This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

Adapting Virtual Embodiment through Reinforcement Learning

Thibault Porssut, Yawen Hou, Olaf Blanke, Bruno Herbelin, and Ronan Boulic, *Senior Member, IEEE*

Abstract

In Virtual Reality, having a virtual body opens a wide range of possibilities as the participant's avatar can appear to be quite different from oneself for the sake of the targeted application (e.g. for perspective-taking). In addition, the system can partially manipulate the displayed avatar movement through some distortion to make the overall experience more enjoyable and effective (e.g. training, exercising, rehabilitation). Despite its potential, an excessive distortion may become noticeable and break the feeling of being embodied into the avatar. Past researches have shown that individuals have a relatively high tolerance to movement distortions and a great variability of individual sensitivities to distortions. In this paper, we propose a method taking advantage of Reinforcement Learning (RL) to efficiently identify the magnitude of the maximum distortion that does not get noticed by an individual (further noted the detection threshold). We show through a controlled experiment with subjects that the RL method finds a more robust detection threshold compared to the adaptive staircase method, i.e. it is more able to prevent subjects from detecting the distortion when its amplitude is equal or below the threshold. Finally, the associated majority voting system makes the RL method able to handle more noise within the forced choices input than adaptive staircase. This last feature is essential for future use with physiological signals as these latter are even more susceptible to noise. It would then allow to calibrate embodiment individually to increase the effectiveness of the proposed interactions.

Index Terms

Virtual Reality, Embodiment, Machine learning, Reinforcement Learning, Motion Capture, Movement Distortion.

EDICS Category: 3-BBND

- *Thibault Porssut is with the Immersive Interaction research Group and Foundation Bertarelli Chair in Cognitive Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, 1015, Switzerland.*
- *Yawen Hou was with the Immersive Interaction research Group , 1015, Switzerland.*
- *Olaf Blanke leads the Laboratory of Cognitive Neuroscience, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne, 1015, Switzerland Switzerland.*
- *Bruno Herbelin is with the Foundation Bertarelli Chair in Cognitive Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, 1015, Switzerland.*
- *Ronan Boulic leads the Immersive Interaction research Group, Ecole Polytechnique Fédérale de Lausanne, 1015, Switzerland.*

Manuscript received to be added; revised to be added. Corresponding author: T. Porssut (email: thibault.porssut@epfl.ch).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

Adapting Virtual Embodiment through Reinforcement Learning

1 Introduction

Immersive VR is an effective complement to classical motor learning and rehabilitation, where participants execute repetitive physical exercises in motivating and controlled conditions [1], [2], [3]. The truly unique feature of VR motor training however lies in the possibility to use *adaptive training* methods. It consists of adjusting the visual feedback given on the executed movement in order to push participants to overcome their own performance limits or to correct for unconscious biases [4], [5], [6], [7], [8], [9]. This is achieved by hiding from view the actual movement of the physical body and by replacing it with a visual feedback that is distorted. Under such conditions, participants adapt their motor commands to compensate for the distortion between actual and virtual movements. It has been shown that the motor adaptation resulting from this dissociation is not consciously perceived (even for rather large differences [10], [11]), and leads to improved motor control and compensation mechanism [12], [13], [14]. In the context of fully immersive VR, an avatar of the participants' body is displayed in a first person and animated in real-time so as to provide participants with the strong sensation of seeing their own body inside the virtual environment. Finding the perceptual thresholds allowing concurrently motor adaptation and feeling of owning the virtual body is delicate, and has been studied only recently [15].

The subjective feeling of owning a virtual body as one's own is defined as the Sense of Embodiment (SoE). According to Kilteni et al. [16], SoE can be decomposed into three sub-components: body ownership, agency, and self-location. The sense of self-location is defined as one's spatial experience of being inside the body. The sense of agency is described as the sensation of being in control of the body, that one's limbs are moving according to one's will. Finally, the sense of body ownership refers to one's self-attribution of a body, implying that the latter is the source of sensations. If one of the three components is disrupted, a Break In Embodiment (BIE) occurs [17], [18], leading to the sensation of loosing control and dissociating from the avatar's body.

To offer a seamless control of the virtual body during distortion while avoiding Break In Embodiments (BIEs), Debarba et al. [19] proposed to use linear distortion functions for adjusting the avatar's hand's position relatively to the virtual object to reach. Such distortions are well tolerated and not noticed until an external event occurs, such as the haptic feedback from the touch of one's body [20]. But in absence of such disruptions and when relying solely on visual feedback, studies have shown that participants tolerate rather large levels of distortion [15], [18], [21], [22], [23]. Furthermore, the detection threshold increases if the external force helps them to reach their goal [19]. To evaluate the perceptual threshold of distortions, Porssut et al. [18] asked participants to perform a non-biological movement with their hand along an elliptic trajectory. While apparently doable, the task is very difficult to accomplish without the help of an additional distortion. Results show that, although extremely high distortions indeed break the sense of embodiment, a large magnitude of deviation is generally accepted and can be further extended if introduced progressively. Indeed, participants still experience body ownership over the virtual body, i.e. no BiE happened, even if they notice that the movement of the virtual body deviates from their actual physical movement [18].

However, avatar-body movement distortions suffer from a large variability in the detection threshold between subjects [18], [19], [20]. It would require to adjust the magnitude of the distortion to each individual to apply a well-controlled distortion in a training or rehabilitation context. Previous studies have tried different evaluation methods, such as the standard staircase and the Point of Subjective Equality (PSE) [18], [19], [24] to find the optimal distortion value for each subject. The optimal distortion is defined as the distortion magnitude for which subjects can have their movement distorted without noticing it. These methods rely on the explicit-feedback from the user (thus interrupting the flow of interactions). Therefore, they rely on a signal which has a good signal-to-noise ratio. Hence they are not applicable in most application contexts using certain physiological measurements with lower signal-to-noise ratio (EEG,GSR..). The use of physiological measurements opens the opportunity to gather implicitly the subjects'feedback, preventing any interruption of the interaction flows. Here, we propose to address this issue by proposing a more flexible (process signals with a lot of noises) and robust (find the right threshold) approach to adapt the distortion to each subject.

RL algorithms (SARSA and Qlearning) have already been used online to discover a choice (among a predefined set) by taking advantage of the detection of error-related potentials with EEG [25], [26]. Based on these prior results RL seems to be also suited to identify each subject's individual threshold that, as opposed to a choice they consciously make, characterizes their unconscious sensitivity to embodiment distortions. To our knowledge, none of the prior studies have used RL to compute such a detection threshold. This is more challenging since we try to evaluate human perception during a continuous movement. The answer is not known in advance and depends on each subject contrary to a simple task choice where the choice is done among a predefined set. Therefore, this study aims to demonstrate that RL is able to find the detection threshold and to do it more efficiently than prior art approaches. It opens up many new possibilities like implicit calibration of VR interactions.

Given the wide range of reinforcement learning (RL) algorithms used in previous studies [25], [26], [27] we first assess the main RL algorithms through computer simulation

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

for efficiency reasons. In a second stage the selected RL algorithm is compared to the staircase algorithm [28] in a controlled experiment with subjects performing the biological movement task with the distortion model from Porssut et al [18].

The contributions of this paper are threefold: (i) Providing a new online, implicit measurement-compatible method to determine the user's perception threshold with regards to distortion; (ii) extend [18] by giving a framework to automatically calibrate an interaction, and (iii) investigate the cause of the detection variability.

The paper is organized as follows. Section 2 recalls the state of the art on the sense of embodiment in Virtual Reality, psychometric methods used to find a threshold and online methods to calibrate a system. Then section 4 presents the experimental framework and section 5 the results. The general discussion and the conclusion end the paper.

2 Related Work

The use of explicit distortion has often been used for 3D interaction without any haptic feedback. Several different hand remapping techniques have been elaborated implying the hand only or the whole body. First studies have focused on the hand only. For instance, Burns et al [24], [29] add a constant offset between the subject real hand and the virtual one. Since only the hand is shown (floating hand), this offset is only applied to the hand. Bowman et al. [30] add the user's arm to the hand and focus on making the subject reach actions more effectively by explicitly stretching the arm while moving the virtual hand toward the object to reach. Since the arm is not tracked, the distortion is still limited to the hand location. Indeed, there is no mapping between the subject's real arm and the virtual one. Debarba et al [19] extend the distortion to the full arm configuration by tracking the subject's full body. The distortion is linear toward the target to reach. Molla et al. [31] introduced a performance animation algorithm allowing to transfer the participant movement onto an avatar of different shape while preserving consistent self-body contacts. Finally, Azmandian et al. [32] redirect the hand using body warping and world mapping extending the distortion to the whole body and world. In their case the virtual body is a means to enable haptic redirection.

