



HAL
open science

Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task

Unnikrishnan Radhakrishnan, Lisheng Kuang, Konstantinos Koumaditis, Francesco Chinello, Claudio Pacchierotti

► **To cite this version:**

Unnikrishnan Radhakrishnan, Lisheng Kuang, Konstantinos Koumaditis, Francesco Chinello, Claudio Pacchierotti. Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task. IEEE Transactions on Haptics (ToH), 2024, 17 (2), pp.249-262. hal-04216606

HAL Id: hal-04216606

<https://inria.hal.science/hal-04216606>

Submitted on 25 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Haptic Feedback, Performance and Arousal: A Comparison Study in an Immersive VR Motor Skill Training Task

Unnikrishnan Radhakrishnan, Lisheng Kuang, Konstantinos Koumaditis,
Francesco Chinello, Claudio Pacchierotti

Abstract—This paper investigates the relationship between fine motor skill training in VR, haptic feedback, and physiological arousal. To do so, we present the design and development of a motor skill task (buzzwire), along with a custom vibrotactile feedback attachment for the Geomagic Touch haptic device. A controlled experiment following a between-subjects design was conducted with 73 participants, studying the role of three feedback conditions – visual/kinesthetic, visual/vibrotactile and visual only – on the learning and performance of the considered task and the arousal levels of the participants. Results indicate that performance improved in all three feedback conditions after the considered training session. However, participants reported no change in self-efficacy and in terms of presence and task load (NASA-TLX). All three feedback conditions also showed similar arousal levels. Further analysis revealed that positive changes in performance were linked to higher arousal levels. These results suggest the potential of haptic feedback to affect arousal levels and encourage further research into using this relationship to improve motor skill training in VR.

Index Terms—Virtual Reality, Haptics, Physiological Arousal, Vibrotactile Feedback, Kinesthetic Feedback

I. INTRODUCTION

VIRTUAL Reality (VR) training is gaining popularity in medicine, rehabilitation, and industry, addressing various psychomotor, procedural, spatial, and decision-making skills [1], [2]. VR training provides virtual environments where trainees receive consistent and replicable training, allowing for an objective assessment of skills. However, evidence supporting the advantages of VR training over other methods is mixed. For example, in a Cochrane meta-analysis of VR-based endoscopy training literature, Khan et al. [3] found that though VR training is better than no training, it is not better than conventional training, whereas Mekbib et al. [4] in a meta-analysis of upper limb rehabilitation literature found that VR was better than conventional therapy. The effectiveness of VR for training may be enhanced by improving the sensory fidelity of VR, i.e., the “immersion” provided by the system through the use of head-mounted displays (HMDs) or CAVE Automatic Virtual Environments (CAVEs) [5]. A more immersive and interactive VR system may indeed lead to higher perceived presence (the subjective response to immersion) [5], which in turn positively affects the effectiveness of VR training [6]. For the same reasons, haptic feedback in VR may increase presence and potentially training performance [7]–[9]. However,

employing haptics remains relatively unexplored in motor skill training literature in immersive VR (IVR) outside the surgical and rehabilitation domains. Therefore, further investigation, not only on the haptic feedback but also on its modalities and variations, can aid the discussion of how VR training might be enhanced [2]). Inspired by examples of motor skill training from immersive VR [10], [11], this study focuses on a buzzwire (or wire loop) task, where the aim is to move a metallic loop across a wire without touching it. Radhakrishnan et al. [11] proved that, in a similar buzzwire training scenario, IVR training was as effective as physical training in improving task performance. A similar setup is used in this study to investigate the following research questions:

RQ 1: *Can vibrotactile or kinesthetic feedback influence VR training performance in a buzzwire motor skill task?*

RQ 2: *Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in arousal levels during training? Is there a link between physiological arousal during training and changes in performance afterward?*

RQ 3: *Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in reported presence, task load, and self-efficacy?*

II. RELATED WORKS

A. Haptics in immersive VR training

In a systematic review and meta-analysis of the use of different types of haptic feedback on VR and box trainers in laparoscopic surgical skills, Overtoom et al. [12] found that the addition of haptics provides only a small positive effect on task performance while providing a better learning curve at the beginning of training as compared to no haptics conditions. Similarly, Rangarajan et al. found that haptics enhanced surgical training further than training without haptics [13]. They also found that the addition of haptic feedback reduced the learning curve for novice trainees. In fact, different modalities of haptic feedback have been documented in the literature, such as kinesthetic, vibrotactile and mid-air haptics. In this paper, we focus on kinesthetic and vibrotactile feedback modalities as they are suitable for providing relevant information regarding mistakes in the buzzwire task. Kinesthetic feedback is closest to recreating real-world physical forces using grounded haptic devices; however, it faces the issue of the relatively high cost of equipment and programming complexity to simulate haptic interaction efficiently. On the other hand, vibrotactile feedback is a form of cutaneous feedback that uses vibrations to

U. Radhakrishnan, K. Koumaditis and F. Chinello are with the Dept. of Business Development and Technology, Aarhus University (Herning, Denmark). E-mail: {unnik, kkoumaditis, chinello}@btech.au.dk

L. Kuang, and C. Pacchierotti are with CNRS, Univ Rennes, Inria, IRISA (Rennes, France). E-mail: {lisheng.kuang, claudio.pacchierotti}@irisa.fr

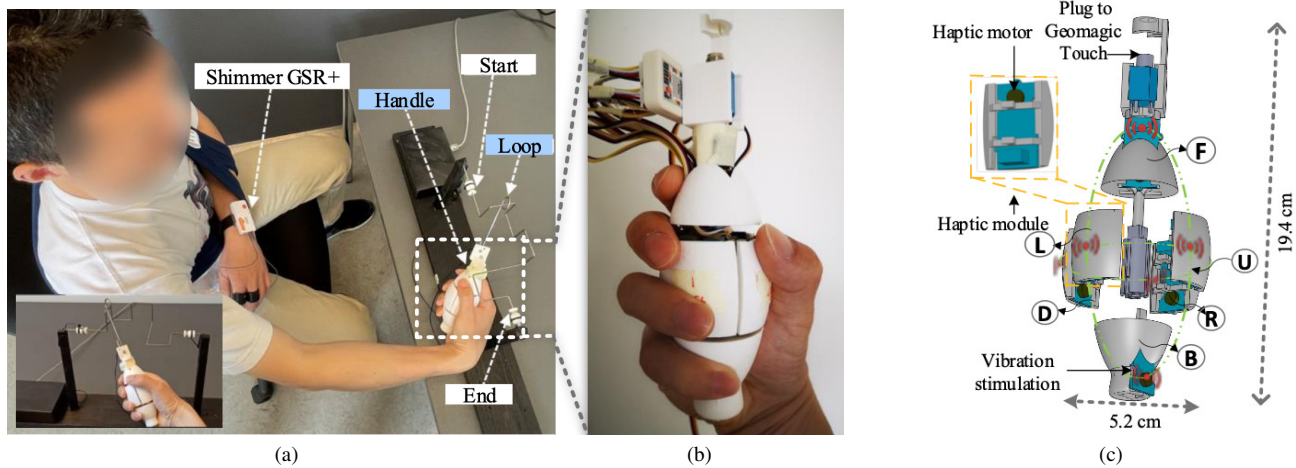


Fig. 1: Experimental setup (physical tasks). (a) Physical experimental environment, where the participant moves the loop across the real wire. The participant wears an Electrodermal Activity (EDA) sensor (Shimmer GSR+). The Heart Rate Variability sensor (Polar H10) is worn around the chest, in contact with the skin (not visible in the picture). (b) The custom handle is held by a participant. It houses six vibrotactile motors that provide distributed vibrations when a mistake happens, i.e., the wire touches the loop. It was attached to a real metallic loop during the physical tasks (as in (a)) and to a Geomagic kinesthetic interface during the VR tasks (as in Fig. 3). (c) CAD representation of the custom handle, highlighting the positioning of the vibrotactile actuators, L: left, U: up, R: right, D: down, F: front, B: back.

convey virtual contact sensations. That is commonly bundled in commercial VR controllers like the Oculus Touch and the Vive controller.

VR training literature has compared vibrotactile and kinesthetic feedback against each other and other modalities. For example, Islam and Lim, in their systematic review of vibrotactile feedback for motor skill training in VR [14], found that vibrotactile feedback used either alone or along with other feedback modalities was effective in most examples they analysed in VR literature. In the case of kinesthetic feedback an indicative study is Carlson et al. [15], where IVR training with kinesthetic haptic feedback was compared to the physical training of a burr puzzle assembly task. In the same work, it was found that although physical training led to better immediate outcomes in terms of task time completion in tests, the VR group showed better performance on a delayed post-test. In a between-subjects comparison of different haptic modalities (vibrotactile and kinesthetic) for powered tool simulation, Yin et al. [16] found that vibrotactile combined with audio feedback led to the greatest increase in performance, whereas the addition of kinesthetic feedback did not improve performance. Kreimeier et al. [7] performed a within-subjects study comparing the effect of vibrotactile feedback (plus visual), kinesthetic feedback (plus visual), and visual only (no haptics) on a throwing task in IVR, finding that kinesthetic feedback led to better task performance compared to the other modalities. Though there are no direct comparisons of kinesthetic vs. vibrotactile feedback for fine motor skill learning, indications from meta-analysis have indicated mostly large effect sizes from studies using kinesthetic [17] and cutaneous feedback [18]. These findings from the literature illustrate the need for further research into investigating the effectiveness of different haptic feedback modalities for motor

skill training.

