'Connor, J. F. Thayer, and K. Vedhara, “Stress and health: A review of psychobiological processes,” *Annual review of psychology*, vol. 72, pp. 663–688, 2021.
- [6] G. S. Kumar and B. Ankayarkanni, “Comparative Study on Mental Stress Detection Using Various Stressors and Classification Techniques,” in *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 2022.
- [7] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Rioniotis, and M. Tsiknakis, “Review on Psychological Stress Detection Using Biosignals,” *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, 2022.
- [8] J. Heard, C. E. Harriott, and J. A. Adams, “A survey of workload assessment algorithms,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 434–451, 2018.
- [9] J. R. Stroop, “Studies of interference in serial verbal reactions,” *Journal of experimental psychology*, vol. 18, no. 6, pp. 643, 1935.
- [10] D. Bibbo, M. Carli, S. Conforto, and F. Battisti, “A Sitting Posture Monitoring Instrument to Assess Different Levels of Cognitive Engagement,” *Sensors*, vol. 19, no. 3, pp. 455, Jan 2019.
- [11] P. Renaud and J.-P. Blondin, “The stress of Stroop performance: Physiological and emotional responses to color–word interference, task pacing, and pacing speed,” *International Journal of Psychophysiology*, vol. 27, no. 2, pp. 87–97, 1997.
- [12] P. Karthikeyan, P. Murugappan, and S. Yaacob, “Analysis of Stroop Color Word Test-Based Human Stress Detection using Electrocardiography and Heart Rate Variability Signals,” *Arabian Journal for Science and Engineering*, vol. 39, pp. 1835–1847, 03 2013.
- [13] F. Scarpina and S. Tagini, “The stroop color and word test,” *Frontiers in psychology*, vol. 8, pp. 557, 2017.
- [14] M.-H. Lee, G. Yang, H.-K. Lee, and S. Bang, “Development stress monitoring system based on personal digital assistant (PDA),” in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, vol. 1, pp. 2364–2367.
- [15] D. Bibbo, S. Conforto, M. Schmid, and F. Battisti, “The Influence of Different Levels of Cognitive Engagement on the Seated Postural Sway,” *Electronics*, vol. 9, no. 4, pp. 601, Mar 2020.
- [16] M. S. Hussain, R. A. Calvo, and F. Chen, “Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference,” *Interacting with computers*, vol. 26, no. 3, pp. 256–268, 2014.
- [17] A. Mohammadi, M. Fakharzadeh, and B. Baraeejad, “An integrated human stress detection sensor using supervised algorithms,” *IEEE Sensors Journal*, vol. 22, no. 8, pp. 8216–8223, 2022.
- [18] L. D. Sharma, V. K. Bohat, M. Habib, A.-Z. Ala'M, H. Faris, and I. Aljarah, “Evolutionary inspired approach for mental stress detection using EEG signal,” *Expert systems with applications*, vol. 197, pp. 116634, 2022.
- [19] T. N. Tombaugh, “A comprehensive review of the paced auditory serial addition test (PASAT),” *Archives of clinical neuropsychology*, vol. 21, no. 1, pp. 53–76, 2006.
- [20] J. De Calheiros Velozo, T. Vaessen, J. Pruessner, I. Van Diest, S. Claes, and I. Myin-Germeys, “The repeated Montreal Imaging Stress Test (rMIST): Testing habituation, sensitization, and anticipation effects to repeated stress induction,” *Psychoneuroendocrinology*, vol. 128, pp. 105217, 2021.
- [21] P. J. Lang, M. M. Bradley, B. N. Cuthbert, et al., “International affective picture system (IAPS): Technical manual and affective ratings,” *NIMH Center for the Study of Emotion and Attention*, vol. 1, no. 39-58, pp. 3, 1997.
- [22] M. Ehrhart, B. Resch, C. Havas, and D. Niederseer, “A Conditional GAN for Generating Time Series Data for Stress Detection in Wearable Physiological Sensor Data,” *Sensors*, vol. 22, no. 16, pp. 5969, 2022.
- [23] M. K. Moser, B. Resch, and M. Ehrhart, “An individual-oriented algorithm for stress detection in wearable sensor measurements,” *IEEE Sensors Journal*, 2023.
- [24] G. Lamotte, C. J. Boes, P. A. Low, and E. A. Coon, “The expanding role of the cold pressor test: a brief history,” *Clinical Autonomic Research*, vol. 31, pp. 153–155, 2021.

