IADIS International Journal on Computer Science and Information Systems
Vol. 18, No. 2, pp. 130-145
ISSN: 1646-3692

ANALYZING ENTROPY OF CONTROLLER MOVEMENTS AND MENTAL WORKLOAD IN YOUNG AND SENIOR USERS: AN APPLIED CASE FROM INDUSTRIAL TELEROBOTICS

Federica Nenna, Davide Zanardi and Luciano Gamberini
Department of General Psychology, University of Padova, Padova, Italy

ABSTRACT

Patterns of human motion were found to mirror cognitive processes in both psychological studies and applied research in Human-Computer Interaction (HCI). Notably, the behavioral entropy of movement trajectories of users was identified as a reflection of workload and fatigue across different settings, including Virtual Reality (VR). In the context of VR particularly, this metric is predominantly derived from the movements of VR controllers, denoted as the Entropy of Controller Movements (ECM). Despite its promising sensitivity and unobtrusive nature as a metric for human workload, ECM's application and proven efficacy in practical and authentic VR-based applications, such as industrial teleoperation platforms, has not been validated yet. Additionally, current literature predominantly features younger experimental samples, leaving unresolved the potential impact of age-related alterations in motor performance on using ECM as a workload metric. This study explored these dimensions by examining the relationship between workload and ECM among 15 young and 15 senior participants who manually operated an industrial robot within a VR environment. Participants were instructed to navigate the robot through a pick-and-place task by using their physical movements in VR. Our research identified unexpected variations in ECM values, particularly in older users, revealing an inverse relationship between movement entropy and task complexity in our scenario. High levels of behavioral entropy were also observed in younger participants. These findings unveil some criticalities in using ECM as a measure of workload in our VR-based industrial contexts, posing new questions regarding its applicability and effectiveness.

KEYWORDS

Behavioral Entropy, Virtual Reality, Workload, Industrial Robotics, Entropy of Controller Movements

1. INTRODUCTION

1.1 Virtual Reality in Manufacturing and Industrial Robotics

Industry 5.0 is enabling greater use of robotic and autonomous systems in manufacturing while bringing peculiar attention to humans. Always more considerations are done regarding human workload in the workplace (Coronado et al., 2022; Panagou et al, 2023) and a great deal of effort is spent on better designing human-machine interactions, to lower users' fatigue and provide workers with the best robotic support.

Lately, physical industrial settings are being transferred into digital platforms also known as digital twins. In digital representation, a "Digital Twin" functions as a virtual counterpart of a real-world entity or system. Harnessing sensor-derived data and historical records offers a high-fidelity simulation that facilitates precise monitoring, analysis, and optimization without direct intervention on the physical entity (Van der Valk et al., 2020). This technological advancement can enhance diagnostic, predictive and operational capabilities within various scientific and industrial applications. Such platforms, can exploit the most various interfaces, including Virtual Reality (VR) (Havard et al., 2019). VR is particularly attractive in this sector for the following reasons. First, VR can faithfully reproduce real scenarios, even the most complex ones, with high levels of immersion. This feature is significantly useful in the robotics and manufacturing industry, where it is crucial for usability test settings to be as close as possible to the real ones. Second, implementing novel features in physical robots for usability tests is time-consuming and requires a great deal of work (Dautenhahn, 2018). Also, interacting with autonomous machines can be particularly dangerous in some situations, or even impossible in others (e.g., Guo et al., 2021). By using simulation software in VR, it is possible to overcome issues related to the hardware, generate usability test settings, or even get immersed in physically unreachable environments in a much faster, more efficient, safer, and cheaper way (Dianatfar et al., 2021; Duguleana et al., 2011). Third, VR allows interactions with any virtual object exploiting human physical and embodied mechanisms. This calls to action human spatiality, which is the innate ability to act in physical space and thus facilitate any interaction with the virtual replica of physical objects. (Villani et al., 2018).

1.2 Methodological Challenge in user Performance and Workload Evaluation

When studying interactions between humans and robots, it is good practice to assess the human stress and workload for avoiding mental overload and the related psychophysical consequences (Fong and Nourbakhsh, 2004; Vidulich and Tsang, 2012). A broad body of literature in the work sector has demonstrated how workload and performance are strictly related, and how high workload levels are associated with mental fatigue, frustration, errors, and distraction (Galy et al., 2012; Young et al., 2015), which can have important consequences in HRI. Ideally, a proper measure of workload should avoid breaks in presence, should be highly sensitive and be unobtrusive (Wierwille and Eggemeier, 1993; Cain et al., 2017). However, because of its abstract nature, measuring workload is still a challenging task.

The most widely used methods for inferring workload are physiological metrics, questionnaires, and performance measures. Particularly, physiological indexes allow to quantify the psychophysical state of a user with high temporal accuracy. Heart Rate Variability (HRV), EEG or eye tracking are just a few examples of objective metrics that have been related to workload and other mental states. Nonetheless, they can show low sensitivity or differential diagnostic (Matthews et al., 2015). Questionnaires are between the most used measurement tools in the HRI field. One gold standard for measuring workload is the NASA Task Load Index (NASA-TLX, Hart and Staveland, 1988), which appears to be sensitive to task type and dual-tasking. However, self-report responses may be subject to biases, or they can miss facets of workload that are inaccessible to consciousness (Matthews et al., 2015). Also, when adopting questionnaires in a validation setting, it must be accounted the temporal gap between the experience and the assessment (Reinhardt et al., 2019). Finally, for performance measures one usually refers to task time and accuracy, which are well suited to continuous monitoring of workload and do not imply breaks in presence. Particularly for users executing tasks in virtual simulation settings, a continuous stream of implicit data can be recorded with high precision and at a high sample rate.