B. Haptic feedback and physiological arousal

The term “arousal” refers to the increase in alertness and attention in response to external or mental stimuli. Subjective methods to measure arousal include questionnaires like the Self-Assessment Manikin (SAM) [19], and the Affective Slider (AS) [20]. Physiological measures of arousal are obtained by measuring signals from the autonomic nervous system (ANS), including EDA (electrodermal activity) caused by sweating in response to arousal, HRV (Heart Rate Variability), respiration volume/rate, pupil diameter variation, and brain activity [21]. EDA is a common choice for measuring ANS activation due to its neuroanatomical simplicity [22]. HRV is also popular as it enables the differentiation of various psychological and physiological states [22]. The relationship between arousal and task performance has been investigated in the literature. Though it has been hypothesized to be linked to performance in an inverted U-shaped curve according to Yerkes-Dodson law [23], this has not been conclusively established by literature due to other factors like task complexity and personality factors [24], [25]. Increases in arousal, for example, can affect memory and cause retrieval of task-irrelevant information, which may affect training outcomes [26]. Though, as discussed, both haptic feedback and arousal may affect task performance, the relationships between these three factors are not well established. There are few explorations linking arousal to haptic feedback in literature; for example, Gatti et al. [27] used a Geomagic Touch (3D Systems, US) to render viscous forces onto participants’ hands. At the same time, they viewed emotional pictures where they found an effect of haptics on subjective arousal (SAM) but not on physiological arousal (EDA, HRV, respiratory rate, and temperature). Akshita et

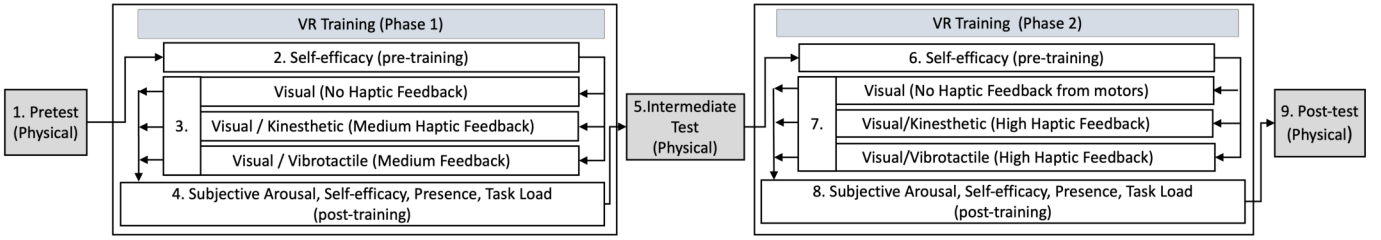


Fig. 2: Flow of the experiment. First, participants perform the buzzwire task in the physical environment (“pretest”), holding the handle attached to a metal loop and interacting with a real buzzwire. Then, participants are given training in an immersive VR environment (“VR Training (Phase 1)”). This time, participants hold the handle attached to a Geomagic Touch haptic interface and interact with a virtual buzzwire, receiving either visual feedback only, visual and kinesthetic feedback, or visual and vibrotactile feedback. After this VR training, participants are asked again to perform the buzzwire task in the physical environment (“intermediate test”). The experimental protocol is then repeated, with the participants again receiving training in an immersive VR environment (“VR Training (Phase 2)”), and then carrying out a final buzzwire physical task (“post-test”). The two phases of VR training (“Phase 1” and “Phase 2”) provide the participants with different intensities of feedback. This experimental organization enables us to study the effect of VR training in the learning of the considered fine motor skill task. To do so, before the training, we asked participants to estimate their perceived efficacy in carrying out the task (“pre-training”), and after the training, we asked participants to estimate their perceived task load, self-efficacy, subjective arousal, and presence (“post-training”). Other metrics, such as task completion time, contact time with the wire, and physiological arousal, are measured throughout the whole experimental session.

al. [28] similarly linked an emotional pictures dataset with vibrotactile haptics and subjective arousal (SAM), finding that high-intensity haptic feedback on the fingers contributes to increases in subjective arousal. The literature on haptic feedback and physiological arousal is not well established in the VR literature, except for Krogmeier et al. [29] for a non-motor skill training context and two pilot studies in the motor skill training domain [30], [31].

In addition to measures of performance and arousal, subjective measures of presence and task load, among others, add additional insight into designing more effective haptics-enabled VR training. There are examples from the literature of investigations into the relationship between presence and haptic feedback, where, for instance, Gibbs et al. [32]; Cooper et al. [33] found that multimodal (i.e., combinations of haptics with audio or visuals) feedback led to better presence than providing feedback in any modality alone. However, it is unclear which feedback modality, between kinesthetic and vibrotactile, can cause the greatest increase in presence. Additionally, research has pointed to the links between task load and haptic modality. For example, kinesthetic feedback has been linked to lower task load in surgical VR training [34] and VR motor skills therapy scenarios [35]. Weber et al. [36], in a desktop VR peg-in-hole experiment, found that kinesthetic feedback led to a lesser overall task load compared to vibrotactile and visual (no haptic feedback) conditions. It remains an open question if this pattern holds for the buzzwire task in immersive VR, which requires finer motor skill control. Self-efficacy is another subjective measure that has been linked to motor skill performance [37]. Though studies have found VR training to be linked to increases in self-efficacy [11], [38], its link to training outcomes is unclear.

III. METHODS

We designed an experiment to address the research questions presented at the end of Sec. I. As a representative

example of the fine motor skill task, we considered a buzzwire (or wire loop) task, where the aim is to move a metallic loop across a wire without touching it (see Fig. 1(a)). The flow of the experiment is shown in Fig. 2, from left to right. First, all participants are asked to perform the buzzwire task in a physical environment, called “pretest” in Fig. 2. During this task, participants hold a handle attached to a metal loop (see Fig. 1), which they move from one end of a physical wire to the other as fast as possible and with the least number of mistakes (see the details in Secs. III-A and III-B). Then, participants are given training in an immersive VR environment to improve their task performance, called “VR Training (Phase 1)”. The VR training environment consists of multiple virtual buzzwires, similar to the one in the pretest environment, but featuring different levels of difficulty (see Fig. 3 and Sec. III-C). This time, participants hold the handle attached to a Geomagic grounded haptic interface. Similarly, as before, they had to move the loop across the considered virtual buzzwire as fast as possible and with the least number of mistakes. During this VR training, participants receive different types of feedback about the contacts of the loop with the wire, according to the group they have been assigned to: one group of participants receiving visual feedback only, one receiving visual and kinesthetic feedback, and one receiving visual and vibrotactile feedback. After this VR training, all participants are asked again to perform the buzzwire task in the physical environment, called “intermediate test.” This post-training task enables us to compare how the VR training affected user performance according to the feedback provided. Finally, this experimental protocol is repeated, with the participants once again receiving training in an immersive VR environment, called “VR Training (Phase 2),” and subsequently analyzing the change in their performance in a final buzzwire physical task, called “post-test” (see the right-hand part of Fig. 2). The two phases of VR training (“Phase 1” and “Phase 2”) provide the participants with different intensities of feedback,

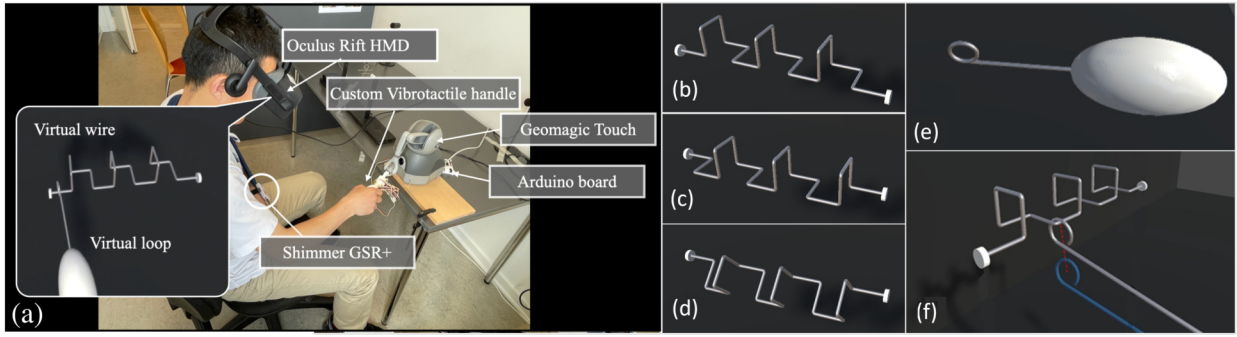


Fig. 3: Experimental setup (Virtual Reality training). (a) The Virtual Reality (VR) training environment, where participants hold the custom handle attached to the Geomagic Touch haptic device. (b) Level 1, (c) level 2, and (d) level 3 of the buzzwire task were presented during the two phases of the VR training. (e) Detail of the loop and handle avatars in VR, which have the same dimensions as their physical counterparts. (f) Visual feedback is provided to the users when the loop contacts the wire: a semi-transparent blue loop indicates the true position of the loop as commanded by the user, while the standard opaque grey loop indicates the proxy/ideal position of the loop inside the wire. Visual feedback is provided in all the feedback conditions.

as detailed in Sec. III-D, enabling us also to analyze whether the intensity of the feedback affects the user’s performance.