- [25] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, "The trier social stress test: principles and practice," *Neurobiology of stress*, vol. 6, pp. 113–126, 2017.
- [26] A. L. Shilton, R. Laycock, and S. G. Crewther, "The Maastricht Acute Stress Test (MAST): Physiological and subjective responses in anticipation, and post-stress," *Frontiers in psychology*, vol. 8, pp. 567, 2017.
- [27] L. Panagi, L. Poole, R. A. Hackett, and A. Steptoe, "Happiness and inflammatory responses to acute stress in people with type 2 diabetes," *Annals of behavioral medicine*, vol. 53, no. 4, pp. 309–320, 2019.
- [28] L. Pepa, A. Sabatelli, L. Ciabattini, A. Monteriù, F. Lamberti, and L. Morra, "Stress Detection in Computer Users From Keyboard and Mouse Dynamics," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 1, pp. 12–19, 2021.
- [29] S. Miles, C. A. Howlett, C. Berryman, M. Nedeljkovic, G. L. Moseley, and A. Phillipou, "Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility," *Behavior research methods*, vol. 53, no. 5, pp. 2083–2091, 2021.
- [30] P. Pillai, B. Balasingam, Y. Hoon Kim, C. Lee, and F. Biondi, "Eye-gaze metrics for cognitive load detection on a driving simulator," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 4, pp. 2134–2141, 2022.
- [31] D. S. Lee, T. W. Chong, and B. G. Lee, "Stress events detection of driver by wearable glove system," *IEEE Sensors Journal*, vol. 17, no. 1, pp. 194–204, 2016.
- [32] J. Lee, H. Lee, and M. Shin, "Driving stress detection using multimodal convolutional neural networks with nonlinear representation of short-term physiological signals," *Sensors*, vol. 21, no. 7, pp. 2381, 2021.
- [33] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, "Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–13, 2019.
- [34] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021.
- [35] G. Giannakakis, M. R. Koujan, A. Roussos, and K. Marias, "Automatic stress detection evaluating models of facial action units," in *15th IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.
- [36] N. Bu, M. Fukami, and O. Fukuda, "Pattern recognition of mental stress levels from differential RRI time series using LSTM networks," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, 2021.
- [37] G. Jun and K. G. Smitha, "EEG based stress level identification," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 003270–003274.
- [38] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. Al-Nashash, "A Review on Mental Stress Assessment Methods Using EEG Signals," *Sensors*, vol. 21, no. 15, pp. 5043, Jul 2021.
- [39] D. Shon, K. Im, J.-H. Park, D.-S. Lim, B. Jang, and J.-M. Kim, "Emotional Stress State Detection Using Genetic Algorithm-Based Feature Selection on EEG Signals," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, pp. 2461, Nov 2018.
- [40] P. Karthikeyan, M. Murugappan, and S. Yaacob, "ECG signals based mental stress assessment using wavelet transform," in *2011 IEEE International Conference on Control System, Computing and Engineering*, 2011, pp. 258–262.
- [41] J. He, K. Li, X. Liao, P. Zhang, and N. Jiang, "Real-Time Detection of Acute Cognitive Stress Using a Convolutional Neural Network From Electrocardiographic Signal," *IEEE Access*, vol. 7, pp. 42710–42717, 2019.
- [42] A. O. Akmandor and N. K. Jha, "Keep the Stress Away with SoDA: Stress Detection and Alleviation System," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 3, no. 4, pp. 269–282, 2017.
- [43] M. Zubair and C. Yoon, "Multilevel mental stress detection using ultra-short pulse rate variability series," *Biomedical Signal Processing and Control*, vol. 57, pp. 101736, 2020.
- [44] G. Shruti and P. Sanchita, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021.
- [45] S. Sriramprakash, V. D. Prasanna, and O.V. Ramana Murthy, "Stress Detection in Working People," *Procedia Computer Science*, vol. 115, pp. 359–366, 2017, 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22–24 August 2017, Cochin, India.
- [46] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-Aware Mental Stress Detection Using Physiological Sensors," in *Mobile Computing, Applications, and Services*, Martin Gris and Guang Yang, Eds., Berlin, Heidelberg, 2012, pp. 211–230, Springer Berlin Heidelberg.
- [47] M. S. Yousefi, F. Reisi, M. R. Daliri, and V. Shalchyan, "Stress Detection Using Eye Tracking Data: An Evaluation of Full Parameters," *IEEE Access*, vol. 10, pp. 118941–118952, 2022.
- [48] D. He, B. Donmez, C. C. Liu, and K. N. Plataniotis, "High Cognitive Load Assessment in Drivers Through Wireless Electroencephalography and the Validation of a Modified N-Back Task," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 4, pp. 362–371, 2019.
- [49] Arushi, R. Dillon, and A. N. Teoh, "Real-time Stress Detection Model and Voice Analysis: An Integrated VR-based Game for Training Public Speaking Skills," in *2021 IEEE Conference on Games (CoG)*, 2021.
- [50] G. Giannakakis, D. Manousos, P. Simos, and M. Tsiknakis, "Head Movements in Context of Speech during Stress Induction," in *13th IEEE International Conference on Automatic Face & Gesture Recognition*, 2018.
- [51] H. Kusano, Y. Horiguchi, Y. Baba, and H. Kashima, "Stress Prediction from Head Motion," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020.
- [52] F. Kern, F. Niebling, and M. E. Latoschik, "Text Input for Non-Stationary XR Workspaces: Investigating Tap and Word-Gesture Keyboards in Virtual and Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2658–2669, 2023.
- [53] I. Radu and B. Schneider, "How Augmented Reality (AR) Can Help and Hinder Collaborative Learning: A Study of AR in Electromagnetism Education," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 9, pp. 3734–3745, 2023.
- [54] P. Reipschlagier, T. Flemisch, and R. Dachsel, "Personal augmented reality for information visualization on large interactive displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 1182–1192, 2020.
- [55] ITU-R, "BT.500-14: Methodologies for the subjective assessment of the quality of television images," https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-14-201910-1!!PDF-E.pdf, 2019, Last accessed 12 April 2023.
- [56] Y. Jiao, X. Wang, C. Liu, G. Du, L. Zhao, H. Dong, S. Zhao, and Y. Liu, "Feasibility study for detection of mental stress and depression using pulse rate variability metrics via various durations," *Biomedical Signal Processing and Control*, vol. 79, pp. 104145, 2023.
- [57] H. Snellen, *Probekbuchstaben zur Bestimmung der Sehschärfe*, Peters, 1873.
- [58] S. Ishihara et al., "Tests for color blindness," *American Journal of Ophthalmology*, vol. 1, no. 5, pp. 376, 1918.
- [59] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.
- [60] S. G. Hart and L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.
- [61] S. A. Nolan and T. Heinzen, *Statistics for the behavioral sciences*, Macmillan, 2011.
- [62] A. Kolmogorov, "On the empirical determination of a distribution function," *Breakthroughs in Statistics: Methodology and Distribution*, pp. 106–113, 1992.
- [63] B. L. Welch, "On the comparison of several mean values: an alternative approach," *Biometrika*, vol. 38, no. 3/4, pp. 330–336, 1951.
- [64] P. A. Games and J. H. Howell, "Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study," *Journal of Educational Statistics*, vol. 1, no. 2, pp. 113–125, 1976.
- [65] J. M. Blackledge, *Digital signal processing: mathematical and computational methods, software development and applications*, Elsevier, 2006.
- [66] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [67] J. K. Kaakinen, U. Ballenghein, G. Tissier, and T. Baccino, "Fluctuation in cognitive engagement during reading: Evidence from concurrent recordings of postural and eye movements," *Journal of Experimental Psychology: Learning Memory and Cognition*, vol. 44(10), pp. 1671–1677, 2018.
- [68] R. A. Grier, "How high is high? A meta-analysis of NASA-TLX global workload scores," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, 2015, vol. 59, pp. 1727–1731.