Such factors can have important consequences in industrial robotics, and therefore, are particularly worth measuring, also in VR. In this respect, all VR hardware (i.e., headset, controllers, body trackers) continuously generates time series data about their position and rotation. Such data can be leveraged to compute, for example, the start time and duration of any interaction with a virtual object, the users' movement velocity within a work setting, their position with respect to the digital replica of a robot or a machine, their task efficiency, or their work peace (Nenna et al., 2022; Nenna et al., 2023). All these metrics can report on the users' behaviors within a VR-based industrial or work context, enhancing our understanding of how people interact with virtual environments.

Remarkably, time series data on the VR controller position were further leveraged to gain insights into users' workload (Reinhardt et al., 2019; 2020). More precisely, the authors computed the Entropy of Controller Movement (ECM), which was demonstrated to be significantly modulated by the users' mental workload. While entropy measures have most likely been computed on gaze data and in desktop and mouse- or joystick-based settings within the human research areas (Goodrich et al., 2004; Reinhardt and Hurtienne, 2018; Stillman et al., 2018; Wu et al., 2020; Diaz-Piedra et al., 2019; Chatzithanos et al., 2021), the evidence of the effectiveness of such a measure within virtual environments opens new possibilities for continuous and indirect workload monitoring in VR-based industrial robotics.

1.3 Behavioral Entropy as a Measure of Workload

The concept of entropy refers to the degree of irregularity, randomness, and disorder in a system. It is typically used to quantify the complexity of different structures or processes or to learn about the randomicity of data or component variations (Wehrl, 1978). While it was initially developed to describe physical phenomena, the concept of entropy can extend to different kinds of data, including time series data, thus applying to various phenomena and application areas. In the HCI and ergonomics fields, entropy has been employed to infer human workload, fatigue, or decisional processes by analyzing the unpredictability of specific movement trajectories. Nakayama, Boer, and colleagues (1999; 2000) first employed measures of entropy in the steering wheel of a vehicle to estimate drivers' workload (i.e., steering entropy). Subsequently, this concept was generalized to human activity at large, which is currently known as behavioral entropy. Some studies computed behavioral entropy on users' movements via mouse (e.g., McKinstry et al., 2008; Reinhardt and Hurtienne, 2018; Stillman et al., 2018), or VR controllers (Reinhardt et al., 2019; 2020); some others also computed the entropy of eye movements as a measure of workload (e.g., Wu et al., 2020).

In all such cases, when individuals interact with technology, they perform task-related movements whose complexity and accuracy can be influenced by the task demand (e.g., complexity, precision). As the task complexity or precision increases, users may experience greater cognitive and physical demands, resulting in more random and unpredictable movement trajectories. More specifically, Goodrich et al. (2004) proposed that, when operators face high workload or other factors causing degraded performance, they might select less efficient behaviors, anticipate less and react more, therefore resulting in more fragmented or exaggerated actions (i.e., reactive behaviors). Under lower workloads, instead, operators are likely to perform anticipatory behaviors, which are smoother with lower magnitudes and less frequent changes. As a result, examining the entropy of human movements can provide insights into the levels of users' workloads. Goodrich et al. (2004) tested these hypotheses in seven users teleoperating a robot through direct or shared control, while additionally performing an arithmetic task. They concluded that behavioral entropy allowed to identify the most complex conditions of the teleoperation task. More recently, Chatzithanos et al. (2021) used behavioral entropy in a remote inspection scenario, whereby a sample of three operators teleoperated a robot to navigate an arena through a joypad. Different workload levels were created via dual- tasking and were directly related to the entropy values. On the measurement of behavioral entropy within immersive virtual environments, Reinhardt et al. (2019) measured the Entropy of Controller Movements (ECM) in twenty-two participants executing a simple rhythm game in VR. They found a clear relation with their workload, indicating ECM as a promising mental workload measurement even in VR. In a subsequent experiment, the same authors tested twenty students performing the e-crossing task in VR, and demonstrated positive relations between the task difficulty, mental workload reported at the NASA-TLX and ECM (Reinhardt et al., 2020). While the literature on behavioral entropy in immersive VR is not that extensive, ECM seems to have great potential as a highly sensitive and unobtrusive measure of workload in various VR scenarios. Nonetheless, to the best of our knowledge, literature assessing the effectiveness of ECM in more applied and realistic instances of VR-based environments (e.g., teleoperation platforms) is missing.

1.4 Age-Related Factors in VR

As described above, ECM represent a prosing technique for measuring workload in fields like VR-based smart manufacturing. However, the prevailing demographic diversity in experimental samples poses pertinent questions about the universal applicability of ECM across different age groups. The current literature predominantly features younger participants, underlining a gap in our understanding of how age influences the effectiveness of ECM as a measure of workload in VR settings. Moreover, a prevalent lack of technological literacy amongst older adults could impede VR's feasibility and applicability. Similarly, numerous barriers, including learning costs and usability, can potentially hinder their relationship with technology (Barnard et al., 2013). Studies on this matter, conducted by Adami and colleagues (2021) and Chen and Or (2017), exemplify the age-related challenges in leveraging VR, highlighting how older participants often derive less benefit from VR-based training programs due to decreased acquired knowledge and elevated risk behavior and struggle with more errors and longer durations in VR environments, respectively. Conversely, research also offers instances of successful VR implementations among older subjects, demonstrating comparable levels of acceptance of VR applications between younger and older populations. The studies by Syed-Abdul et al. (2019) and Ijaz et al. (2019) exemplify how perceptions of utility and ease of use can elevate the willingness among older adults to adopt VR technologies, contributing to a reduction in stress levels and an enhancement in task performance and intention to use VR. The physical and intuitive interactivity offered by VR makes it a natural interface, which could actually be particularly beneficial for older individuals. Findings in this direction were by studies showcasing superior gestural performance in older adults (Carvalho et al., 2017). Nonetheless, the trajectory towards more natural, body-dependent interaction modalities is fraught with impediments, specifically for older workers in manufacturing systems who often exhibit deteriorations in physical performance, manual dexterity, movement speed, and motor flexibility (Ketcham et al., 2002; Pennathur et al., 2003; Verrel et al., 2012). The onset of physical declines can present significant obstacles when adopting modalities of gestural interactivity. These modalities often require a more refined level of visual-motor coordination than what is needed with more conventional devices, which could be potentially altered in older individuals, also altering the potential feasibility of measures such as ECM. In fact, applying ECM as a metric to assess cognitive load can also reflect the repercussions of these age-related motor changes. The observed higher levels of movement irregularity and dispersion may predominantly be a consequence of diminished motor control performance rather than a direct result of increased fatigue experienced by the users. This understanding is required as it underlines the importance of consider the impacts of motor control impairments from the effects of fatigue when assessing gestural interactivity, ensuring a more precise and comprehensive interpretation of user experience and interaction efficacy in diverse age groups and operational contexts.