Most of these studies could identify a threshold under which participants do not perceive that a distortion has been applied between the apparent movement of the avatar (seen in first-person perspective in VR) and the actual movement they performed. Indeed, Burns et al. [24] show that, in absence of haptic feedback, visual feedback prevails on the proprioceptive feedback whenever there is a perceptual conflict [24], [29]. However, the SoE's critical importance for controlling and owning a virtual body is not explicitly studied in these prior works. Since the SoE has an important impact on the user experience [33], [34], most of the current studies use distortion including the SoE as a central part of the evaluation of the subjective experience. For example, Kokkinara et al. [35] perform a spatio-temporal distortion of the participant's movement (2 to 4 times faster) during a reaching task with the arm fully extended. They observe a visuo–proprioceptive remapping, and a significant dip in agency, but not in body ownership. Debarba et al. propose refined spatio-temporal distortions for finger pointing movements on a tangible surface [36]. Similarly, Zenner et al. [37] show that even without any haptic feedback the hand position can be redirected during a pointing task. Esmaeili et al. [21] show that the detection threshold for hand distortion changes according to the movement direction. Bovet et al. [20] go further by using Molla et al [31]'s algorithm to show that a self-contact distortion can be very detrimental for body ownership. Finally, Ogawa et al. [22] show that a realistic avatar (a floating virtual hand) increases the self-attribution compared to an abstract avatar (a sphere), suggesting that subjects notice the distortion less when a realistic avatar is used.

	Staircase	Psychometric	S+P	RL
Robustness to Noise				X
Online	X			X
Model Transfer		X	X	X
Dynamic Data Size	X		X	X
Previous Studies	[20], [36] [38], [39] [40]	[38], [39] [40], [41] [18], [36]	[39], [40] [29], [42] [43], [44]	[25], [26] [27], [45] [46], [47]

TABLE 1: Key features of the threshold identification methods (S+P= Staircase + Psychometric)./ Robustness to Noise: characterizes the performance despite the noise in subject's answers (subject's mistakes); Online: allows direct threshold computation without the need of an extra method like psychometric function (better for offline computation); Model Transfer: model can be updated during a new session with additional data; Dynamic data size: no need to know in advance the number of data needed to find a subject's threshold; Previous Studies: are the main studies using these methods, except for RL for which these studies demonstrate its potential as none have used RL to identify such a detection threshold yet.

Table 1 compares the three main families of prior art methods together with reinforcement Learning (RL). As highlighted in this table Reinforcement Learning (RL) algorithms seem suited for the online distortion value adjustment, as the calibration of the distortion is completely subjective and needs to be adapted over time. Past studies using RL in robotics [25] [26] [27] [46] demonstrated that a single negative feedback (transmitted when subjects feel an error) is enough for certain RL algorithms to train a satisfying model. For example, Kim et al. [45] had their robot learning user-defined gestures and the association of each gesture with a predefined robot action using intrinsic RL techniques [45]. Luo et al. [27] manage to train a RL agent to perform a binary-choice task in an online experiment using a single type of negative reward as feedback. In [26], the task was to have a robotic arm learn the position of the basket chosen by the user. The accuracy varies depending on the position of the chosen basket, but in general the system managed to learn the correct position. However our context is different in the sense that the unknown threshold to discover with RL is not even explicitly known to the subjects themselves. That's why, for obvious usability reasons and

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

given the variety of RL algorithms [25], [26], [27], [45], [46] we first have to identify the one converging with the smallest number of iterations. To this end, the present study first compares several RL algorithms in Section 3.

Another major concern is related to the inherent noise in subjects'answers. Indeed, over time subjects may tolerate more the distortion and change their threshold or they might get tired and answer wrongly. Adaptive staircase and RL algorithms are, in general, sensitive to perturbed answers. They will take more time or have difficulties to converge. However, RL algorithms are more robust to noise [47] and can even be used with signal like EEG [25], [26]. To our knowledge no study has used adaptive staircase with EEG signals.

3 Reinforcement Learning

Before running an actual experiment, we need to ensure that the selected RL algorithm is usable with subjects. It has to converge at least below 200 iterations. Indeed, based on a pilot study with 5 subjects, a subject spends on average around thirty minutes to repeat the task (see 4.4) 200 times. Based on [25], [26], [27], [45], [46], we first chose to test with Qlearning and SARSA. To test the performances of RL algorithms, the reward function is designed as follows: when the distortion value is equal or below the subjects' threshold, subjects do not notice the distortion in their movements, which yields a positive reward. Once the distortion value exceeds their threshold, the subjects experience a Break in Embodiment (BIE), which yields a negative reward. The absolute value of the reward is the distortion gain. For instance, if a subject performs the task with a distortion gain of 2.5 and does not notice the distortion, then the RL agent receives a reward of 2.5. On the contrary, if the subject feels a BIE, then the RL agent receives a negative reward of -2.5. In the ideal situation, their threshold remains stable, thus this reward function entice the RL agent to learn the highest possible gain to find how much distortion subjects can tolerate. Then, the actions are defined as increasing or diminishing the gain by steps of 0.5 or 0.25. However, under this design, SARSA and Qlearning took more than 1000 iterations to converge ; the combination of the action space and the state space is too large to be fully explored in a small numbers of iterations.

Therefore, we have restricted our problem to a non-stationary multi-armed bandit (MAB) which is more adapted to find a threshold. Indeed, the k-armed bandit model omits the state-space and the transition probabilities to focus solely on finding the best action from a total of k choices by maximizing the total rewards over a predefined period of time [48]. Three different algorithms are tested to find the threshold.

3.1 Comparisons

A k-armed bandit problem is chosen with each arm corresponding to one of the k distortion gains. The agent receives a positive or a negative reward chosen from a non-stationary probability distribution (i.e. the reaction of subjects to the same distortion value might change over time) depending on the reaction of subjects to the selected gain. The threshold is found by maximizing the expected reward, namely the value function, conditioned by the action: $Q_*(a) = E[R_t|A_t = a]$ where R_t is the reward at time t and A_t is the action chosen at time t. The optimal detection threshold (action) is the value with the highest Q-value: $A_{opt} = \arg\max_a Q_*(a)$. As the Q-value of each action ($Q_*(a)$) is unknown at the beginning, an estimate of $Q_*(a)$ is updated for each action at each time step (update the Q-table). As the problem is non-stationary, the update rule at time step t is : $Q_{t+1} = Q_t + \alpha[R_t - Q_t]$ where α is the learning rate that diminishes over time. Each of them are evaluated with their cumulative rewards over 1000 steps. The best algorithm is the one converging to the desired cumulative rewards in the least iterations. Three different algorithms have been tested and compared: a simple ϵ-Greedy algorithm, Upper-Confidence-Bound (UCB) combined to ϵ-Greedy algorithm and Policy Gradient.

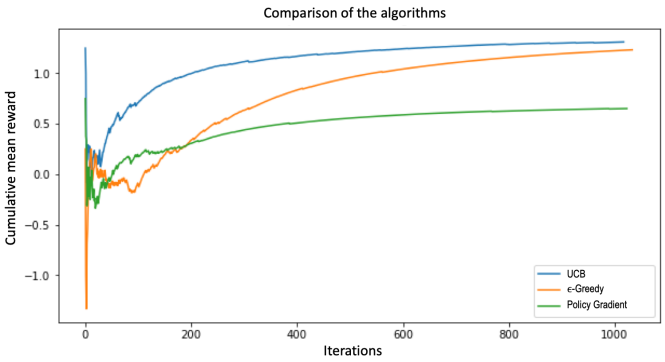


Fig. 1: Accumulated rewards of the three algorithms with UCB converging faster than the two other algorithms(1000 steps).

The results indicate that UCB outperforms the other two algorithms (see Fig.1) by converging within a relatively small number of iterations (150-200) as opposed to (500-550) for ϵ-Greedy and more than 1000 for Policy Gradient to reach the same cumulative mean reward of 1. For this reason we retain UCB for the controlled study with subjects.