Performance of the task is measured using both objective metrics (completion time, mistakes, physiological arousal) and subjective measures (presence, task load, self-efficacy).

The following sections detail each of the aforementioned parts of our experiment.

A. Physical experimental environment

The physical environment is shown in Fig. 1. It is an adaptation of the physical test of motor skills used by Radhakrishnan et al. [11]. The metal wire is long 52 cm and rests on two 20 cm-high pillars. The start and end positions are denoted by grooves on two white plastic cylinders on either side of the wire, designed so that the loop can rest on them. An Arduino UNO is used to detect the mistake signals, i.e., when the loop touches the wire, and transmit them to an external computer through a serial connection. Similar circuits are used to detect when the loop is lifted off the groove at the beginning of the task as well as when the loop reaches the groove at the end of the wire. A video showing the physical setup can be found at <https://youtu.be/qZ6FBP-poAs?t=106> (also available in the supplementary video).

Participants were asked to carry out the task in this physical environment three times throughout the overall experience (“pretest,” “intermediate test,” and “post-test”), as described before and summarized in Fig. 2.

B. Custom vibrotactile handle

Figure 1 shows the custom handle designed to provide vibrotactile sensations about the (undesired) contacts between the loop and the wire, according to the feedback condition at hand. It is shaped as an ellipsoid and houses six vibrotactile modules, inspired from [39], [40]. The vibrotactile modules are positioned around the handle: four are placed symmetrically around the plane perpendicular to the main axis, and two are placed at the ends of the main axis (see Fig. 1).

Small gaps around where the haptic modules are positioned weaken the transmission of vibrations, making it easier to

recognize the source of the vibration. In this regard, we evaluated the vibration propagation throughout the handle by carrying out the following experiment. Each haptic module was sequentially activated at the maximum nominal power and vibration amplitude. Accelerometers were positioned onto all modules so as to measure the vibration propagation across different modules. Results show that the vibration amplitude was reduced by $67.19\% \pm 4.78\%$ (mean \pm std) from the vibrating module to the others.

To render the cue of contact between the rod ring and the virtual wire using vibrotactile feedback, we activate the corresponding vibrotactile motor. For example, if the contact happens in the superior part of the ring (Fig. 4, position U), the upper motor is activated. We carried out a perceptual experiment, enrolling 12 participants, to evaluate the capability of the handle to provide spatialized vibrotactile sensations. We activated one random motor at a time and asked the participants to indicate which one was vibrating (30 trials in total, 5 per motor). Participants were able to correctly recognize the activated motor 68.0% of the times (see Fig. 5). On the other hand, 13.6% of the times participants confused adjacent motors, while 18.4% they confused motors located further away.

C. VR training environment

The VR training environment is shown in Fig. 3. It was composed of three virtual buzzwires levels, shown in Figs. 3(b), 3(c), and 3(d). Level 1 was 57 cm long (end to end) with a horizontal span of 21 cm and eighteen 90° bends (see Fig. 3(b); level 2 was designed to be the mirrored version of level 1 so that the beginning and the end were inverted (see Fig. 3(c); and level 3 was designed to be the same as level 1 but tilted 45° around the main axis of the wire (Fig. 3(d)).

Participants were asked to wear a Oculus Rift Head Mounted Display (HMD) and hold the custom handle attached to a Geomagic Touch interface, as shown in Fig. 3(a). The latter is used to track the movement of the user to animate the virtual loop in all feedback conditions (see Figs. 3(e) and 3(f)).

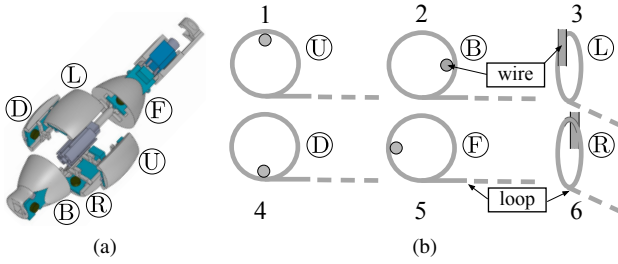


Fig. 4: Examples of vibrotactile motor activation in loop-wire contact. To render the vibrotactile stimuli generated by the contact between the ring and the wire (in the VR scenario) using the vibrotactile feedback (examples number 1-6 in (b)), we activate the corresponding vibrotactile motor. With reference to (b), when the superior internal part of the ring touches the wire (example 1 in (b)), the up-motor (U) in the handle is activated (as shown in (a)). Vice-versa, if the inferior internal part of the ring touches the wire (as in the example 4 in (b)), the motor down (D) of the handle is activated. The rest of the cases are represented by the examples number 2, 3, 5, and 6 in (b).

		Recognized pattern						Recognition Rate (%)
		L	U	R	D	F	B	
Rendered pattern	L	60	5	5	18.33	1.67	10	100%
	U	13.33	60	8.33	13.33	0	5	90%
	R	3.33	8.33	75	8.33	0	5	80%
	D	3.33	15	16.67	55	1.67	8.33	70%
	F	1.67	5	0	1.67	90	1.67	60%
	B	5	3.33	1.67	21.67	0	68.33	50%

Fig. 5: Custom vibrotactile handle: perceptual experiment. Confusion matrix showing the recognition rates when activating each of the six motors (see also Fig. 1).

D. Feedback conditions during VR training

Participants undergo two training phases in VR to become better at the buzzwire task. They are randomly assigned to one of the three VR training conditions: visual feedback only, visual and kinesthetic feedback, and visual and vibrotactile feedback about the contacts between the loop and the wire, that users are asked to minimize. Each subject carries out the VR training in only *one* feedback condition.

Across the two VR training phases (“Phase 1” and “Phase 2”, see Fig. 2) the intensity of the haptic feedback, kinesthetic and vibrotactile, changes. Haptic feedback provided during Phase 2 is 50% stronger than that provided during Phase 1 for kinesthetic and vibrotactile feedback, respectively. On the other hand, visual feedback does not change across the two VR training phases.

We carried out a short preliminary experiment to ensure that this difference in haptic intensity was noticeable. The change was indicated as “clearly noticeable” by the 12 participants

of the preliminary test, who were asked to carry out the VR buzzwire task four times, one per feedback intensity (medium, high) and type of haptic feedback (kinesthetic, vibrotactile), and describe their experience. The indications and results of this preliminary study have been used to set the parameters of the provided feedback, i.e., the stiffness/damping coefficients for the kinesthetic feedback, the vibration amplitude for the vibrotactile feedback.

Below we detail the three types of feedback conditions.

1) Visual feedback only

In this feedback condition, whenever the loop touches the wire, the virtual representation of the loop doubles, as shown in Fig. 3(f) and at <https://youtu.be/qZ6fBP-poAs?t=65> (also available in the supplementary video). A semi-transparent blue loop indicates the true position of the loop as commanded by the user, while the standard opaque grey loop indicates the proxy/ideal position of the loop, still inside the wire. It is worth noticing that here both the motors in the custom vibrotactile handle and the kinesthetic force feedback provided by the Geomagic Touch actuators are not active, while the user still holds the same handle attached to the kinesthetic interface to perform the task. During the contact situation, the opacity of the semi-transparent blue loop increases as it moves away from the proxy position of the loop inside the wire. Finally, a dotted line indicates the direction where the user should move to rejoin the wire and continue the task; the color of the line changes from black to red as the user moves the loop away from its proxy position inside the wire.

2) Visual and kinesthetic haptic feedback

In this case, whenever the loop touches the wire, participants receive visual feedback as well as kinesthetic feedback forces provided by the Geomagic Touch interface. Specifically, the loop-wire haptic interaction was rendered using a simple elastic model with stiffness 56.7 N/m and 85 N/m for the two phases of the VR training, respectively. The linear damping was kept constant in both phases at 8 Ns/m. These values were chosen following the indications of the preliminary study, so as to resemble as much as possible the interaction with the physical buzzwire.

3) Visual and vibrotactile haptic feedback

Again, whenever the loop touches the wire, participants receive visual feedback as well as vibrotactile feedback stimuli provided by the custom handle. Specifically, vibrations were provided along the direction where the contact between the virtual loop and wire happened, e.g., if the loop touched the wire in its upper sector (as in Fig. 3(f)), the vibration was provided by the motor U (up) (see Fig. 1(c)); if the loop touched the wire in its lower sector, motor D (down) would be activated. Only one motor is activated at a time, based on the directional component of the contact that has the highest magnitude. The vibration amplitude was fixed and set to 0.33 g for phase 1 and 0.49 g for phase 2 of VR training. These values were chosen again following the indications of the preliminary study reported in section III-D, so as to resemble as much as possible the interaction with the physical buzzwire.