Anna Ferrarotti is a Ph.D. student in Applied Electronics at the Department of Industrial, Electronic and Mechanical Engineering of Roma Tre University. She received her Master degree in Information and Communication Technology Engineering from Roma Tre University in 2021 and 2022, respectively. Her main research interests are in the area of quality of experience and immersive technologies.



Marco Carli (SM'05) is Full Professor at the Department of Industrial, Electronic and Mechanical Engineering of Roma Tre University. His research interests are in the area of digital signal and image processing with applications to multimedia communications, digital watermarking, multimedia quality evaluation, and information security. He is Associate Editor for IEEE Transaction on Image Processing and for ELSEVIER Signal Processing: Image Communication.



Sara Baldoni is Assistant Professor at the Department of Information Engineering of the University of Padua. She received her PhD in Applied Electronics from Roma Tre University in 2022. Her main research interests are in the area of quality of experience and immersive technologies. She serves as Associate Editor for ELSEVIER Displays Journal and for ELSEVIER Signal Processing: Image Communication.



Federica Battisti (SM'15) is Associate Professor at the Department of Information Engineering of the University of Padua. Her research interests are in multimedia quality assessment, quality of experience and immersive technologies. She is Editor-in-chief of ELSEVIER Signal Processing: Image Communication, and she serves as Associate Editor of IEEE Transactions on Multimedia, of EURASIP Journal on Image and Video Processing. She is vice-chair of the EURASIP Technical Area Committee on Visual Information Processing.