In conclusion, while the integration of ECM in VR-based interaction provides promising insights into workload measurement and interactivity, it is necessary to conduct a more comprehensive and more profound investigation of older individuals' diverse needs and capabilities within this domain. Conducting such an investigation would facilitate the development of more inclusive and adaptable environments in smart manufacturing.

2. THE PRESENT STUDY

With this study, we aim at covering the identified gaps and learn more about the effectiveness of ECM as a measure of workload in VR. We thus used a VR-based industrial robotics scenario (Nenna et al., 2022, 2023), and compared results obtained from participants under 30 years of age (young) with those over 50 years of age (senior), as illustrated in Figure 1. Specifically, a total of thirty participants - divided into two age groups - guided an industrial robotic arm through a pick-and-place task in VR by physically moving their arms, under low (single-task) and high (dual-task) mental demands. We also administered the NASA-TLX questionnaire to reference the participants' perceived workload after both task conditions.

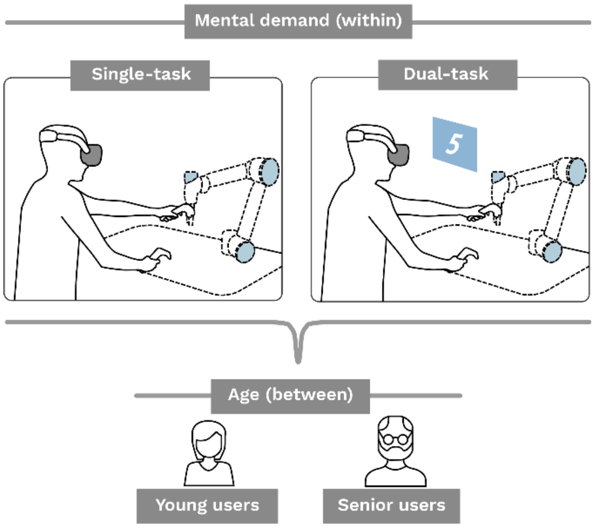

Figure 1. Illustration of the experimental design, utilizing a 2x2 mixed factorial model, where two distinct participant groups, categorized as young and senior, executed a pick-and-place task under varying conditions of mental demand—low (single-task) and high (dual-task)

Our hypotheses can be outlines as follows.

H1) We expect to observe elevated ECM values in the dual-task condition compared to the single-task condition, reflecting different workloads (Nenna et al., 2023). This hypothesis is based on existing literature, which illustrates a correlation between heightened task complexity and increased ECM, reflecting cognitive and physical demands (McKinstry et al., 2008; Reinhardt and Hurtienne, 2018; Stillman et al., 2018; Reinhardt et al., 2019; 2020).

H2) We posit a generally higher ECM in the senior compared to the young group. This hypothesis is substantiated by extensive literature indicating that older individuals are prone to decline in physical performance and motor skills (Ketcham et al., 2002; Pennathur et al., 2003; Verrel et al., 2012), which might unveil less smooth motor paths when driving the robot via physical actions and also higher demands. Additionally, literature suggests how older users often encounter more difficulties leveraging VR interfaces, particularly when compared to their younger counterparts (Chen and Or, 2017; Adami et al., 2021), which might influence their ECM trends as well.

3. METHODOLOGY

3.1 Participants

The experimental sample consisted of 15 individuals (9 females) composed the senior group, who reported being more than 50 years old (M_{age}= 57.1, SD_{age}= 6.2), and 15 individuals (6 females) the young group, who reported being less than 30 years old (M_{age}= 27.8, SD_{age}= 6.4). All participants signed informed consent. The inclusion criteria were the following:

absence of past or present neurological/psychiatric disorders, being right-handers, possessing normal or corrected-to-normal vision with contact lenses, and normal color vision. The local ethics committee approved the research methodology, and the study was conducted following the principles of the Declaration of Helsinki. One senior participant was excluded due to limited proficiency with technological devices, which caused extremely long training and a severe inability to perform the task. All participants reported to be inexperienced with VR and telerobotics, particularly the senior once.

3.2 Technical Setup and Experimental Procedure

Participants were provided with HTC VIVE Pro Eye and both its controllers. The virtual environment was programmed in Unity (version 2020.2.1f1) and was validated in previous studies (Nenna et al., 2022, 2023). All data was automatically saved on the internal storage of the local laboratory computer at the end of each experimental session.