E. Metrics

1) Objective performance metrics

In their review of haptic feedback for motor skill training, Basalp et al. [41] listed three types of metrics that can be used

TABLE I: Relationship between increases in physiological arousal and EDA and ECG metrics.

Effect of Arousal (↑: increase, ↓: decrease)	
EDA	Skin Conductance (SC) ↑
	Skin Conductance Response Amplitude (SCRamp) ↑
	Skin Conductance Response Peaks Rate (SCRPeaks) ↑
ECG	Heart Rate (HR) ↑
	Inter-Beat Interval (IBI) ↓
	Root Mean Square of Successive Differences (RMSSD) ↓
	Standard Deviation of NN Intervals (SDNN) ↓
	Normalized High-Frequency Component (HFN) ↓
	LF/HF (Low Frequency/High Frequency) Ratio ↑

to objectively measure performance, i.e., spatial, temporal, and spatiotemporal metrics. We considered three metrics:

- task completion time (TCT), which is the time taken to move the loop from start to end.
- loop-wire contact time (CT), which is the total time the loop is in contact with the wire.
- the performance change score (PCS), which is calculated by rank ordering changes in performance among all participants into ten equal quantiles for TCT and CT separately [41] (see Appendix, Section 2, for the quantile ranges). We first convert the change in the performance metrics, TCT and CT, into scores ranging from 1 to 10 (with 1 indicating the least change in performance and 10 indicating the highest) by dividing them into 10 evenly-distributed groups for all participants. The PCS for each participant is then determined by adding these two scores together, e.g., a participant who has improved the most in both TCT (rank order = 10) and CT (rank order = 10) would get a PCS of 20.

2) Subjective measures

We considered subjective metrics related to perceived task load, self-efficacy, subjective arousal and presence. To measure task load, we used NASA-TLX [42], across the six standard dimensions - Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. A combined task load score was then taken by averaging all six dimensions. To measure self-efficacy, the participants are asked the question “How confident are you that you can perform a similar task effectively (go from start to finish as fast as you can with minimal mistakes) on a scale from 1 to 7?”, both before and after the training (see also Fig. 2). To measure subjective arousal, we combined the arousal sub-scale of Self-Assessment Manikin (SAM) [19] with the Affective Slider [20], as seen in Granato et al. [43]. To measure presence, we used the five dimensions in the physical subscale of the Multimodal Presence Scale [44].

3) Physiological arousal metrics

To investigate the participants’ arousal levels, heart rate variability, and electrodermal activity, we used the Polar H10 (Polar Electro Oy, Finland) and Shimmer GSR+ (Shimmer Research Ltd., Ireland) sensors. Baseline values were subtracted where applicable from the physiological arousal metrics to control for individual physiological differences [45].

All physiological sensor data were streamed and stored using the iMotions platform (iMotions A/S, Denmark).

a) Heart Rate Variability

The Polar H10 is an Electrocardiogram (ECG) heart rate monitor often used in VR studies [11]. The sensor is worn around the chest with direct contact with the skin. Data in the form of heart rate and Inter-beat Intervals (R-R intervals) are transmitted via Bluetooth at a rate between 1 and 2 Hz to the PC and then recorded by the iMotions application. Based on the raw R-R interval data, different measures of heart rate variability, including time and frequency domain metrics, were calculated using the hrv-analysis Python library¹. Table I shows the link between increases in arousal and its effect on ECG based metrics. Increases in arousal are indicated by increases in Heart Rate and Low Frequency/High Frequency (LF/HF) Ratio [46]. On the other hand, decreases in HRV measures like IBI (Inter-Beat Interval), SDNN (Standard Deviation of NN Intervals), RMSSD (Root Mean Square of Successive Difference), and the frequency domain measure HFN (Normalized High-Frequency Component), indicate an increase in arousal [47].

b) Electrodermal activity (EDA)

The Shimmer GSR+ (Galvanic Skin Resistance) measures EDA by passing a small current through electrodes placed on the index and middle fingers in the left hand [11], to allow the participants to use their right hand for carrying out the buzzwire task. The sampling rate was 128 Hz. EDA measures used in this study include SC (Skin Conductance), which increases in response to an increase in arousal [48]. Table I also shows the link between increases in arousal and its effect on EDA metrics. An increase in arousal also leads to a higher rate of Skin Conductance Response Peaks per minute (SCRPeaks) which are peaks in the SC amplitude lasting between 1-5 seconds after onset [29]. Similarly, the mean peak amplitude of all SCR peaks (SCRamp) is also a positive measure of arousal [29]. All EDA signals were processed using the Neurokit2 Python library [49].

F. Experimental procedure

Participants signed up for the study using the University’s online participant recruitment system and were randomly assigned to one of the three feedback conditions. They had to satisfy these criteria: (a) right-handed (to minimize variation in the setup), (b) normal vision or corrected-to-normal vision with contact lenses (no glasses), and c) no mental illnesses or sensitivity to nausea. Participants signed up for 35-minutes time slots of their choosing and were paid the equivalent of 12 EUR. Approval for this experiment was obtained from the Cognition and Behavior Lab’s Human Subjects Committee (approval code: 339), Aarhus University.

After reading and signing the consent forms, participants were familiarized with the experimental task by an experimenter, who demonstrated the task. Thereafter, participants were given privacy to place the Polar H10 around their chest. Afterwards, the Shimmer GSR electrodes were placed on the index and middle fingers of the participant’s left hand. Finally, participants were assisted in wearing a sling around the neck

¹<https://github.com/Aura-healthcare/hrv-analysis>

so that they could rest their left hand with the fingers relaxed, as seen in Fig. 1(a). Participants were asked to keep their left hand still, so as to minimize the noise in the recorded signals. The signal quality for both sensors was verified before the experiment started. Baseline data from these biosensors were measured with the participants seated quietly, with their eyes closed, breathing normally, and without wearing the HMD.

The overall flow of the experiment is summarized in Fig. 2 and has been described at the beginning of Sec. III. 73 participants enrolled in the study, randomly divided between visual/kinesthetic (n=26), visual/vibrotactile (n=22), and visual only (n=25) feedback conditions. 37 participants identified themselves as female, 35 as male, and 1 as other. 36 participants indicated their age group in the 25-34 range, 33 in the range 18-24, and 4 in the range 35-44. More than half of the participants in the study (n=49) indicated that they had tried VR using a head-mounted display more than once, while 24 reported never having used it before the experiment. An analysis of the quality of the HRV and EDA data using iMotions led to discarding of HRV and EDA data from 3 and 20 participants, respectively, due to signal quality issues. The data loss in the heart rate variability metrics (HR, IBI, RMSSD, SDNN, HFN, LF/HF Ratio) was almost equally distributed between the conditions, as can be observed in Section 3 of the Appendix, with losses ranging from 0 to 2 per metric. However, with regards to EDA (electro-dermal activity) metrics (SC, SCRamp, SCRPeaks), the losses are unevenly distributed, with minimal losses in the visual/vibrotactile condition and higher losses in the visual/kinesthetic and visual only conditions.

In the process of analysing the gathered data, we employed different statistical techniques. Firstly, two-way mixed ANOVA tests were conducted to investigate the main effects of the between-subjects factor – feedback condition (visual, visual/kinesthetic, visual/vibrotactile) – and the within-subjects factor – time (pretest, intermediate test, post-test) – on performance metrics (task completion time, contact time, performance change score) and subjective measures (self-efficacy, task load, subjective arousal and presence). Furthermore, Mauchly’s test of sphericity and associated Greenhouse-Geisser corrections were applied where required. To discern the influence of feedback conditions on individual variables such as the NASA-TLX dimensions, SAM (Self-Assessment Manikin) ratings, and specific EDA and HRV metrics, one-way ANOVA or Kruskal-Wallis tests were utilised.

IV. RESULTS

A. Objective performance

We analyzed the task completion time (TCT), contact time (CT), and performance change scores (PCS), as described in Sec. III-E1. Changes in performance metrics due to training are evaluated as changes

- from the pretest to the post-test physical tasks (referred to as “overall training”);
- from the pretest to the intermediate test physical tasks (referred to as “Phase 1 training”);
- from the intermediate test to the post-test physical tasks (referred to as “Phase 2 training”).