3.2 Upper Confidence Bounds

UCB combined with the ϵ-Greedy policy is the RL algorithm chosen to find the detection threshold. As the best action is unknown at the beginning of the experiment, the next action is taken in the next time step using a mix of the UCB strategy and the ϵ-Greedy policy. For ϵ % of time, a random action is chosen, and for the $(1-\epsilon)$ % of time the action is chosen according to the UCB strategy: $A_t = \arg\max_a[Q_t(a) + c\sqrt{\frac{\ln t}{N_t(a)}}]$, where $c > 0$ controls the degree of exploration and $N_t(a)$ denotes the number of times action a has been chosen until the time step t. c is set to 2 for this experiment. Since the highest Q-value stops to change long before the Q-table converges, the convergence condition has been set to 15 unchanged consecutive iterations. These iterations are counting from the 35th trial. If UCB reaches 100 iterations, the algorithm is terminated. In this case, the gain with the highest Q-value is chosen as the detection threshold. The exploration ratio ϵ is set to 1 at start in order to let the algorithm fully explore the actions,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

and decay following $\epsilon_{t+1} = \epsilon_t - \log_{10} \frac{t+1}{20}$ (t is the number of iterations), until it diminishes to the minimal exploration ratio 0.01. Similarly, the initial learning rate α for the Q value update $Q_{t+1} = Q_t + \alpha[R_t - Q_t]$ is set to 0.5, and decays in the same fashion: $\alpha_{t+1} = \alpha_t - \log_{10} \frac{t+1}{40}$ until it reaches the minimal learning rate 0.001. The denominators of these two decaying rules are found through grid search.

4 Experiment

The experiment uses an implementation of the Attraction Well distortion [18] mechanism with the following parameter: R as the radius of the real tennis ball, d_{range} as the radius of the circular trajectory (0.35m) and G as the value returned by UCB / staircase.

In the following sections, the robustness of an algorithm is defined as its ability to prevent subjects from detecting the distortion when the applied gain is equal or below the threshold found by this algorithm. The conservativeness of the algorithm is defined as giving a too small threshold. In that case, subjects wouldn't notice the distortion when the applied gain is somewhat above this small threshold. For each subject and each algorithm, the robustness and the conservativeness of these algorithms are quantified using the true positive rate (TPR) (or recall) (see Eq.1) and the positive predictive value (PPV) (or precision) (see Eq.2). We also define the different components of the confusion matrix (see Sec. 5.2) as follows: the true positives (TP) are subjects noticing the distortion above the threshold, the true negatives (TN) are subjects not noticing the distortion below the threshold, the false positives (FP) are subjects not noticing the distortion above the threshold, the false negatives (FN) are subjects noticing the distortion below the threshold.

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

The purpose of this study is to find the detection threshold with adaptive staircase and UCB, and compare their performance. The ideal adaptive algorithm should be robust, not conservative, and fast-converging. The following hypotheses have been formulated:

First, the UCB algorithm is more robust than the staircase algorithm (H1). The "robustness" of the two algorithms is quantified by computing the TPR for each subject and algorithm. The lower the FN are, the higher the TPR is and the more robust an algorithm is for our application. Robustness is the most important criterion for this experiment. The UCB is hypothesized to be more robust, because a mechanism has been implemented to correct a portion of the subjects' reaction (Sec. 4.2). If subjects have never detected a distortion at a certain level of gain and suddenly detect it, their response is considered noisy and corrected.

Secondly, the UCB algorithm is more conservative than the staircase algorithm (H2). The "conservativeness" of the two algorithms is quantified by computing the PPV for each subject and algorithm. The lower the FP are, the higher the PPV is and the less conservative an algorithm is for our application. Ideally the conservativeness of the algorithm has to be as low as possible while keeping a high robustness. However, it is hard to have an algorithm that is very robust and not conservative simultaneously. Due to the way the reward function is defined ($R = \pm$ gain of distortion depending on if subjects experience BIE or not (Sec. 3.2)), the UCB might converge to a lower value than the staircase since negative rewards are very detrimental for UCB.

Finally, subjects who spent more time using an immersive 3D application have a higher detection threshold (H3). This hypothesis is more exploratory and tries to explain the subject inter-variability found in [18], [19]. The experiment has been designed around this idea with a demographic information survey added at the start of the experiment. Three subtopics are about the subject's experience with immersive 3D application (see Sec.4.4). A pilot study of 5 subjects was conducted to adjust the task of the experiment and the questions. Since several participants did not experience any BIE at the original maximal distortion value of 4, it has been increased to 10.

4.1 Distortion Function: the Attraction Well

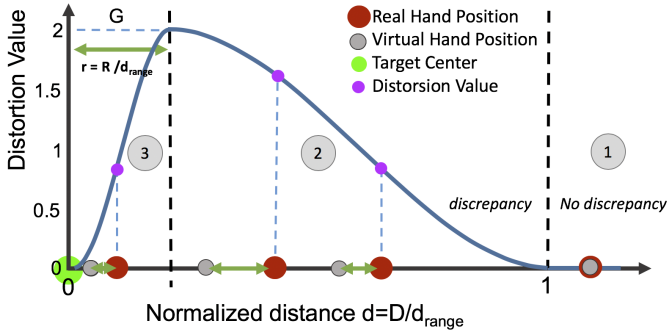


Fig. 2: Overview of the well-shaped distortion function. No distortion is applied when d is greater than 1, i.e. the virtual hand is placed exactly where the real hand is situated (region 1). When d is below 1, the virtual hand starts to get attracted to the target (its position is closer to the target than the position of the subjects' real hand). The attraction amplitude increases as d decreases from 1 to r (region 2), then diminishes to zero as d decreases from r to 0 (region 3).

We recall here the main features of the tracking task and its associated distortion function introduced in [18] and used for the controlled experiment of distortion detection. The distortion is designed to help a subject for tracking a target moving on an elliptic trajectory while preserving their sense of agency. The avatar's hand is first attracted towards the target until it reaches the outer boundary of the moving target (i.e. a sphere slightly bigger than a tennis ball). Then, once the virtual hand is inside the moving target, the attraction is progressively reduced to zero until the hand arrives at the target center.

The following notations are set: $\vec{P}_{target}$ is the 3D position of the moving target of radius R, $\vec{P}_{real}$ is the real 3D position of the subject's hand (more precisely, it is the position of the center of the tennis ball held by subjects as can be seen on Fig.6), D is the distance between the target and the hand with $D = \|\vec{P}_{target} - \vec{P}_{real}\|$, and d_{range} is the distance range

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

of the attraction force centered on the moving target. Finally, equations are expressed with the normalized variables d and r respectively given by $d = \frac{D}{d_{range}}$ and $r = \frac{R}{d_{range}}$.

Fig. 2 presents an overview of the well-shaped attraction profile from [18] ; its amplitude is expressed as a function of the normalized distance d. For $d > 1$, no distortion occurs, hence the virtual hand position coincides with the real hand position. An attraction is enforced whenever $d < 1$ thereby bringing the avatar hand closer to the target compared to the real hand. More precisely, this attraction is decomposed into an outer and an inner region delimited by the normalized moving target radius (respectively labeled 2 and 3 in Fig. 2). The outer region offers an increasing attraction until a maximum when the real hand reaches the boundary of the moving target. Conversely, the inner region reduces the attraction amplitude from this maximum down to zero when it coincides with the target. The maximum amplitude of the attraction force is denoted as G (illustrated with the value 2 in Fig. 2). Then:

$$f(d) = \begin{cases} d \in [0, r] & G \times (-2 \times (\frac{d}{r})^3 + 3 \times (\frac{d}{r})^2) \\ d \in [r, 1] & G \times (2 \times (\frac{(d-r)}{(1-r)})^3 - 3 \times (\frac{(d-r)}{(1-r)})^2 + 1) \end{cases} \quad (3)$$

Given the distortion value provided by the attraction profile $f(d)$, an attraction coefficient is computed $1/(1 + f(d)$ to build the distorted hand position $\vec{P}_{distorted}$, shown to subjects, from the knowledge of the current positions of the mobile target $\vec{P}_{target}$ and of the real hand $\vec{P}_{real}$. Then:

$$\vec{P}_{distorted} = \vec{P}_{target} + (\frac{1}{1 + f(d)}) \times (\vec{P}_{real} - \vec{P}_{target}) \quad (4)$$

The distortion amplitude $f(d)$ being always positive, Equation 4 ensures that the distorted hand position $\vec{P}_{distorted}$ always lies in-between the current target position $\vec{P}_{target}$ and the real hand position $\vec{P}_{real}$. Moreover both the real and the distorted positions coincide for the boundaries $[0, 1]$ of the normalized distance d.