Before starting the experiment, participants conducted a training session based on the same tasks used in the experimental phase to familiarize themselves with the virtual environment and minimize individual differences related to the ability to use the virtual system. Afterward, during the experimental session, all subjects controlled a robotic arm in VR to execute a pick-and-place task under different demands (single- and dual-task) presented in random order. The young group completed 40 trials for each experimental condition. Differently, for the senior group, the number of trials was lowered to 20, as they showed longer training duration and greater difficulties in familiarizing with the tasks. This may be due to their lack of technological literacy (Wildenbos et al., 2018), which may pose a barrier to the repeated execution of VR-based tasks for older adults.

In the single-task, participants were called to physically drive the robot to pick a bolt from the workstation and place it into a blue box. Figure 2 depicts one complete task trial, which was split into two task phases: the Pick phase, which required higher movement precision to accurately align the robot effector with the bolt, and the Place phase, which required less movement precision as the box where to release the bolt is provided with a larger area. Given that the two task phases demanded different levels of precision of the movement trajectories, we computed the ECM within each of them independently. Oppositely to the single-task, in the dual-task participants additionally performed an arithmetic task to create a higher level of task demand. Specifically, they were presented with a series of digits randomly ranging between 1 and 10 every 2sec, with a jitter of 0.3sec. They were thus asked to sum the numbers all the way through the trial, and then report the result of the mental calculations on a virtual numeric keyboard once the trial was completed. Such paradigm was already validated in previous investigations (Nenna et al., 2022, 2023).

For guiding the robot, participants approached their right hand to the robot effector and then grasped it by pressing the grip button on the controller. Therefore, they dragged the robot to the desired position by physically moving their own arm within the virtual space, producing a movement trajectory. To enable the picking or placing operations, they then pressed the pad button on the left controller and the robot automatically went down on the workstation to either pick or place the bolt. Once the bolt entered the box, a new bolt and box randomly appeared on the workstation, and a new trial started.

3.3 Measurements

The levels of perceived mental workload were measured through the NASA-TLX questionnaire (Hart and Staveland, 1988) after the single- and the dual-task. The global workload score was calculated by averaging the scores in its six sub-scales (mental demand, physical demand, temporal demand, frustration, effort, and performance).

For the computation of behavioral entropy, and more specifically ECM, we opted for individually addressing the pick and place phases trajectories, as they require different levels of movement precision which might impact the entropy values. Therefore, we isolated the movement trajectories performed during the pick and the place action. We only included trials exempt from movement disruptions (e.g., the participant lost the robot grip and returned to grasp the robot again, causing fragmented trajectories). In this way, all actions under examination were considered smooth, continuous, and uninterrupted.

For the entropy computation, we then choose to determine sample entropy using the method of Richman and Moorman (2000). This method best fits the randomness intrinsic to systems behaving in real-world or complex environments and has been demonstrated to be the preferred method in applied research for mental workload assessments, also in VR scenarios (Reinhardt et al., 2019). Sample entropy can be defined as the negative logarithm of the probability that if two sets of data points of length *m* are similar, they will remain similar at *m*+1 (Richman and Moorman, 2000). Therefore, if areas in a trajectory that appear similar at one length are no longer similar at a greater length, greater dispersion and complexity are observed in a trajectory, which increases the sample entropy (Hehman et al., 2015). For its computation, we converted the data to normalized time and shifted the absolute controller positions on the three axes (x, y, z) to always start from zero (0, 0, 0) in each trial. Based on the overview of statistical tools for calculating sample entropy of Chen and colleagues (2019), we used the function *sample_entropy* in Rstudio (R Core Team, 2022) from the package *pracma* (Borchers, 2023), which allows setting specific parameters like the embedding dimension *m* (length of sequences to be compared for similarity) and the tolerance *r* (the threshold for determining similarity between windows). By following the approaches of Hehman et al. (2015) and Reinhardt et al. (2019), we set m=2 and r=0.2; thus, we compared windows with a length of 2, and each sequence was determined to be similar if it was within a tolerance of 0.2 multiplied by the standard deviation of the data. We finally calculated the ECM on the three axes (ECM-X, ECM-Y, ECM-Z) and the ECM-total by averaging the three individual ECMs (Reinhardt et al., 2019).

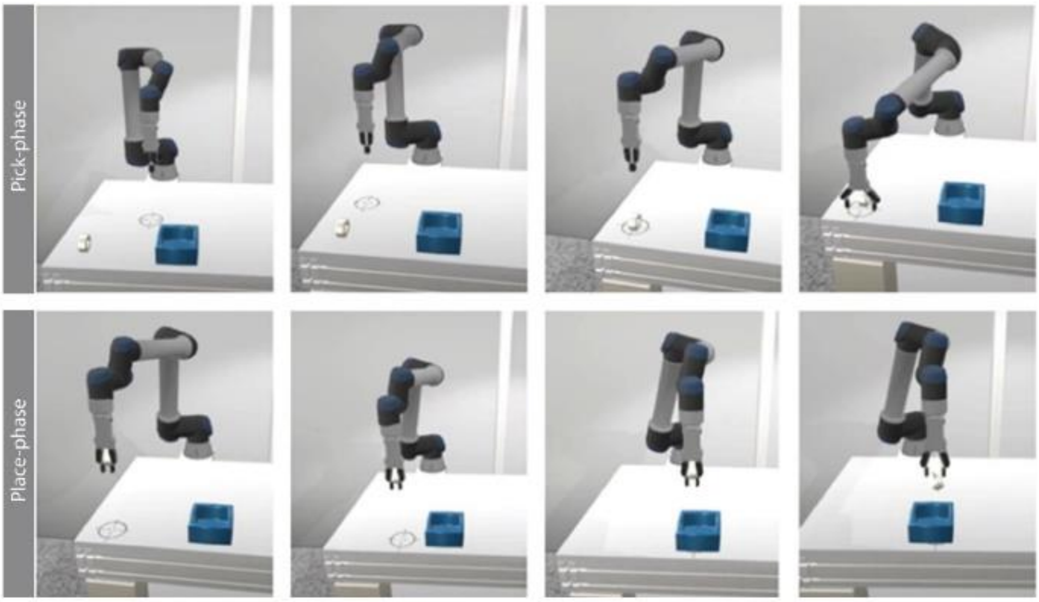

Figure 2. Depiction of one trial of the pick-and-place task, divided into two task phases (i.e., pick and place)