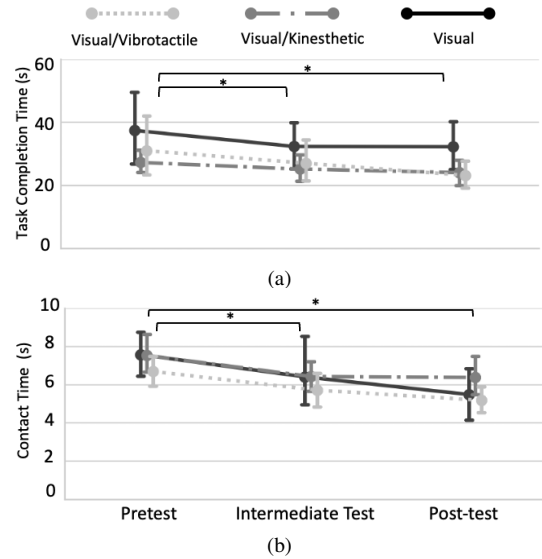


Fig. 6: Mean performance metrics with error bars (CI : 95%) across the three physical buzzwire tests for Visual/Vibrotactile, Visual/Kinesthetic, and Visual feedback conditions for (a) task completion time (TCT) and (b) contact time (CT). * denotes significant difference at $\alpha=0.05$.

a) Task completion time (TCT)

Fig. 6(a) shows the task completion time across the three physical buzzwire tests. Table II summarizes the results of the analysis. Mauchly’s test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2)=30.089$, $p<.001$. Therefore, a Greenhouse-Geisser correction was applied ($\epsilon=0.737$). There was no statistically significant interaction between the feedback condition and test iteration on TCT. The main effect of time showed a statistically significant decrease in TCT over the three test iterations, with TCT decreasing from 31.53 ± 21.57 s (mean \pm one standard deviation) in the pretest, to 27.98 ± 15.64 s in the intermediate test and to 26.2 ± 13.97 s in the post-test. Post-hoc analysis with a Bonferroni adjustment revealed that TCT decreased significantly from pretest to intermediate test (3.63 s, $p=.024$), and from pretest to post-test (5.34 s, $p=.01$) but not from intermediate test to post-test (1.7 s, $p=.304$). The main effect of group showed that there was no statistically significant difference in task completion time between feedback conditions.

b) Contact time (CT)

Fig. 6(b) shows the task completion time across the three physical buzzwire tests. Table II summarizes the results of the analysis. Mauchly’s test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2)=4.754$, $p=.093$. There was no statistically significant interaction between the feedback condition and test on CT. The main effect of time showed a statistically significant difference in CT over the test iterations, with CT decreasing from 7.25 ± 2.49 s (mean \pm one standard deviation) during the pretest, to 6.18 ± 2.99 s in the intermediate test, and to 5.68 ± 2.58 s in the post-test. Post-hoc analysis with a Bonferroni adjustment revealed that CT decreased significantly from pretest to intermediate test (1.07 s, $p=.002$), and from pretest to post-test (1.61 s, $p<.001$) but not from intermediate

TABLE II: Summary of Two-way Mixed ANOVA Results for Task Completion Time and Contact time by Feedback Condition (between-subjects) and Test Iteration (within-subjects)

Dependent variable	Source of variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-value	p-value
Task Completion Time	Feedback condition	2883.949	2	1441.974	1.930	.153
	Test iteration	1060.274	1.473	719.692	7.582	.003*
	Feedback condition * Test iteration	181.935	2.946	61.747	.650	.582
Contact Time	Feedback condition	32.951	2	16.476	.992	.376
	Test iteration	95.662	2	47.831	18.113	<.001*
	Feedback condition * Test iteration	7.940	4	1.985	.752	.559

* denotes significant difference at $\alpha=0.05$

test to post-test (0.53 s, $p=.153$). The main effect of feedback condition showed that there was no statistically significant difference in contact time between the feedback conditions.

c) Performance Change Score (PCS)

The mean PCS among the three conditions was calculated to be 11.19 for visual/vibrotactile, 10.81 for visual/kinesthetic, and 11.04 for the visual condition. There was no statistically significant difference between the PCS for the three conditions ($F(2,69)=0.047$, $p=0.95$).

B. Self-Efficacy, Task load, Presence

There was no statistically significant interaction between the feedback condition and time on self-efficacy, $F(4.729,165.52)=1.143$, $p=.339$, partial $\eta^2=.032$. The main effect of time did not show a statistically significant difference in mean SE at the different time points, $F(2.365,165.52)=1.549$, $p<.0005$, partial $\eta^2=.022$. The main effect of feedback condition showed that there was no statistically significant difference in mean SE between feedback conditions $F(2,70)=1.135$, $p=.327$, partial $\eta^2=.031$.

Fig. 7(a) shows the overall NASA-TLX score obtained by calculating the average of the six NASA-TLX task load dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration) reported by the participants. There were no significant differences between the feedback conditions in terms of individual NASA-TLX dimensions.

Fig. 7(b) shows the overall presence score reported by the participants, calculated by taking the average of the responses to the five questions in the presence questionnaire. There were no significant differences in the presence score between the feedback conditions ($F(2,70)=1.5$, $p=.22$).

C. Physiological arousal

a) Arousal and feedback condition

There was no effect of feedback condition on subjective arousal (Self-Assessment Manikin) during training ($H(2)=0.47$, $p=.79$). No significant effect of feedback condition was found on each of the EDA and HRV metrics. To measure differences in arousal immediately after the participants commit a mistake during VR training, the physiological arousal levels were averaged for 10-second windows starting at the moment of contact between the loop at the wire. No significant differences were found between the three feedback conditions in terms of these ‘‘immediate’’ arousal metrics.

TABLE III: Physiological Arousal Metrics across High ($n=15$) and Low Performance Change groups ($n=16$).

Arousal Metric	High Change (Mean \pm SD)	Low Change (Mean \pm SD)	p-value	
HRV	HR	4.82 \pm 5.59	-1.16 \pm 3.28	0.002*
	IBI	-42.27 \pm 55.81	4.7 \pm 34.69	0.02*
	RMSSD	-58.41 \pm 101.15	17.68 \pm 35.54	0.001*
	SDNN	-47.91 \pm 63.07	4.08 \pm 32.14	0.003*
	LF/HF Ratio	-1.47 \pm 3.97	-1.44 \pm 4.97	0.98
	HFN	1.33 \pm 17.83	7.41 \pm 14.89	0.35
EDA	SC	2.35 \pm 3.18	1.84 \pm 1.81	1.0
	SCRamp	0.14 \pm 0.22	0.14 \pm 0.16	0.99
	SCRPeaks	10.37 \pm 4.76	8.84 \pm 1.84	0.36

* denotes significant difference at $\alpha=0.05$

b) Arousal and task performance

The link between task performance and arousal was assessed using co-relation tests between performance change score (PCS) and arousal, as measured by both subjective (SAM) and physiological (EDA, HRV) measures. Statistically significant (but weak) negative correlations were observed between PCS and three HRV metrics: RMSSD ($\rho=-0.35$, $p=.003$), SDNN ($\rho=-0.32$, $p=.008$) and HR ($\rho=0.245$, $p=.043$). As decreases in RMSSD and SDNN, and increases in HR, are correlated with increases in arousal, the results of the correlations tests show that changes in performance are correlated with increases in arousal. To further investigate this point, we grouped participants into two groups: high-performance ($n=15$) and low-performance ($n=15$), respectively. Following

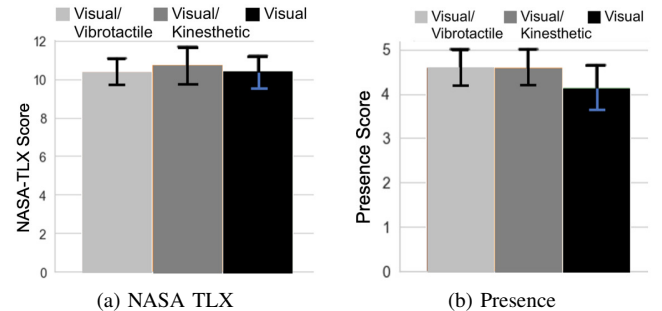


Fig. 7: (a) NASA TLX and (b) Presence scores across Visual/Vibrotactile, Visual/Kinesthetic, and Visual conditions. Mean and 95% confidence interval are reported.

the technique described by Radhakrishnan et al. [11], the high-performance group was defined as participants who had a PCS above the upper range of the inter-quartile range (i.e., above the 75% mark) and the low-change group had participants from the lower quartile range (i.e., below the 25% mark). As shown in Table III, Mann-Whitney U tests were performed for HRV and EDA metrics between these two groups. It can be observed from the table that participants in the high-performance group had a higher baseline adjusted heart rate (HR) compared to the low-performance group. It can also be observed that participants in the high-performance group had a lower baseline-adjusted inter-beat interval (IBI), Root Squared Mean of Successive Differences (RMSSD), Standard Deviation of NN intervals (SDNN). This result supports the patterns observed in the correlation tests, i.e., changes in performance are correlated with increases in arousal.

To further investigate the trends in the relationship between arousal and performance, participants were also divided into high-arousal and low-arousal groups regarding, separately, the HRV and EDA metrics. This split followed the same principle we used for splitting the participants into high and low-performance groups, i.e., according to their position in the inter-quartile range. Subsequently, the PCS were compared for each of these pairs of high and low-arousal groups using Mann-Whitney U tests. In terms of high and low-arousal groups split according to RMSSD, participants in the high-arousal group had an IS of 9.06, which was lower than the IS of 12.06 observed in the low-arousal group ($p=.01$, $U=66$). Similarly, for groups split according to SDNN, participants in the high-arousal group were observed to have an IS of 9.06, which was lower than the IS of 12.31 observed in the low-arousal group ($p=.017$, $U=70$). Among groups split according to HR, participants in the high-arousal group had an HR of 12.39 as compared to the HR of 9.78 observed in the low-arousal group, and this difference approached the threshold of statistical significance ($p=.09$, $U=221$). This further confirms the trends observed earlier of performance changes as arousal increases. To investigate potential interactions between arousal and NASA-TLX task load measures, correlation tests were performed between the arousal metrics on one side and task load measures on the other (frustration, mental load, physical load, temporal load, effort, performance, NASA-TLX score). A statistically significant and moderate positive correlation was found between SCRamp and temporal demand ($\rho=0.323$, $p=.018$), i.e., as temporal demand increased the amplitude of skin conductance responses increased.