The distortion is tuned through the three following parameters. R is the radius of the moving target (tennis ball). d_{range} is the distance range of the attraction force centered on the moving target, which corresponds to the radius (0.35m) of the circular trajectory the target follows (Sec. 4.4). Lastly, G (referred to as the "distortion gain" or "gain" in the following paragraphs) is the maximum amplitude of the attraction (G = 2 in Fig. 2). Thanks to the well-shaped attraction profile, subjects who are successful at closely pursuing the target may notice that the small imperfections of their trajectory are reflected in the movement of the virtual hand, hence conveying a sense of being in full control of the performed movement. Conversely, if subjects have some issues following the target, their tracking error makes them fall within the region with a high attraction, which makes following the target easier. Based on previous results [18] and pilot testing, the following discrete values of distortion gain are used: {0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 4, 5, 7, 10}. The last values are separated by a larger step-size because, as the gain increases, it becomes harder for subjects to detect a difference between two distorted movements.

4.2 Implementation

Equipment and software

Fig. 3: Equipment: subjects have a tracker on each shoulder, elbow, and hand, as well as a tracker on their chest. Subjects hold the tennis ball in the right hand and the HTC Vive Controller in the left hand.

The HTC Vive Pro Eye, a Head Mounted Display (HMD) with 1440 x 1600 pixels per eye, 110 ° field of view and 90 Hz refresh rate, is used for display. This headset has a 120 Hz eye tracking system with 0.5-1.1 ° of accuracy. The eye tracking is used to ensure that subjects are always looking at their right hand in the second phase of the task. Bose QuietComfort 35 wireless headphones with active noise canceling are used to play a non-localized white noise to the subject during the experiment. The white noise is interrupted when communicating with subjects. For the motion capture 8 HTC Vive Trackers V2 are placed at the origin of the room (in front of the chair where subjects sit) (1), on the subject's chest (1), shoulders (2), elbows (2) and hands (2). Subjects also hold an HTC Vive Controller in their left hand to answer questions while being immersed in the virtual environment. To ensure a good tracking during all the experiment, one HTC's base station is placed in each corner of the room.

The virtual environment is a square room of $6 \times 6 \times 3m^3$ with a chair in the middle of the room. An avatar holding a tennis ball in the right hand is calibrated to collocate with the subject's body. This maintains a visuo-proprioceptive and tactile coherence between the real and virtual hands in the absence of finger tracking. The application is implemented using Unity 3D 2019.2.0f1. Subjects' movements are reproduced through animation by the avatar using LimbIk from FinalIK[1] package. Subjects are seated for the whole

1. root-motion.com

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

duration of the experiment and only need to perform simple movements with their right hand. Finally, the performance of the application is monitored through the Unity Profiler to maintain a constant frame rate (90 FPS).

The avatar is globally scaled based on the subject's height. After manually calibrating the position of the chest and the shoulders, subjects have to remain seated with their arms extended horizontally in the forward direction to align the trackers attached to their hands with simple rectangles attached to the avatar's hands. The length of the arms and the position of the hands are manually adjusted during this step as well. The avatar mesh is not visible during the calibration to prevent subjects from viewing visual interpenetration.

Staircase Design

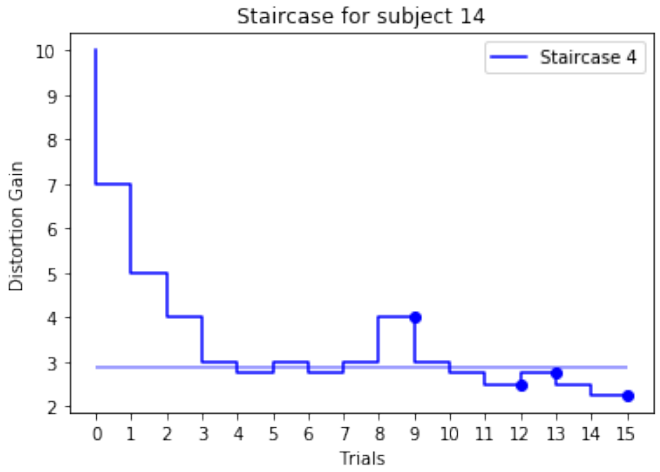


Fig. 4: Example of a converging staircase (7 turns in direction). The dots indicate the last 4 changes in direction (used to obtain the threshold). The horizontal line indicates the detection threshold, calculated using the mean value of the distortion gains of the last 4 turns in direction.

The staircase implementation follows the one of Bovet et al. [20] with slight modifications. As the step size between our gain values is not constant, the step value used in [20] is dismissed. The previous/next step of a gain value is its previous/next direct neighbor. The staircase block runs 4 staircases in parallel, each having a different starting gain (two starting low (0, 1) and two starting high(7, 10)). The staircases are presented at each iteration in a random order to prevent subjects from getting used to the distortion. If subjects detect the distortion, the gain is lowered, and vice-versa. Once a staircase converges, it returns a random gain until all the staircases converge or terminate. The stopping and convergence criteria remain unchanged: the staircase converges when the direction changes seven times (seven staircase turns) or when it reaches 20 iterations. The detection threshold is calculated using the mean of the distortion gains of the last four turns in direction.

Perturbed rewards

Since subjects are not aware of their threshold, their reaction to the same distortion gain is not stable. Additionally, the reaction of the subjects might be noisy due to fatigue or to the training effect. RL algorithms are, in general, vulnerable to perturbed rewards. The UCB algorithm might have difficulty converging when the reaction of the subjects to the same distortion gain is very unstable. Sometimes, it might also converge to a wrong threshold (e.g. 0, aka no distortion, when subjects don't experience a BIE at a higher distortion gain). Wang et al. [47] proposed a method to counter this problem. When the RL agent receives the perturbed reward from subjects' answer, it predicts the true reward based on the accumulated history of rewards for the corresponding distortion gain (majority voting). For instance, subjects have encountered the gain value g four times and have only detected it once. Then in the next iteration if they detect the gain g, the response is considered noisy and is corrected to "No detection".

4.3 Participants

A pilot study was run with 5 participants. The experiment has been conducted with 22 subjects. Before starting the experiment, subjects signed a consent form approved by the ethical committee of the canton of Vaud. Then they were asked to fill in a form with questions about their background (gaming experience, previous experience with VR applications, etc.) The 22 subjects were paid 20 CHF per hour for their participation. One subject's data were discarded due to technical issues.

The 21 subjects included in the analysis are aged from 18 to 25, in average 21.14 ± 1.9 years old. Five of them are female, and sixteen of them are male. Twenty of them are right-handed and one of them is left-handed. Subjects have normal or corrected to normal vision.

4.4 Methods

The experiment is divided into two blocks: one using the staircase method, and the other one using the UCB method. The block order is counterbalanced between subject to avoid any bias.

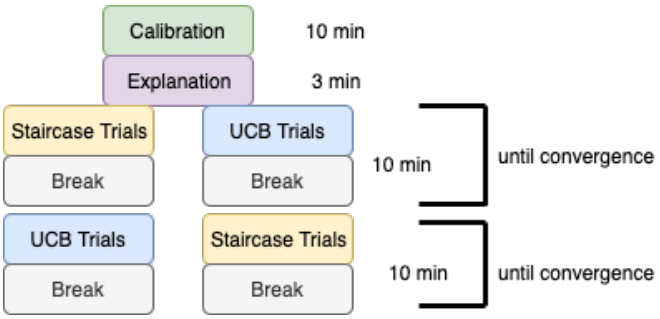


Fig. 5: Protocol Overview

The experiment proceeds in three steps. The first step is calibration (Sec.4.2) (approximately 10 min). The second step is explanation. The explanation consists of six trials alternating between no distortion gain (0) and maximal distortion gain (10) to show the distortion to the subjects. The last step is the task, divided into two blocks (Sec.4.4). The number of trials for each block is not predefined; it depends on the reaction of the subjects. Each block terminates

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

when the corresponding algorithm in charge of finding the detection threshold converges or terminates. There is a break every ten minutes to prevent from getting too tired. Subjects are asked to remain seated at all times to avoid technical issues. The experiment is resumed when they want. Each block is done only once. The entire experiment lasts for about one hour and ends with a short debriefing with subjects.