3.4 Statistical Analysis

All data were analyzed through Generalized Linear Models (GLMs from lme4 package, Bates et al., 2014), with Participant as a random effect. We computed a model for each ECM calculation (ECM-X, ECM-Y, ECM-Z, ECM-total) over the factors Task (single-task, dual-task) and Age (young, senior). For the analysis of the responses to the NASA-TLX questionnaire, we additionally included the factor Item (mental demand, temporal demand, physical demand, performance, effort, frustration) to further explore possible differences within each of the questionnaire sub-scales. Specifically, each model was chosen after first fitting the data through the function descdist() of the package fitdistrplus (Delignette-Muller and Dutang, 2015), which allowed choosing the appropriate model setting based on data distribution. Post hoc contrasts were performed on each significant interaction with the application of the Bonferroni correction for multiple comparisons (Bonferroni, 1936).

4. RESULTS

The analysis of the NASA-TLX revealed a main effect for Task ($X^2 = 121.58$, $p < 0.001$) and Item $X^2 = 81.26$, $p < 0.001$). Significant interactions were observed between Task and Item ($X^2 = 41.78$, $p < 0.001$) and Item and Age ($X^2 = 16.57$, $p < 0.01$). No significant main effect was found for Age ($p = 0.19$).

Table 1. Descriptive statistics of the NASA-TLX questionnaire scores

Dimension	Young Single-task M ± SD	Young Dual-task M ± SD	Senior Single-task M ± SD	Senior Dual-task M ± SD
Global score	5.79 ± 4.85	10.30 ± 5.28	5.54 ± 4.87	11.5 ± 5.89
Mental demand	2.67 ± 1.35	13.80 ± 4.25	4.14 ± 4.44	15.6 ± 2.71
Physical demand	8.40 ± 5.47	11.9 ± 4.48	5.21 ± 4.02	8.14 ± 5.63
Temporal demand	7.27 ± 3.47	7.00 ± 3.40	7.14 ± 4.75	11.2 ± 5.58
Performance	4.33 ± 4.75	7.93 ± 5.64	4.43 ± 3.39	9.36 ± 5.49
Effort	8.27 ± 5.18	13.9 ± 3.83	9.71 ± 6.70	16.8 ± 2.94
Frustration	2.73 ± 2.76	7.47 ± 5.01	2.57 ± 1.40	8.14 ± 5.95

Table 2. Descriptive statistics of the ECMs

Task phase	Axis	Young Single-task M ± SD	Young Dual-task M ± SD	Senior Single-task M ± SD	Senior Dual task M ± SD
Pick	ECM-x	0.026 ± 0.013	0.017 ± 0.015	0.021 ± 0.026	0.010 ± 0.008
	ECM-y	0.199 ± 0.099	0.137 ± 0.086	0.112 ± 0.089	0.104 ± 0.085
	ECM-z	0.093 ± 0.084	0.080 ± 0.041	0.052 ± 0.060	0.033 ± 0.037
	ECM-tot	0.106 ± 0.048	0.078 ± 0.028	0.062 ± 0.040	0.049 ± 0.034
Place	ECM-x	0.044 ± 0.038	0.045 ± 0.042	0.023 ± 0.029	0.019 ± 0.025
	ECM-y	0.216 ± 0.144	0.205 ± 0.145	0.145 ± 0.100	0.111 ± 0.090
	ECM-z	0.064 ± 0.065	0.068 ± 0.072	0.037 ± 0.050	0.028 ± 0.048
	ECM-tot	0.108 ± 0.059	0.106 ± 0.058	0.068 ± 0.042	0.053 ± 0.038

Post-hoc tests on the Task-Item interaction showed significantly higher levels of mental demand ($p<.0001$), effort ($p < 0.0001$), and frustration ($p < 0.001$) in the dual- compared to the single-task. Differently, no significant contrasts were found for the Age-Item interaction. Descriptive statistics are resumed in Table 1. For the analysis of behavioral entropy, we resumed the descriptive statistics in Table 2, the results of the GLMs in Table 3, and all the post-hoc contrasts are depicted in Figure 3.

Table 3. Results of the GLMs performed in the pick and place phases on each ECM measure

Axis	Task phase	Task	Age	Task * Age
ECM-x	Pick	$X^2 = 66.66$ ***	$X^2 = 1.62$ ns	$X^2 = 4.03$ *
	Place	$X^2 = 0.04$ ns	$X^2 = 28.53$ ***	$X^2 = 3.07$ ns
ECM-y	Pick	$X^2 = 1.47$ ns	$X^2 = 7.79$ **	$X^2 = 1.33$ ns
	Place	$X^2 = 7.61$ **	$X^2 = 13.22$ ***	$X^2 = 10.53$ **
ECM-z	Pick	$X^2 = 7.27$ **	$X^2 = 15.44$ ***	$X^2 = 0.28$ ns
	Place	$X^2 = .008$ ns	$X^2 = 36.49$ ***	$X^2 = 8.76$ **
ECM-tot	Pick	$X^2 = 15.73$ ***	$X^2 = 17.38$ ***	$X^2 = 0.05$ ns
	Place	$X^2 = 5.52$ *	$X^2 = 22.85$ ***	$X^2 = 18.58$ ***