V. DISCUSSION

Here we discuss the results from Sec. IV in the context of the three research questions introduced at the end of Sec. I.

A. Can vibrotactile or kinesthetic feedback influence VR training performance in a buzzwire motor skill task?

Reductions in task completion time were seen after overall training (pretest to post-test) as well as Phase 1 training (pretest to intermediate test) for all participants. Similarly, concerning contact time, participants in all feedback conditions showed reductions after overall training (pretest to post-test).

The absence of significant changes in both task completion and contact time after Phase 2 training (intermediate test to post-test) is possibly due to the participants reaching a "performance plateau" [50]. Furthermore, statistical tests comparing the performance change scores across the three feedback conditions were inconclusive in finding any difference between them. This implies that, while IVR training with vibrotactile and kinesthetic feedback improves performance, it is not clear which haptic modality is better or if they are better than visual feedback only. However, analysis described in Sec. V-B points to the potential for further investigation on the use of haptic feedback to modulate arousal and thereby improve performance. Furthermore, alternate uses of haptic feedback could be explored in the context of training in the buzzwire task, for example in the form of virtual fixtures or guidance [51]–[54].

B. Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in arousal levels during training? Is there a link between physiological arousal during training and changes in performance afterward?

Analysis of the data revealed no effect of feedback condition on EDA and HRV. Analysis of SAM responses did not reveal any effect of feedback on subjective arousal. Correlation tests revealed a weak negative correlation between changes in performance and two HRV metrics (RMSSD, SDNN), i.e., as arousal increased (decrease in RMSSD and SDNN is correlated with increases in arousal) during VR training, so did performance. This link was further supported by statistical tests which showed higher arousal levels indicated by four HRV metrics (HR, IBI, RMSSD, and SDNN) in the high-change group as compared to the arousal levels of participants in the low-change group. This trend was again confirmed by statistical tests comparing performance between groups of participants split according to the degree of arousal, which showed that participants in high-arousal groups (defined by the HRV metrics RMSSD and SDNN) demonstrated greater performance change scores compared to those in low-arousal groups. However, the question of whether kinesthetic or vibrotactile feedback can cause variations in arousal is still unresolved and future research might explore this question further by, e.g., investigating novel haptic modalities or research designs which can elicit such variations. Even though our study points to higher levels of arousal linked to positive changes in performance, we also acknowledge the view that increasing arousal or stress may have detrimental effects on performance. For example, in a similar buzzwire training scenario, Radhakrishnan et al. [11] found that participants across VR and physical training conditions who had the highest arousal levels during training showed the lowest changes in performance (and vice-versa). Though the motor task is similar (a buzzwire scenario), that study had participants undergo training in a different setup, i.e., physical training and VR training with simple non-directional vibrotactile feedback from the Oculus Touch controller. There are also non-VR studies that point out that high-arousal levels may lead to lower performance [55], [56]. Wu et al. [57] in their VR driving

scenario found that moderate arousal levels (neither too high nor too low) correlated with the best performance. Perhaps, such contradictions may be solved if variables in addition to performance and arousal are considered. For example, a high-arousal level may be linked to higher enjoyment or motivation, which in turn affects performance positively, but similar arousal levels due to anxiety may negatively affect performance. Another possible line of inquiry opened by this study is the indication of an increase in the amplitude of skin conductance responses accompanied by a simultaneous and moderate increase in temporal demand as measured by NASA-TLX. As temporal demand is negatively correlated with performance in prior research [58], future research might explore the effectiveness of using cost-effective EDA and ECG sensors to measure temporal demand and other factors contributing to cognitive/task load and explore their use in improving IVR training.

C. Can motor skill training in VR with different haptic feedback (kinesthetic vs. vibrotactile) cause variations in reported presence, task load, and self-efficacy?

The analysis showed no effect of feedback condition on overall presence. This might be caused by the subjective experience of presence being dominated in this study by the visual aspect which was the same across the three feedback conditions. As Grassini et al. [59] had linked increased presence to better training outcomes, the lack of an effect of feedback condition on changes in performance may thus be partially explained by the lack of differences in terms of presence. However, there are inspirations from literature for further investigation into improving presence and thereby training outcomes. For example, some studies have found vibrotactile feedback coupled with visual feedback to result in better-reported presence compared to visual alone [32] or kinesthetic feedback alone [7].

The analysis also revealed no effect of feedback conditions on the overall task load. This is important as increases in workload have been linked to decreased motor skill performance as described by Yurko et al. [60] in their study on simulator-based laparoscopy training. In light of this, the lack of differences in performance metrics between the feedback conditions in our study may be linked to the lack of differences in perceived task load. In a study using a similar buzzwire test setup, Radhakrishnan et al. [11] found participants who received physical training to report more temporal demand than those who underwent IVR training (with non-directional vibrotactile feedback), and similar to our study, overall task load and performance changes were indistinguishable between the training conditions. On the other hand, Weber et al. [36], in a peg-in-hole task using desktop VR, found that vibrotactile feedback resulted in a higher task load compared to kinesthetic feedback, and the authors attributed this to "feedback ambiguity" arising from the design of the vibrotactile feedback. Therefore, it is possible that the lack of differences in overall task load between the kinesthetic and vibrotactile feedback conditions in our study may arise from the lack of directional ambiguity as shown in the perceptual experiment described in Sec. III-B.

There were no statistically significant changes in self-efficacy during VR training for all the feedback conditions. This is surprising, as prior literature points to the links between self-efficacy and learning/training performance [6], [37]. Our study has shown that there are changes in performance across all the feedback conditions, in spite of this lack of increase in self-efficacy. Radhakrishnan et al. [11] presents a contrasting example with a similar buzzwire motor skill task, where it was found that significant changes in self-efficacy during VR training was accompanied by corresponding changes in performance. Stevens et al. [61] in a motor skill training task showed that increased task-difficulty level is one factor which leads to impairments in both self-efficacy and performance. Therefore future research could consider methods to improve self-efficacy, for example by adapting the difficulty level of training.

D. Summary

In summary, VR training is effective in improving performance regardless of the considered haptic feedback modality. So one may wonder if kinesthetic or vibrotactile feedback were to be used, are they interchangeable, or should additional parameters be considered? Our analysis of the data showed that there is a correlation between higher arousal levels and higher performance across feedback conditions. These findings should encourage further research in using different types of haptic feedback to potentially affect arousal levels and therefore performance. Future research could explore variations of kinesthetic feedback, for example with wearable exoskeletons, which might also combine aspects of cutaneous feedback to provide a good trade-off between cost and performance [62].

VI. LIMITATIONS

Though the current study confirms previous findings [11], where participants in a VR training condition showed positive changes in performance indistinguishable from those who underwent physical training. However, it cannot be concluded that there was a long term effect from VR training due to the lack of retention tests in this study, as short-term changes in performance can either lead to better or worse learning outcomes over time [63].

Arousal in the present study may be linked to multiple factors including skill level, task complexity, visual feedback, the novelty of VR, and haptic rendering and hardware. Pre-existing skill levels and arousal levels were controlled by using a baseline phase and novelty effects were controlled by a tutorial phase. However, there is scope for further improvement to obtain a more fine-grained link between haptic feedback and physiological arousal. Future studies investigating the effect of haptic feedback on arousal may try to incorporate a baseline haptic feedback condition where the participants receive haptic feedback in regular intervals as seen in Krogmeier et al. [29]. The corresponding physiological arousal levels can then be measured, for a baseline level of correlation between haptic feedback and arousal to be established. Another shortcoming of the present study was the loss of arousal data from the Shimmer GSR+ sensor, which might have led to potential differences between the conditions *not* being detected. Future

studies could consider either not using such sensors due to their sensitivity to movement or using them on locations on the body like the feet or the shoulders, which are less convenient but more reliable.

The Geomagic Touch’s workspace is another limiting factor in its use in immersive VR training scenarios. Future studies may explore either wearable haptic devices [18], [62], [64]–[66] or grounded kinesthetic devices with larger workspaces. It is also worth observing that the two haptic feedback conditions have some shared properties. For example, in the kinesthetic condition, there is an element of cutaneous sensation felt while the participants hold the handle and when they receive feedback. Similarly, in the vibrotactile condition, participants feel some resistance due to the inertia inherent in the Geomagic Touch. However, since these properties are shared across the feedback conditions, they have been controlled in this experiment.

Another potential limitation of this study is the between-subjects methodology followed. Though it is better suited for measuring learning effects, this might partly explain the lack of significant findings in subjective metrics. None of the participants in any of the feedback condition could experience any of the other feedbacks available, therefore subjective measures may not reveal the actual effect of the haptic modality on these subjective metrics. One solution is to have a larger sample size in between-subjects studies, which can reveal smaller effect sizes. On the other hand, Richard et al. [67], in a simulation study comparing within and between-subject methods for evaluating embodiment, concluded that the within-subjects method is more sensitive in detecting small effect sizes while having smaller sample sizes. However, as within-subjects methodology may potentially cause recency effects [68] affecting the analysis of objective performance measures, future studies might control for this aspect by having longitudinal cross-over studies.