Task

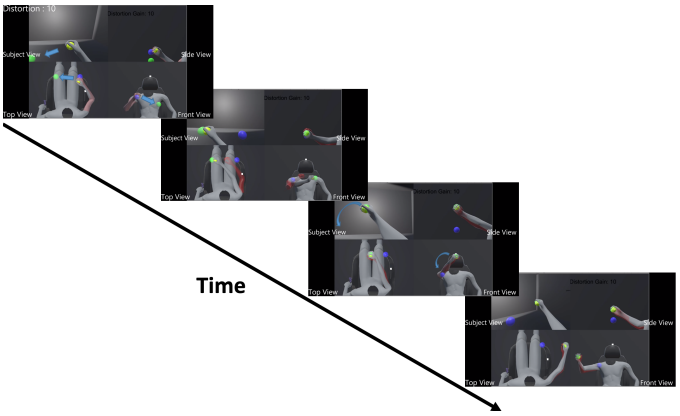


Fig. 6: Overview of a trial (Top Left: First person view; Top right: Side view; Bottom right: Top view; Bottom Left: Front view): The subject first puts the tennis ball inside the blue sphere, then follows the green sphere. The trial ends after the subject answers the question. In this figure, the red arm illustrates the subject's actual movement when following the target, while the grey arm belongs to the virtual avatar. Subjects are not shown the red arm during the experiment. (see the video for a sample of the experiment)

The task consists of a two-phase movement and a question. The task stays the same throughout the experiment. Subjects start with the tennis ball held in their right hand, in contact with their chest. In the first part of the movement, subjects have to put the tennis ball inside a semitransparent blue sphere in front of them and wait for a timer to finish. From this point in time, the eye tracking system is activated to track the eyes of the participants. They need to stay focused on the movement of their right hand, or else the trial is restarted. If the gaze of the subject is not fixed on the hand for 0.5s, the trial restarts after showing a warning message to them. While staying focused on their right hand, subjects need to move the tennis ball from the blue sphere to the center of a semitransparent green sphere. Then, they need to follow the green sphere that moves along a circular trajectory with a radius of 0.35m for a few seconds. This moment of transition is illustrated in Fig.7. The task is completed if they have followed their right hand's movement with their eyes and have maintained the tennis ball inside the green sphere during at least 4 seconds. Otherwise, the trial restarts from the beginning. Once the movement is completed, the target spheres and the avatar become invisible.

Then, a question and a cursor (white circle) appear on the virtual wall in front of the subjects. Subjects need to direct the cursor with their head orientation into the "Yes" / "No" circles, and press the controller's trigger for about

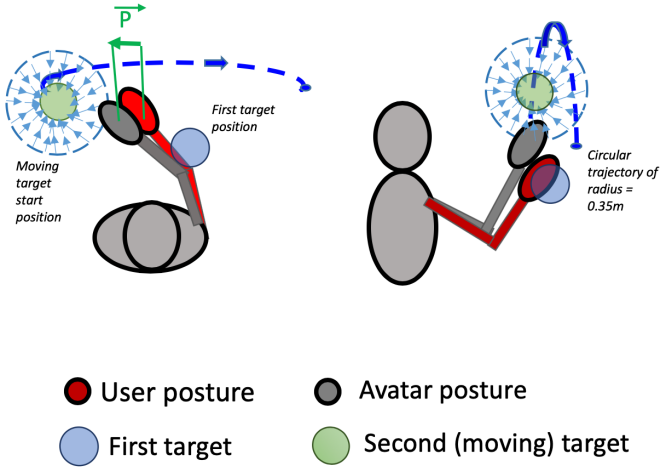


Fig. 7: This depicts the moment when the subject leaves the first target (blue sphere) and tries to reach the moving target (green sphere). (Left) Top view. In dark blue: the circular trajectory of the moving target (green sphere). The vector $\overrightarrow{P}$ is the discrepancy induced by the distortion. (Right) View from the right. The circular trajectory has a radius of 0.35m.

two seconds to answer the question. Once the question is answered, the objects in the virtual scene appear again at their initial position and the next trial starts.

Question

The same description has been given as in [49] to the subjects to explain the concept of Break In Embodiment (BIE). Kokkinara et al. [49] count the number of BIEs per session and make them say "Now" every time they experience a BIE. In our context, we want the subjects to keep focused on the task and we only need their feedback about the most recent trial. Hence They have to answer at the end of each trial with "Yes" or "No" to the question: "I felt at least once I didn't own the virtual body." This method allows the participants to answer without breaking the immersion.

Demographic Information

At the beginning of the experiment participants have to fill a demographic information survey. It is composed of seven subtopics (age, gender, handness, Main occupation, VR experience, Action video game experience, video game experience more generally). The last three are about the subject's experience with immersive 3D application ; they allow to explore the cause of subject inter-variability for the detection threshold and check the H3 hypothesis.

5 Results

A within subjects comparisons of UCB and staircase performance is performed using a pairwise t-test. Then the relationship between UCB and staircase detection threshold is examined using Pearson correlation. The assumption of normality of residuals is tested with the Shapiro-Wilk test. If residuals are deemed not normal the response is transformed with a Box-Cox transformation $y\lambda$, which does not alter the order of the response values. If the normality hypothesis is still rejected, a two-sided Wilcoxon signed-rank

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

test is performed. The exploratory analyses conducted with the data got from the demographic information survey is evaluated with a Spearman correlation. Since the same tests is used repeatedly on dependent datasets, FDR correction is applied for each Spearman correlation. For the statistical analysis, differences are deemed statistically significant for p-values below the threshold $\alpha = 0.05$. The analysis is conducted using Python.

5.1 Detection thresholds

5.1.1 Comparison

The threshold values are averaged across all subjects for UCB and Staircase (Fig.9d). On average, the UCB ($M = 1.93$ with $SD = 1.1$) set a significantly ($t_{21} = -2.2$ $p < 0.05$) lower threshold than the staircase ($M = 2.75$ with $SD = 1.45$). Even if both algorithms obtain quite a high average threshold, the UCB theshold is lower. These results may explain why the UCB is more robust. Indeed, if the threshold is too high compared to the real subject's threshold, there are more risks to produce FN. The subjects may experience breaks in embodiment more often, due to the more frequent perception of the distortion and also due to its bigger amplitude. Finally, the high standard deviation emphasizes the high variability of this detection threshold between the subjects. These results are in accordance with Porssut et al. [18].

5.1.2 Correlation

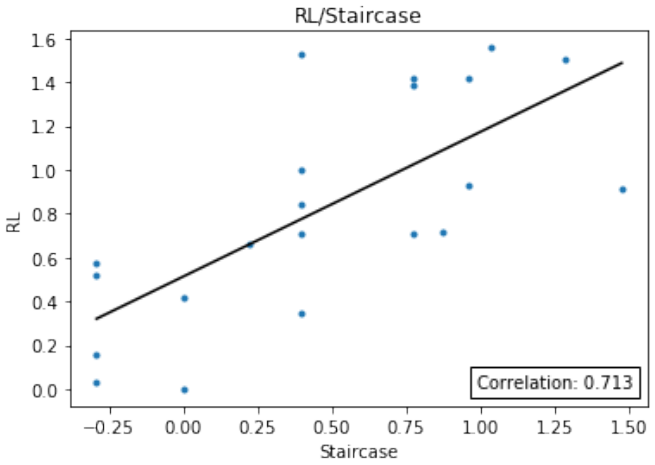


Fig. 8: Correlation between Staircase and UCB.

To explore the link between the staircase and UCB algorithms, Pearson correlations between their thresholds were conducted across subjects. The correlation analysis revealed a positive correlation $(r = 0.71, p < 0.001)$(see Figure 8). Even though previously a significant difference has been observed between the two average thresholds, this correlation ensures that the same type of threshold is measured and the difference is accounted for mostly by the robustness of UCB, i.e. the reduced number of false negatives. Indeed, the staircase's threshold increases at the same pace as UCB's threshold, they follow the same trend. This is consistent with the previous results and confirms that both thresholds are of the same nature.