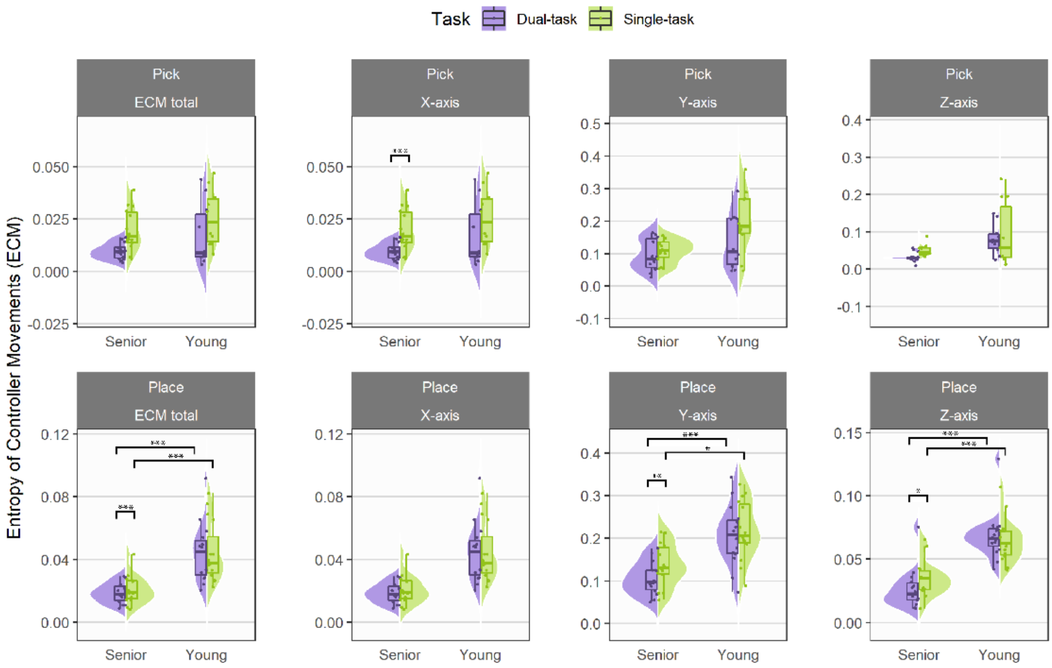

Figure 2. Averaged behavioral entropy in each task phase (pick, place) and task condition (dual-task, single-task), divided by age (senior, young). The dots on the boxplots indicate each participant's averaged entropy. The half violins show the data distribution. The stars indicate the significance levels of the post hoc tests (*p ≤ .05; **p ≤ .01; ***p ≤ .001)

5. DISCUSSION

In this study, we applied the concept of behavioral entropy to the trajectory of VR controller movements, exploring its effectiveness in measuring human workload within a VR-based robotic teleoperation scenario. We further explored possible differences between young and senior users, who are prone to decreasing smoothness of movements (Seidler et al., 2002), which might affect the effectiveness of ECM as a measure of workload. We thus defined different mental demands through the dual-task methodology, asking participants to physically drive a robotic arm once as a single task, and once concurrently with an arithmetic task (dual-task). As indicated by the NASA-TLX questionnaire, our dual-task manipulation effectively induced different workload levels through the task. Specifically, regardless of participants' age, the dual-task elicited significantly higher mental demand, effort, and frustration than the single-task. Interestingly, both young and senior participants self-reported similar workload levels, suggesting no age-dependent differences in perceived workload.

However, unexpectedly, our findings on the ECM deviate from what previously observed (i.e., Chatzithanos et al., 2021; Goodrich et al., 2004; Reinhardt et al., 2019). Specifically, we first hypothesized (H1) higher ECM in the most demanding task condition (i.e., dual-task).

Contrary to our hypotheses, our observations revealed divergent trends across the different age groups. Our data about the senior group unveiled elevated ECM values in the single-task condition compared to the dual-task condition. ECM-x predominantly evidenced this tendency during the pick phase of the experiment. Concurrently, this observation was corroborated by the ECM-total, ECM-y, and ECM-z during the place phase, highlighting a coherent pattern across different ECM measurements. These collective findings explain an inverse relationship between the entropy of movement trajectories and the complexity of the experimental task, differently from what was previously evidenced in different VR scenarios (Reinhardt et al., 2019). On this matter, we argue that the secondary task may have intrusively interfered with the primary pick-and-place, influencing seniors' motor trajectories. Specifically, in the dual-task, participants summed a series of numbers throughout each pick-and-place action and were always instructed to be as fast and accurately as possible in both tasks. However, the faster they placed the bolt into the box, the fewer mental calculations they had to compute. Differently, in the single-task, they were free to act impulsively while following the instruction of being both fast and accurate. This may have led to more conscious, regular, and smooth motor trajectories in the dual-task to correctly pick and then place the bolt as fast as possible, and more impulsive and dispersive trajectories in the single-task.

Conversely, the younger group exhibited no significant difference in ECM across various task conditions. Despite observing variations in mental workload, as evidenced by the self-report measurements in our VR-based telerobotic framework, such distinctions were not identifiable when utilizing ECM measurements. On this regard: previous research, which demonstrated the efficacy of behavioral entropy in detecting varying levels of workload resulting from dual-tasking, was conducted in desktop environments using either a mouse or joypad (Chatzithanos et al., 2021; Goodrich et al., 2004). However, to the best of our knowledge, no study investigated the effectiveness of behavioral entropy in VR contexts leveraging dual-tasking. In their VR-based experiments on behavioral entropy, Reinhardt and colleagues (2019, 2020) always used some inhibition tasks to increase the levels of demand on the user, while no instances of ECM under VR dual-tasking were ever provided. It is thus possible that the presence of a secondary task changes the nature of the primary task, particularly when young users physical motion with higher degrees of freedom is involved, like in our VR contexts. This underlines the necessity of a better understanding of behavioral entropy sensitivity in VR, by also considering the impact of secondary tasks on human motion trajectories.