VII. CONCLUSION

In this study, a fine motor skill training buzzwire task in immersive VR was used to investigate the effect of haptic feedback and physiological arousal on performance. The experiment revealed the effectiveness of motor skill training in VR regardless of the haptic feedback. The investigation also found links between arousal and performance, with increases in arousal being accompanied by increases in performance. Further research is required to investigate whether the inclusion of haptic feedback holds potential for other motor skill training scenarios in VR, as well as the trade-offs between kinesthetic feedback and varieties of cutaneous feedback, such as the vibrotactile feedback considered here. With regards to investigating the relationship between haptic feedback and arousal, additional physiological metrics can be considered, such as EEG (electroencephalography), EMG (electromyography), and pupil dilation, particularly since biosensors to measure these signals are being increasingly integrated into commercial VR head-mounted displays.

ACKNOWLEDGMENTS

The authors want to thank the Cognition and Behavior Lab, Aarhus University, and iMotions A/S for facilitating the

experiment.

REFERENCES

- [1] J. Abich, J. Parker, J. S. Murphy, and M. Eudy, “A review of the evidence for training effectiveness with virtual reality technology,” *Virtual Reality*, vol. 25, no. 4, pp. 919–933, 2021.
- [2] U. Radhakrishnan, K. Koumaditis, and F. Chinello, “A systematic review of immersive virtual reality for industrial skills training,” *Behaviour & Information Technology*, vol. 40, no. 12, pp. 1310–1339, 2021.
- [3] R. Khan, J. Plahouras, B. C. Johnston, M. A. Scaffidi, S. C. Grover, and C. M. Walsh, “Virtual reality simulation training in endoscopy: a cochrane review and meta-analysis,” *Endoscopy*, vol. 51, no. 07, pp. 653–664, 2019.
- [4] D. B. Mekbib, J. Han, L. Zhang, S. Fang, H. Jiang, J. Zhu, A. W. Roe, and D. Xu, “Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials,” *Brain injury*, vol. 34, no. 4, pp. 456–465, 2020.
- [5] M. Slater, “Immersion and the illusion of presence in virtual reality,” *British Journal of Psychology*, vol. 109, no. 3, pp. 431–433, 2018.
- [6] G. Makransky and G. B. Petersen, “The cognitive affective model of learning in immersive virtual reality,” *Educational Psychology Review*, vol. 33, no. 3, pp. 937–958, 2021.
- [7] J. Kreimeier, S. Hammer, D. Friedmann, P. Karg, C. Bühner, L. Bankel, and T. Götzelmann, “Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality,” in *Proc. 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2019, pp. 289–298.
- [8] T. Howard, M. Marchal, A. Lécuycy, and C. Pacchierotti, “Pumah: Pantilt ultrasound mid-air haptics for larger interaction workspace in virtual reality,” *IEEE Trans. Haptics*, vol. 13, no. 1, pp. 38–44, 2019.
- [9] L. Meli, S. Scheggi, C. Pacchierotti, and D. Prattichizzo, “Wearable haptics and hand tracking via an rgb-d camera for immersive tactile experiences,” in *Proc. ACM SIGGRAPH Posters*, 2014.
- [10] C. G. Christou, D. Michael-Grigoriou, D. Sokratous, and M. Tsiakoulia, “Buzzwirevr: An immersive game to supplement fine-motor movement therapy,” in *ICAT-EGVE*, 2018, pp. 149–156.
- [11] U. Radhakrishnan, F. Chinello, and K. Koumaditis, “Investigating the effectiveness of immersive vr skill training and its link to physiological arousal,” *Virtual Reality*, pp. 1–25, 2022.
- [12] E. M. Overtoom, T. Horeman, F.-W. Jansen, J. Dankelman, and H. W. Schreuder, “Haptic feedback, force feedback, and force-sensing in simulation training for laparoscopy: A systematic overview,” *Journal of surgical education*, vol. 76, no. 1, pp. 242–261, 2019.
- [13] K. Rangarajan, H. Davis, and P. H. Pucher, “Systematic review of virtual haptics in surgical simulation: a valid educational tool?” *Journal of surgical education*, vol. 77, no. 2, pp. 337–347, 2020.
- [14] M. S. Islam and S. Lim, “Vibrotactile feedback in virtual motor learning: A systematic review,” *Applied Ergonomics*, vol. 101, 2022.
- [15] P. Carlson, A. Peters, S. B. Gilbert, J. M. Vance, and A. Luse, “Virtual training: Learning transfer of assembly tasks,” *IEEE Trans. Visualization and Computer Graphics*, vol. 21, no. 6, pp. 770–782, 2015.
- [16] G. Yin, M. J.-D. Otis, P. E. Fortin, and J. R. Cooperstock, “Evaluating multimodal feedback for assembly tasks in a virtual environment,” *Proc. ACM on Human-Computer Interaction*, vol. 3, no. EICS, pp. 1–11, 2019.
- [17] B. Weber and C. Eichberger, “The benefits of haptic feedback in telesurgery and other teleoperation systems: a meta-analysis,” in *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being: 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III 9*. Springer, 2015, pp. 394–405.
- [18] P. Kourtesis, F. Argelaguet, S. Vizcay, M. Marchal, and C. Pacchierotti, “Electrotactile feedback applications for hand and arm interactions: A systematic review, meta-analysis, and future directions,” *IEEE Trans. Haptics*, 2022.
- [19] M. M. Bradley and P. J. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [20] A. Betella and P. F. Verschure, “The affective slider: A digital self-assessment scale for the measurement of human emotions,” *PLoS one*, vol. 11, no. 2, 2016.
- [21] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of psychophysiology*. Cambridge university press, 2007.
- [22] M. E. Dawson, A. M. Schell, and D. L. Filion, *The Electrodermal System*, 4th ed., ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2016, p. 217–243.