		Actual Class	
		Distortion Detected	Distortion Not Detected
Predicted Class	Distortion Detected	TP=28.6/33.3	FP=14.6/21.7
	Distortion Not Detected	FN=18.3/13.6	TN=39.2/32.1

TABLE 2: Confusion Matrix Staircase/UCB (Average across all subjects divided by the number of trials per subject (%))

5.2 Confusion Matrix

The detection threshold is used to compute the confusion matrix for each subject. From this confusion matrix the TPR and PPV are computed for each user and each algorithm. The average confusion matrix across all subjects for staircase and UCB 2 is computed to give an overview of the algorithm answers. A total of 22 staircases that did not converge are discarded as in [19]. UCB gives more true positive than staircase but less true negative. On the contrary staircase gives more false negative but less false positive than UCB. Thus UCB seems more robust but less conservative than staircase. The PPV and TPR are computed in the following sections to check these first results.

5.3 Performance

5.3.1 Recall

The TPR is averaged across all subjects for UCB and Staircase (Fig.9b). Since the residuals are deemed not normal even after the Box-Cox transformation, a two-sided Wilcoxon signed-rank test is performed. The UCB ($M = 0.74$ with $SD = 0.27$) has a significantly ($t = 43$ $p = 0.008$) higher TPR than the staircase ($M = 0.62$ with $SD = 0.28$). Even if they both have quite a high TPR, the UCB is more robust. Thus, the hypothesis (H1) is validated. This means that the threshold obtained thanks to the UCB better prevents subjects from detecting the distortion.

5.3.2 Precision

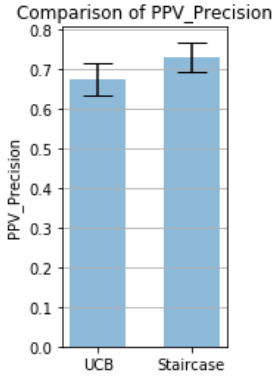
The PPV is averaged across all subjects for UCB and Staircase (Fig.9a). The UCB ($M = 0.67$ with $SD = 0.19$) has a lower PPV than the staircase ($M = 0.73$ with $SD = 0.18$). However, this difference is not significant ($t_{21} = -1.31$ $p = 0.37$). Surprisingly, both algorithms have a quite high PPV, which means they are not especially conservative. Thus, the hypothesis (H2) is not validated. However, UCB is at least equivalent if not less conservative than the staircase. Consequently, the threshold obtained is not much smaller than the real threshold of subjects. These thresholds prevent many false positive i.e. subjects not detecting the distortion when the applied gain is above the found threshold.

5.3.3 Classification Accuracy

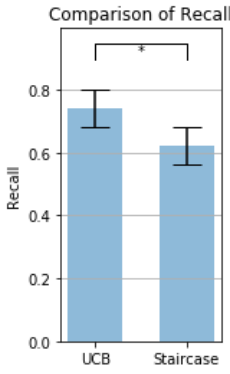
The classification accuracy is averaged across all subjects for UCB and Staircase (Fig.9c). The UCB ($M = 0.65$ with $SD = 0.095$) has a lower classification accuracy than the staircase ($M = 0.67$ with $SD = 0.11$). However, this difference is not significant ($t_{21} = -0.75$ $p = 0.46$). They both have quite a high classification accuracy. This means they both manage to set a threshold which is close enough to the real subject's threshold.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

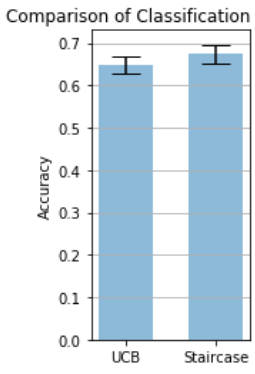
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics



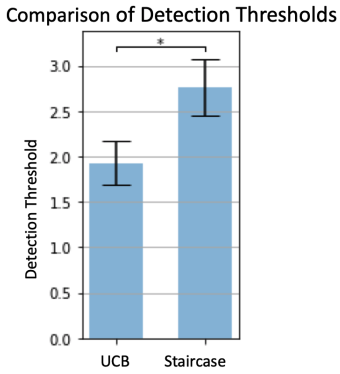
(a) Bar graph for the pairwise comparison of precision/PPV. Error bars represent the standard error of the mean.



(b) Bar graph for the pairwise comparison of recall/TPR. Error bars represent the standard error of the mean.



(c) Bar graph for the pairwise comparison of recall/TPR. Error bars represent the standard error of the mean.



(d) Bar graph for the pairwise comparison of detection thresholds. Error bars represent the standard error of the mean.

Fig. 9: Performance and Detection Threshold Results

5.4 3D Immersive Experience

We examined whether the subjects' thresholds variability (Sec.5.1.1) could be linked to the subject's experience with 3D immersive application by computing their correlation.

Firstly, the correlation analysis revealed no correlation ($r = 0.205, p = 0.37, Q_{FDR} = 0.44$) between hours of video games and Staircase convergence threshold. However, the correlation analysis revealed a positive correlation ($r = 0.45, p = 0.041, Q_{FDR} = 0.25$) for UCB convergence threshold but it doesn't survive the correction for multiple comparison. This difference of results between UCB and Staircase can be explained again because of the TPR/Recall value which is higher for UCB. The higher number of FN for Staircase in contrast to UCB's limited amount of FN, may hide the influence of this new factor on the threshold value. Even though the correlation between video games experiences and perception threshold didn't survive, it gives relevant information to study this high variability. Identifying more of those factors could be an interesting extension to this study.

Secondly, the link between action video games experiences and perception threshold, for both UCB and staircase, is analyzed. The correlation analysis revealed no correlation between hours of video games and Staircase convergence threshold ($r = 0.148, p = 0.52, Q_{FDR} = 0.52$) and no correlation also with UCB convergence threshold ($r = 0.326, p = 0.15, Q_{FDR} = 0.26$) for UCB convergence. Action video game doesn't seem to be one of the factors causing this variability.

Finally, the link between VR experiences and perception threshold, for both UCB and staircase, is studied. The correlation analysis revealed no correlation between hours of video games and Staircase convergence threshold ($r = 0.310, p = 0.17, Q_{FDR} = 0.26$) and no correlation also with UCB convergence threshold ($r = 0.343, p = 0.13, Q_{FDR} = 0.26$) for UCB convergence. VR experience doesn't seem to be one of the factors causing this variability.

Since no significant correlation has been found between the three scores, H3 is not validated. Indeed, thanks to all these results it may be interesting to do a new study with a more complete questionnaire about video game background (habit, exposure..) [50], [51].

6 Discussion

This study explores the problem of finding the detection threshold for individual subjects. To determine what is the best threshold, three criteria are used: 1) Robustness (Precision); 2) Conservativeness (Recall); 3) Accuracy. The robustness is evaluated thanks to the TPR and the PPV averaged across subjects computed thanks to the confusion matrix. The ideal algorithm should be both robust and the least possible conservative, necessitating a low number of iterations to converge. It also needs to adapt to the future variations of subjects' threshold. Upper-Confidence-Bounds (UCB) is the reinforcement learning method selected and staircase is the method used as a reference. The two algorithms are compared with the three previous criteria.

Due to the design of the reward function and the correction mechanism of the reward, the UCB is hypothesized to be more robust (H1), but also more conservative than the staircase (H2). The result shows that UCB is more robust (higher TPR) than staircase, and that their conservativeness is comparable (no significant differences between their PPV). The UCB and staircase also have a similar classification accuracy. Thus, UCB would be a better choice considering the fact that robustness is more important than conservativeness to avoid provoking BIEs. Finally, even though both algorithms defined a significantly different threshold, the correlation analyses showed that they are, indeed, measuring the same threshold. Using reinforcement learning algorithms to solve the problem of finding the detection threshold seems to be a more robust alternative than a classic staircase.

Porssut et al. [18] mentioned a high variability of the thresholds among subjects. This phenomenon is also found in this study. Moreover, the observation that their subjects had a relatively high detection thresholds is coherent to what has been found. For the 21 subjects, the average of the threshold found by all the converged staircases is

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

2.75 ± 1.45, and 1.93 ± 1.10 for UCB. The averages of the distortion gain are higher compared to Porssut et al. [18] (1.43 ± 0.41), which might explain why four of their subjects had a detection threshold above 2. Indeed, some subjects can tolerate a distortion gain up to 5, which largely exceeds the maximal threshold (G = 2) Porssut et al. [18] had set for their experiment. To conclude, our results are consistent with the previous findings.