As a second hypothesis (H2), we expected the senior group to exhibit higher ECM values, potentially mirroring their age-related modifications in movement trajectories (Seidler et al., 2002). In contrast, our data shown a consistently elevated behavioral entropy in young participants compared to senior participants, particularly prominent in the place phase. This observation contrasts our initial expectations, leaving us unable to conclusively determine whether the reduced ECM in senior participants derived from a lack of age-related motor impairments or a differing behavioral approach to task execution relative to younger users. It is to be noted that this acknowledgment should not be interpreted as indicative of disparities in perceived workload; both young and senior participants recorded analogous levels of workload on the NASA-TLX. Hence, it is plausible that other external factors, such as unfamiliarity with VR environments and lower technological literacy, significantly influenced their motor behavior. From this perspective, senior users, presumably interacting with virtual environments for the first time, might have opted for more smoothed and cautious maneuvers to operate the virtual robot, reflecting their tentative approach. Contrastingly, the younger group, generally more acclimated to VR technologies and, in some instances, with prior exposure to advanced technological interfaces, adapted swiftly to the VR environment. Consequently, their comfort and familiarity possibly allowed them to exhibit more impulsive and dispersive actions, culminating in elevated behavioral entropy.

6. CONCLUSIONS

In this study, we explored the feasibility of behavioral entropy to assess human workload in VR-based robotic teleoperations, considering potential differences between young and senior users. By employing a dual-task methodology, we defined conditions of low and high mental demands, requiring participants to operate a robotic arm both independently and concurrently with an arithmetic task.

Deviating from findings in existing literature, our results exhibited an increase in ECM values under conditions of low mental demand for the senior group, and no observable difference among young participants. Furthermore, we noted a generally higher ECM in younger participants compared to the senior ones.

While our results contribute to shed light on the feasibility of behavioral entropy across aging in the VR context, we must acknowledge certain limitations of our study. For instance, we did not analyze thoroughly the behavioral aspects related to our participants' movements. By incorporating such biomechanical metrics, we could deepen our understanding of the groups' diverse motor strategies when physically driving the robot via VR controllers. Also, the chosen dual task paradigm might have affected the users' motor strategies, and consequently, their ECM. Further studies might explore how various tasks shape the ECM.

Overall, behavioral entropy, and ECM in VR, represent valuable unobtrusive and easily computable measures to trace the dispersion of human motion trajectories in digital environments. However, new questions were raised about its effectiveness in reflecting users' workload in VR. Future research should delve into these questions, embracing a holistic evaluation approach to understand better the factors that may influence human motion strategies in VR and the degree to which behavioral entropy can inform users' workloads across varied task scenarios.

ACKNOWLEDGEMENT

We thank Davide Gobbo for his support on the development of the virtual environment. This study was carried out within the scope of the project "use-inspired basic research", for which the Department of General Psychology of the University of Padova has been recognized as "Dipartimento di eccellenza" by the Italian Ministry of University and Research.

REFERENCES

Adami, P. et al. (2021). Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Advanced Engineering Informatics*, Vol. 50, 101431.

Barnard, Y., Bradley, M. D., Hodgson, F. and Lloyd, A. D. (2013). Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. *Computers in human behavior*, Vol. 29, No. 4, pp. 1715-1724.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, Vol. 67, No. 1, pp. 1-48. https://doi.org/10.18637/jss.v067.i01

Boer, E. R. (2000). Behavioral entropy as an index of workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Los Angeles, CA, USA, Vol. 44, No. 17, pp. 125-128.

Bonferroni, C. (1936). *Teoria statistica delle classi e calcolo delle probabilita*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze (8th ed.), Seeber.

Borchers, H. W. (2023). *Pracma: Practical Numerical Math Functions* (2.4.4). https://CRAN.R-project.org/package=pracma

Cain, C. L. et al. (2017). A mixed methods study of emotional exhaustion: Energizing and depleting work within an innovative healthcare team. *Journal of Interprofessional Care*, Vol. 31, No. 6, pp. 714-724.

Carvalho, D. et al. (2017). Age group differences in performance using distinct input modalities: A target acquisition performance evaluation. *2017 24º Encontro Português de Computação Gráfica e Interação (EPCGI)*, pp. 1-8.

Chatzithanos, P., Nikolaou, G., Stolkin, R. and Chiou, M. (2021). Fessonia: A method for real-time estimation of human operator workload using behavioural entropy. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1325-1331.

Chen, C. et al. (2019). A comprehensive comparison and overview of R packages for calculating sample entropy. *Biology Methods and Protocols*, Vol. 4, No. 1, bpz016.

Chen, J. and Or, C. (2017). Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. *Applied Ergonomics*, Vol. 65, pp. 437-448.

Coronado, E. et al. (2022). Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0. *Journal of Manufacturing Systems*, Vol. 63, pp. 392-410.

Dautenhahn, K. (2018). Some brief thoughts on the past and future of human-robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)*, Vol. 7, No. 1, pp. 1-3.

Delignette-Muller, M. L. and Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, Vol. 64, pp. 1-34.

Dianatfar, M., Latokartano, J. and Lanz, M. (2021). Review on existing VR/AR solutions in human–robot collaboration. *Procedia CIRP*, Vol. 97, pp. 407-411.

Diaz-Piedra, C. et al. (2019). The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Applied Ergonomics*, Vol. 77, pp. 92-99.

Duguleana, M., Barbuceanu, F. G. and Mogan, G. (2011). Evaluating human-robot interaction during a manipulation experiment conducted in immersive virtual reality. In R. Shumaker (eds) *Virtual and Mixed Reality - New Trends. VMR 2011.* Lecture Notes in Computer Science, vol 6773. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22021-0_19

Fong, T. and Nourbakhsh, I. R. (2004, October). *Peer-to-Peer Human-Robot Interaction for Space Exploration* (AAAI Technical Report No. 5, pp. 87-90).