- [23] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology*, vol. 18, no. 5, pp. 459–482, 1908.
- [24] J. Storbeck and G. L. Clore, "Affective arousal as information: How affective arousal influences judgments, learning, and memory," *Social and personality psychology compass*, vol. 2, no. 5, pp. 1824–1843, 2008.
- [25] J. A. Bargh and J. L. Cohen, "Mediating factors in the arousal-performance relationship," *Motivation and Emotion*, vol. 2, no. 3, pp. 243–257, 1978.
- [26] J. L. Plass and S. Kalyuga, "Four ways of considering emotion in cognitive load theory," *Educational Psychology Review*, vol. 31, no. 2, pp. 339–359, 2019.
- [27] E. Gatti, G. Caruso, M. Bordegoni, and C. Spence, "Can the feel of the haptic interaction modify a user's emotional state?" in *World Haptics Conference (WHC)*, 2013, pp. 247–252.
- [28] Akshita, H. Alagarai Sampath, B. Indurkha, E. Lee, and Y. Bae, "Towards multimodal affective feedback: Interaction between visual and haptic modalities," in *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, p. 2043–2052.
- [29] C. Krogmeier, C. Mousas, and D. Whittinghill, "Human-virtual character interaction: Toward understanding the influence of haptic feedback," *Computer Animation and Virtual Worlds*, vol. 30, no. 3–4, 2019.
- [30] K. Koumaditis, F. Chinello, and S. Venckute, "Design of a virtual reality and haptic setup linking arousals to training scenarios: A preliminary stage," in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 1–2.
- [31] U. Radhakrishnan, K. Koumaditis, and F. Chinello, "Investigating haptic feedback and arousal for motor skill training in virtual reality," in *Haptics: Science, Technology, Applications: 13th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications (EuroHaptics), Proceedings*. Springer Nature, 2022, p. 472.
- [32] J. K. Gibbs, M. Gillies, and X. Pan, "A comparison of the effects of haptic and visual feedback on presence in virtual reality," *International Journal of Human-Computer Studies*, vol. 157, 2022.
- [33] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer, "The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment," *PLoS one*, vol. 13, no. 2, 2018.
- [34] M. Zhou, D. Jones, S. Schwaitzberg, and C. Cao, "Role of haptic feedback and cognitive load in surgical skill acquisition," in *Proc. Human Factors and Ergonomics Society Annual Meeting*, vol. 51, no. 11, 2007, pp. 631–635.
- [35] C. Ramírez-Fernández, A. L. Morán, and E. García-Canseco, "Haptic feedback in motor hand virtual therapy increases precision and generates less mental workload," in *Conf. Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 2015, pp. 280–286.
- [36] B. Weber, M. Sagardia, T. Hulin, and C. Preusche, "Visual, vibrotactile, and force feedback of collisions in virtual environments: effects on performance, mental workload and spatial orientation," in *International Conference on Virtual, Augmented and Mixed Reality*, 2013, pp. 241–250.
- [37] A. Bandura, "The explanatory and predictive scope of self-efficacy theory," *J. social and clinical psychology*, vol. 4, no. 3, p. 359, 1986.
- [38] Y. Shu, Y.-Z. Huang, S.-H. Chang, and M.-Y. Chen, "Do virtual reality head-mounted displays make a difference? a comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments," *Virtual Reality*, vol. 23, no. 4, pp. 437–446, 2019.
- [39] P.-A. Cabaret, T. Howard, C. Pacchierotti, M. Babel, and M. Marchal, "Perception of spatialized vibrotactile impacts in a hand-held tangible for virtual reality," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, 2022, pp. 264–273.
- [40] M. Aggravi, F. Pausé, P. R. Giordano, and C. Pacchierotti, "Design and evaluation of a wearable haptic device for skin stretch, pressure, and vibrotactile stimuli," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2166–2173, 2018.
- [41] E. Basalp, P. Wolf, and L. Marchal-Crespo, "Haptic training: which types facilitate (re) learning of which motor task and for whom answers by a review," *IEEE Trans. Haptics*, 2021.
- [42] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.
- [43] M. Granato, D. Gadia, D. Maggiorini, and L. A. Ripamonti, "Software and hardware setup for emotion recognition during video game fruition," in *Proc. 4th EAI International Conference on Smart Objects and Technologies for Social Good*, 2018, p. 19–24.
- [44] G. Makransky, L. Lilleholt, and A. Aaby, "Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach," *Computers in Human Behavior*, vol. 72, pp. 276–285, 2017.
- [45] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments," *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.
- [46] R. Orsila, M. Virtanen, T. Luukkaala, M. Tarvainen, P. Karjalainen, J. Viik, M. Savinainen, and C.-H. Nygård, "Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company," *International Journal of Occupational Safety and Ergonomics*, vol. 14, no. 3, pp. 275–283, 2008.
- [47] D. Narciso, M. Melo, J. V. Raposo, J. Cunha, and M. Bessa, "Virtual reality in training: An experimental study with firefighters," *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6227–6245, 2020.
- [48] C. Collet, C. Petit, A. Priez, and A. Dittmar, "Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation," *Biological psychology*, vol. 69, no. 2, pp. 195–203, 2005.
- [49] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," *Behavior research methods*, vol. 53, no. 4, pp. 1689–1696, 2021.
- [50] R. Magill and D. I. Anderson, *Motor learning and control: Defining and assessing learning*. McGraw-Hill Publishing New York, 2016, pp. 256–272.
- [51] J. J. Abbott, P. Marayong, and A. M. Okamura, "Haptic virtual fixtures for robot-assisted manipulation," in *Robotics research*. Springer, 2007, pp. 49–64.
- [52] F. Chinello, C. Pacchierotti, J. Bimbo, N. G. Tsagarakis, and D. Prattichizzo, "Design and evaluation of a wearable skin stretch device for haptic guidance," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 524–531, 2017.
- [53] L. Devigne, M. Aggravi, M. Bivaud, N. Balix, C. S. Teodorescu, T. Carlson, T. Spreters, C. Pacchierotti, and M. Babel, "Power wheelchair navigation assistance using wearable vibrotactile haptics," *IEEE Trans. Haptics*, vol. 13, no. 1, pp. 52–58, 2020.
- [54] L. Kuang, M. Aggravi, P. R. Giordano, and C. Pacchierotti, "Wearable cutaneous device for applying position/location haptic feedback in navigation applications," in *Proc. IEEE Haptics Symposium*, 2022, pp. 1–6.
- [55] A. Prabhu, W. Smith, Y. Yurko, C. Acker, and D. Stefanidis, "Increased stress levels may explain the incomplete transfer of simulator-acquired skill to the operating room," *Surgery*, vol. 147, no. 5, pp. 640–645, 2010.
- [56] J. A. Quick, A. D. Bukoski, J. Doty, B. J. Bennett, M. Crane, and S. L. Barnes, "Objective measurement of clinical competency in surgical education using electrodermal activity," *Journal of Surgical Education*, vol. 74, no. 4, pp. 674–680, 2017.
- [57] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, "Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task," *IEEE Trans. Affective Computing*, vol. 1, no. 2, pp. 109–118, 2010.
- [58] L. Arsintescu, R. Chachad, K. B. Gregory, J. B. Mulligan, and E. E. Flynn-Evans, "The relationship between workload, performance and fatigue in a short-haul airline," *Chronobiology International*, vol. 37, no. 9–10, pp. 1492–1494, 2020.
- [59] S. Grassini, K. Laumann, and M. Rasmussen Skogstad, "The use of virtual reality alone does not promote training performance (but sense of presence does)," *Frontiers in psychology*, vol. 11, p. 1743, 2020.
- [60] Y. Y. Yurko, M. W. Scerbo, A. S. Prabhu, C. E. Acker, and D. Stefanidis, "Higher mental workload is associated with poorer laparoscopic performance as measured by the nasa-tlx tool," *Simulation in healthcare*, vol. 5, no. 5, pp. 267–271, 2010.
- [61] D. Stevens, D. I. Anderson, N. J. O'Dwyer, and A. M. Williams, "Does self-efficacy mediate transfer effects in the learning of easy and difficult motor skills?" *Consciousness and cognition*, vol. 21, no. 3, pp. 1122–1128, 2012.
- [62] C. Pacchierotti, "Cutaneous haptic feedback for robotics and Virtual Reality," *Habilitation à diriger des recherches*, Université de Rennes 1, Rennes, 2022.
- [63] S. S. Kantak and C. J. Winstein, "Learning-performance distinction and memory processes for motor skills: A focused review and perspective," *Behavioural brain research*, vol. 228, no. 1, pp. 219–231, 2012.
- [64] C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo, "Wearable haptic systems for the fingertip and the hand: taxonomy, review, and perspectives," *IEEE Trans. Haptics*, vol. 10, no. 4, pp. 580–600, 2017.
- [65] X. De Tinguy, C. Pacchierotti, M. Marchal, and A. Lécuyer, "Enhancing the stiffness perception of tangible objects in mixed reality using

wearable haptics,” in *Proc. IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 81–90.

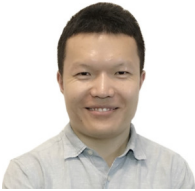
- [66] F. Chinello, M. Malvezzi, D. Prattichizzo, and C. Pacchierotti, “A modular wearable finger interface for cutaneous and kinesthetic interaction: control and evaluation,” *IEEE Trans. Industrial Electronics*, vol. 67, no. 1, pp. 706–716, 2019.
- [67] G. Richard, T. Pietrzak, F. Argelaguet, A. Lécuyer, and G. Casiez, “Within or between? comparing experimental designs for virtual embodiment studies,” in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 186–195.
- [68] R. A. Bjork and W. B. Whitten, “Recency-sensitive retrieval processes in long-term free recall,” *Cognitive Psychology*, vol. 6, no. 2, pp. 173–189, 1974.



Claudio Pacchierotti (SM’20) is a tenured researcher at CNRS-IRISA in Rennes, France, since 2016. He was previously a postdoctoral researcher at the Italian Institute of Technology, Genova, Italy. Pacchierotti earned his PhD at the University of Siena in 2014. He was Visiting Researcher at the University of Pennsylvania in 2014, at the University of Padua in 2013, at the University of Twente in 2014, and at Sapienza University of Rome in 2022. Pacchierotti received the 2014 EuroHaptics Best PhD Thesis Award and the 2022 CNRS Bronze Medal. He is Senior Chair of the IEEE Technical Committee on Haptics, Co-Chair of the IEEE Technical Committee on Telerobotics, and Secretary of the Eurohaptics Society.



Unnikrishnan Radhakrishnan is a Ph.D. Fellow at the Department of Business Development and Technology, Aarhus University. Previously, he worked as a software engineer at AMMACHI labs building skill training simulators for government and the industry. He holds a Master of Computer Applications degree from Amrita University. His current research interests are in the applications of VR/AR, haptic feedback and biosensors to support education and training.



Lisheng Kuang is a PhD candidate at CNRS-IRISA in Rennes, France. He got a master degree of control engineering from Harbin Institute of Technology (Shenzhen), in 2017. He was a research associate at the Italian Institute of Technology, Genova, Italy in 2019. He had five months visiting the Extended Reality and Robotics (xR²) Lab at Aarhus University in 2022. His research focuses on the design, prototyping, testing, and control of wearable robots for applications in human mobility aids.



Konstantinos Koumaditis is an Associate Professor at the Department of Business Development and Technology of Aarhus University and co-director of the Extended Reality and Robotics lab (xR²). Holds a bachelor’s and Master’s degrees on Engineering and management UK and a PhD Greece. In the past worked as a freelance engineer and participated and managed teams of developers from design to production in EU projects. Currently Konstantinos leads Danish funded projects on immersive technologies.

Konstantinos has published several research articles in international conferences/journals, served as guest editor in journals and chaired conference venues (like NordiCHI tutorials and CHI journals).



Francesco Chinello He got in 2014 his PhD degree on Information Engineering and Robotics, at the University of Siena (Italy). After his Post-Doc at the University of Siena and at the Italian Institute of Technology (Italy), he continued his career at the Aarhus University (BTECH department) (Denmark) as a teacher and researcher, co-responsible of the Extended Reality and Robotics lab. His research interests are in Haptics, Robotics applications and VR technologies.