To identify the cause of the variability, the link between subjects' engagement level with 3D immersive application and the detection threshold is explored. A trend suggests that the more subjects play video games, the more the perception threshold is high. However this correlation didn't survive the correction for multiple comparison. This trend was only observed for the results obtained using the UCB algorithm. This trend may be observed for the first time because no other studies have used reinforcement learning to find the detection threshold. It gives an interesting hint for future experiment. It would be interesting to use and build a new questionnaire based on previous video game surveys [50], [51] to isolate more factors and understand the influence of video games on the detection threshold. Indeed, the main issue to calibrate a distortion is the high variability of the threshold. If we were able to link this variability with subjects'video games background, we might be able to adjust the threshold based only on this questionnaire.

In the debriefing with the subjects, they were asked at which moment they experienced a BIE and which factors helped them to notice the distortion. Globally, the subjects reported to have observed the "jump" when they moved the tennis ball from inside the blue sphere to follow the green sphere. They described feeling helped by an external force, or feeling that the avatar's hand is automatically attracted to the green ball. Some also reported having noticed the distortion when the movement of the avatar's hand is too steady compared to the movement of their physical hand during the movement. These remarks correspond to our expectations as they are coherent with the behaviour of the distortion function. They also underline the importance of controlling the distortion value to avoid altering subjects' experience.

There are several points of this experiment which could be improved upon. Instead of determining one threshold per staircase, the Point of Subjective Equality (PSE) could have been used as in Burns et al. [29] to find one single threshold for all the staircases and using a classical psychometric function [52], [53], [54], [55]. However, most of the time, these methods are used offline, as opposed to our approach, which was online. No experimentation has been performed with subjects to verify whether the threshold was consistent across several days. It could have been worthwhile to conduct the study during the span of several days with the same subjects and methods. Finally, the sample used for the experiment was not gender balanced. To our knowledge, no study has shown an effect of gender on movement distortion perception. However, the motion range of the elbow changes according to gender [56]. These anatomical differences might explain a part of our subjects' inter-variability. It would be interesting to control for this parameter in future studies.

7 Conclusion

In conclusion, reinforcement learning works well to find the detection threshold. It has the same accuracy and conservativeness as the staircase algorithm, but it is more robust. Overall, considering that robustness is the most important evaluation criterion, the reinforcement learning method used in this study (UCB) should be the method to retain for future applications.

In this experiment, subjects were asked to answer a question at the end of each trial to indicate if they had experienced a break in their sense of embodiment (Break in Embodiment (BiE)). Having explicit human feedback is very advantageous to update the reinforcement algorithms. However, it is very demanding and tiresome for a human to simultaneously stay focused on the virtual scene while continuously generating feedback. Therefore, developing an approach to obtain implicit feedback is highly relevant. Studies from [25], [26], [27], [45], [46], [57] demonstrated the possibility to use supervised learning models to detect the presence of an EEG signal generated by subjects' brain when they observe an error. In particular, Salazar-Gomez et al. [57] state that the accuracy of classifying secondary EEG - defined as the EEG signal provoked by the misclassification of the primary EEG signal - has a higher accuracy than solely using the primary EEG. However, none of previous works have used EEG with RL for assessing the sensitivity to embodiment distortions. Indeed, as opposed to a choice they consciously make among a predefined set, the detection threshold is highly individual and not consciously known to the subjects themselves. With the proposed approach, it may be possible to implicitly adjust the detection threshold within a VR experience. The used reinforcement learning method (UCB), is robust enough to find the maximum magnitude of distortion that each subject can tolerate without provoking a BIE in real-time for a VR application, and is compatible with implicit feedback like EEG signals.

At last, our study could be useful for exercising and motor rehabilitation. In particular it has been shown by Cameirao et al. [58] that task-oriented rehabilitation combined with the observation of virtual limbs facilitate the functional recovery of the arms. In this context, finding the correct detection threshold of the subjects would help them consider the distorted movement as their own, which could positively impact their recovery process. However, our study has been restricted to homogeneous (age, video game) healthy subjects. Therefore, for future clinical application, the study requires to be tested with a more diverse population and several parameters must be adapted like the duration (1 hour might be too long) and the type of movement. Besides, our method has only been evaluated for a predefined movement. Testing the performance of the algorithm when the movement is only partially known or entirely unknown would be necessary in such an application.

8 Supplementary Material

A video is provided with this paper. The dataset is hosted in a Zenodo repository (10.5281/zenodo.4298840), the Unity project is hosted in a Gitlab

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

repository (gitlab.epfl.ch/iig/research/adapting-virtual-embodiment-through-reinforcement-learning) and the RL algorithm is hosted in a Gitlab repository (gitlab.epfl.ch/iig/research/rlalgorithms).

ACKNOWLEDGMENTS

We would like to thank Ms. Francesca Gieruć and M. Mathias Delahaye for their invaluable contributions. This work has been supported by the SNFS project 'Immersive Embodied Interactions' with grant 200020_178790.

REFERENCES

[1] A. Henderson, N. Korner-Bitensky, and M. Levin, "Virtual reality in stroke rehabilitation: A systematic review of its effectiveness for upper limb motor recovery," pp. 52–61, mar 2007.

[2] S. V. Adamovich, G. G. Fluet, E. Tunik, and A. S. Merians, "Sensorimotor training in virtual reality: A review," pp. 29–44, jan 2009.

[3] H. Sveistrup, "Motor rehabilitation using virtual reality," pp. 1–8, dec 2004.

[4] B. R. Brewer, R. Klatzky, and Y. Matsuoka, "Visual feedback distortion in a robotic environment for hand rehabilitation," *Brain Research Bulletin*, vol. 75, no. 6, pp. 804–813, apr 2008.

[5] J. L. Patton, M. E. Stoykov, M. Kovic, and F. A. Mussa-Ivaldi, "Evaluation of robotic training forces that either enhance or reduce error in chronic hemiparetic stroke survivors," *Experimental Brain Research*, vol. 168, no. 3, pp. 368–383, jan 2006.

[6] J. L. Emken and D. J. Reinkensmeyer, "Robot-enhanced motor learning: Accelerating internal model formation during locomotion by transient dynamic amplification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 33–39, mar 2005.

[7] Y. Wei, P. Bajaj, R. Scheldt, and J. Patton, "Visual error augmentation for enhancing motor learning and rehabilitative relearning," in *Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics*, vol. 2005, 2005, pp. 505–510.

[8] J. L. Patton and F. A. Mussa-Ivaldi, "Robot-Assisted Adaptive Training: Custom Force Fields for Teaching Movement Patterns," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 636–646, apr 2004.

[9] C. D. Takahashi and D. J. Reinkensmeyer, "Hemiparetic stroke impairs anticipatory control of arm movement," *Exp Brain Res*, vol. 149, pp. 131–140, 2003.

[10] T. I. Nielsen, "Volition: a New Experimental Approach," *Scandinavian Journal of Psychology*, vol. 4, no. 1, pp. 225–230, mar 1963.

[11] O. A. Kannape, L. Schwabe, T. Tadi, and O. Blanke, "The limits of agency in walking humans," *Neuropsychologia*, vol. 48, no. 6, pp. 1628–1636, may 2010.

[12] J. Fasola, O. A. Kannape, M. Bouri, H. Bleuler, and O. Blanke, "Error Augmentation Improves Visuomotor Adaptation during a Full-Body Balance Task," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., jul 2019, pp. 1529–1533.

[13] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review," pp. 21–53, nov 2013.

[14] K. O'Brien, C. R. Crowell, and J. Schmiedeler, "Error augmentation feedback for lateral weight shifting," *Gait and Posture*, vol. 54, pp. 178–182, may 2017.

[15] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic, "Characterizing embodied interaction in First and Third Person Perspective viewpoints," in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, mar 2015, pp. 67–72.

[16] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence Teleoperators —& Virtual Environments*, vol. 21, 11 2012.

[17] E. Kokkinara and M. Slater, "Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion," *Perception*, vol. 43, no. 1, pp. 43–58, 2014.

[18] T. Porssut, B. Herbelln, and R. Boulic, "Reconciling being in-control vs. being helped for the execution of complex movements in VR," in *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., mar 2019, pp. 529–537.

[19] H. G. Debarba, B. Herbelin, and R. Boulic, "Self-Attribution of Distorted Reaching Movements in Immersive Virtual Reality," *Computers & Graphics*, no. and to be presented at SVR in 2018, 2018.