Galy, E., Cariou, M. and Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, Vol. 83, No. 3, pp. 269-275.

Goodrich, M. A. et al. (2004). *Behavioral entropy in human-robot interaction* (Report). BRIGHAM YOUNG UNIV, PROVO, UT. https://apps.dtic.mil/sti/pdfs/ADA446467.pdf

Guo, Y., Freer, D., Deligianni, F. and Yang, G. Z. (2021). Eye-tracking for performance evaluation and workload estimation in space telerobotic training. *IEEE Transactions on Human-Machine Systems*, Vol. 52, No. 1, pp. 1-11.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, Vol. 52, pp. 139-183.

Havard, V., Jeanne, B., Lacomblez, M. and Baudry, D. (2019). Digital twin and virtual reality: A co-simulation environment for design and assessment of industrial workstations. *Production & Manufacturing Research*, Vol. 7, No. 1, pp. 472-489.

Hehman, E., Stolier, R. M. and Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, Vol. 18, No. 3, pp. 384-401.

Ijaz, K., Ahmadpour, N., Naismith, S. L. and Calvo, R. A. (2019). An immersive virtual reality platform for assessing spatial navigation memory in predementia screening: feasibility and usability study. *JMIR mental health*, Vol. 6, No. 9, e13887.

Ketcham, C. J., Seidler, R. D., Van Gemmert, A. W. and Stelmach, G. E. (2002). Age-related kinematic differences as influenced by task difficulty, target size, and movement amplitude. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 57, No. 1, pp. P54-P64.

Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., & Abich, J. (2015). Workload is multidimensional, not unitary: what now? *Proceedings of the Foundations of Augmented Cognition: 9th International Conference, AC 2015, held as Part of HCI International 2015*. Springer International Publishing, Los Angeles, CA, USA, pp. 44-55.

McKinstry, C., Dale, R. and Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, Vol. 19, No. 1, pp. 22-24.

Nakayama, O., Futami, T., Nakamura, T. and Boer, E. R. (1999). Development of a steering entropy method for evaluating driver workload. *SAE Transactions*, Vol. 108, pp. 1686-1695.

Nenna, F., Orso, V., Zanardi, D. and Gamberini, L. (2022). The virtualization of human–robot interactions: A user-centric workload assessment. *Virtual Reality*, 1-19.

Nenna, F., Zanardi, D. and Gamberini, L. (2023). Enhanced Interactivity in VR-based Telerobotics: An Eye-tracking Investigation of Human Performance and Workload. *International Journal of Human-Computer Studies*, 103079. https://doi.org/10.1016/j.ijhcs.2023.103079

Panagou, S., Neumann, W. P. and Fruggiero, F. (2023). A scoping review of human robot interaction research towards Industry 5.0 human-centric workplaces. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2023.2172473

Pennathur, A., Contreras, L. R., Arcaute, K. and Dowling, W. (2003). Manual dexterity of older Mexican American adults: A cross-sectional pilot experimental investigation. *International Journal of Industrial Ergonomics*, Vol. 32, No. 6, pp. 419-431.

Reinhardt, D. and Hurtienne, J. (2018). Cursor entropy reveals decision fatigue. *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. Article No. 31, pp. 1-2.

Reinhardt, D., Haesler, S., Hurtienne, J. and Wienrich, C. (2019). Entropy of controller movements reflects mental workload in virtual reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 802-808.

Reinhardt, D., Hurtienne, J. and Wienrich, C. (2020). Measuring Mental Effort via Entropy in VR. *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, pp. 43-44).

Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology-heart and circulatory physiology*, Vol. 278, No. 6, pp. H2039–H2049.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Seidler, R. D., Alberts, J. L. and Stelmach, G. E. (2002) Changes in multi-joint performance with age. *Motor control*, Vol. 6, No. 1, pp. 19-31.

Syed-Abdul, S. et al. (2019). Virtual reality among the elderly: A usefulness and acceptance study from Taiwan. *BMC geriatrics*, Vol. 19, No. 1, pp. 1-10.

Stillman, P. E., Shen, X. and Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in cognitive sciences*, Vol. 22, No. 6, pp. 531-543.

Van der Valk, H. et al. (2020). A Taxonomy of Digital Twins. *26th Americas Conference on Information Systems (AMCIS)*, Salt Lake City, USA.

Verrel, J., Lövdén, M. and Lindenberger, U. (2012). Normal aging reduces motor synergies in manual pointing. *Neurobiology of Aging*, Vol. 33, Issue 1, pp. 200.e1-200.e10.

Vidulich, M. A. and Tsang, P. S. (2012). Mental workload and situation awareness. *Handbook of human factors and ergonomics*, pp. 243-273.

Villani, V., Capelli, B. and Sabattini, L. (2018). Use of virtual reality for the evaluation of human-robot interaction systems in complex scenarios. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 422-427.

Wehrl, A. (1978). General properties of entropy. *Reviews of Modern Physics*, Vol. 50, No. 2.

Wierwille, W. W. and Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human factors*, Vol. 35, No. 2, pp. 263-281.

Wildenbos, G. A., Peute, L. and Jaspers, M. (2018). Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). *International Journal of Medical Informatics*, Vol. 114, pp. 66-75.

Wu, C. et al. (2020). Eye-tracking metrics predict perceived workload in robotic surgical skills training. *Human factors*, Vol. 62, No. 8, pp. 1365-1386.

Young, M. S., Brookhuis, K. A., Wickens, C. D. and Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, Vol. 58, No. 1, pp. 1-17.