[20] S. Bovet, H. G. Debarba, B. Herbelin, E. Molla, and R. Boulic, "The critical role of self-contact for embodiment in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1428–1436, 2018.

[21] S. Esmaeili, B. Benda, and E. D. Ragan, "Detection of Scaled Hand Interactions in Virtual Reality: The Effects of Motion Direction and Task Complexity."

[22] N. Ogawa, T. Narumi, and M. Hirose, "Effect of Avatar Appearance on Detection Thresholds for Remapped Hand Movements," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[23] K. Matsumoto, E. Langbehn, T. Narumi, and F. Steinicke, "Detection Thresholds for Vertical Gains in VR and Drone-based Telepresence Systems."

[24] E. Burns, S. Razzaque, A. T. Panter, M. C. Whitton, M. R. McCallus, and F. P. Brooks, "The hand is more easily fooled than the eye: Users are more sensitive to visual interpenetration than to visual-proprioceptive discrepancy," in *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 1, feb 2006, pp. 1–15.

[25] I. n. Iturrate, J. Omedes, and L. Montesano, "Shared control of a robot using eeg-based feedback signals," in *Proceedings of the 2Nd Workshop on Machine Learning for Interactive Systems: Bridging the Gap Between Perception, Action and Communication*, ser. MLIS '13. New York, NY, USA: ACM, 2013, pp. 45–50.

[26] I. Iturrate, L. Montesano, and J. Minguez, "Robot reinforcement learning using eeg-based reward signals," 05 2010, pp. 4822–4829.

[27] T.-j. Luo, Y.-c. Fan, J.-t. Lv, and C. Zhou, "Deep reinforcement learning from error-related potentials via an eeg-based brain-computer interface," 12 2018, pp. 697–701.

[28] T. Meese, "Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations," *Perception —& psychophysics*, vol. 57, pp. 267–81, 04 1995.

[29] E. Burns, S. Razzaque, and M. Whitton, "MACBETH: Management of avatar conflict by employment of a technique hybrid," *Differences*, vol. 6, no. 2, pp. 11–20, 2007.

[30] D. A. Bowman and L. F. Hodges, "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments," in *Proceedings of the 1997 symposium on Interactive 3D graphics - SI3D '97*. New York, New York, USA: ACM Press, 1997, pp. 35–ff.

[31] E. Molla, H. Galvan Debarba, and R. Boulic, "Egocentric Mapping of Body Surface Constraints," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2017.

[32] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson, "Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. New York, New York, USA: ACM Press, 2016, pp. 1968–1979.

[33] L. Bréchet, R. Mange, B. Herbelin, Q. Theillaud, B. Gauthier, A. Serino, and O. Blanke, "First-person view of one's body in immersive virtual reality: Influence on episodic memory," *PLoS ONE*, vol. 14, no. 3, p. e0197763, mar 2019.

[34] P. Pozeg, E. Palluel, R. Ronchi, M. Solcà, A. W. Al-Khodairy, X. Jordan, A. Kassouha, and O. Blanke, "Virtual reality improves embodiment and neuropathic pain caused by spinal cord injury," *Neurology*, vol. 89, no. 18, pp. 1894–1903, oct 2017.

[35] E. Kokkinara, M. Slater, and J. López-Moliner, "The Effects of Visuomotor Calibration to the Perceived Space and Body, through Embodiment in Immersive Virtual Reality," *ACM Transactions on Applied Perception*, vol. 13, no. 1, pp. 1–22, oct 2015.

[36] H. G. Debarba, S. Perrin, and R. Boulic, "Perception of Redirected Pointing Precision in Immersive Virtual Reality," *IEEE Virtual Reality IEEE IEEE Virtual Reality*, no. March, 2018.

[37] A. Zenner and A. Kruqer, "Estimating detection thresholds for desktop-scale hand redirection in virtual reality," in *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., mar 2019, pp. 47–55.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3057797, IEEE Transactions on Visualization and Computer Graphics

[38] P. Marvit, M. Florentine, and S. Buus, "A comparison of psychophysical procedures for level-discrimination thresholds," *The Journal of the Acoustical Society of America*, vol. 113, no. 6, p. 3348, 2003.

[39] R. Madigan and D. Williams, "Maximum-likelihood psychometric procedures in two-alternative forced-choice: Evaluation and recommendations," Tech. Rep. 3, 1987.

[40] P. R. Jones, S. Kalwarowsky, O. J. Braddick, J. Atkinson, and M. Nardini, "Optimizing the rapid measurement of detection thresholds in infants," *Journal of Vision*, vol. 15, no. 11, pp. 2–2, aug 2015.

[41] J. L. Hall, "A procedure for detecting variability of psychophysical thresholds," *Journal of the Acoustical Society of America*, vol. 73, no. 2, pp. 663–667, 1983.

[42] B. Kollmeier, R. H. Gilkey, and U. K. Sieben, "Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model," *Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1852–1862, 1988.

[43] T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2801–2810, 2002.

[44] M. R. Leek, "Adaptive procedures in psychophysical research," *Perception and Psychophysics*, vol. 63, no. 8, pp. 1279–1292, 2001.

[45] S.-K. Kim, E. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning – using error-related potentials for real world human-robot interaction," *Scientific Reports*, vol. 7, 12 2017.

[46] R. Chavarriaga, A. Biasiucci, K. Forster, D. Roggen, G. Troster, and J. d. R. Millan, "Adaptation of hybrid human-computer interaction systems using eeg error-related potentials," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2010, pp. 4226–9, 08 2010.

[47] J. Wang, Y. Liu, and B. Li, "Reinforcement learning with perturbed rewards," *CoRR*, vol. abs/1810.01032, 2018.

[48] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[49] E. Kokkinara and M. Slater, "Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion," *Perception*, vol. 43, no. 1, pp. 43–58, 2014, pMID: 24689131.

[50] E. Altintas, Y. Karaca, T. Hullaert, and P. Tassi, "Sleep quality and video game playing: Effect of intensity of video game playing and mental health," *Psychiatry Research*, vol. 273, pp. 487–492, mar 2019.

[51] S. M. Grüsser, R. Thalemann, and M. D. Griffiths, "Excessive computer game playing: Evidence for addiction and aggression?" *Cyberpsychology and Behavior*, vol. 10, no. 2, pp. 290–292, 2007.

[52] M. C. Picardi and D. M. Green, "Comparison of three adaptive psychophysical procedures," *Journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1527–1533, 1982.

[53] J. J. Remus and L. M. Collins, "Comparison of adaptive psychometric procedures motivated by the Theory of Optimal Experiments: Simulated and experimental results," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 315–326, jan 2008.

[54] S. Amitay, A. Irwin, D. J. C. Hawkey, J. A. Cowan, and D. R. Moore, "A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1616–1625, mar 2006.

[55] J. A. Stillman, "A comparison of three adaptive psychophysical procedures using inexperienced listeners," *Perception & Psychophysics*, vol. 46, no. 4, pp. 345–350, 1989.

[56] A. Sawant, *Measurement of Joint Motion: A Guide to Goniometry, Third Edition*. F.A. Davis, 2004, vol. 56, no. 04.

[57] A. Salazar-Gomez, J. DelPreto, S. Gil, F. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," 05 2017, pp. 6570–6577.

[58] M. Cameirão, S. Bermúdez i Badia, E. Duarte, and P. Verschure, "Virtual reality based rehabilitation speeds up functional recovery of the upper extremities after stroke: A randomized controlled pilot study in the acute phase of stroke using the rehabilitation gaming system," *Restorative neurology and neuroscience*, vol. 29, pp. 287–98, 05 2011.

Thibault Porssut is a senior researcher in virtual reality and cognitive neuroscience in the laboratory of Dr. Boulic. He obtained his Master's degree in Mechanical Engineering and Industrial Engineering from l'Ecole des Arts et Métiers and a Research Master's degree in Digital Mock-up and Virtual Immersion from Institut Image in France. He did his Master's project at EVL in Chicago in partnership with Mechdyne. He obtained his PhD from EPFL School of Robotics, Control, and Intelligent Systems in 2020 for his research work on "Embodiment in Virtual Reality" under the supervision of Dr. Boulic and Dr. Herbelin.

Yawen Hou is currently pursuing a Master degree in Computer Science at the École Polytechnique Fédérale de Lausanne (EPFL, Switzerland). She participated to this research as part of her semester project in her third semester of Master under the supervision of Thibault Porssut and Ronan Boulic.

Olaf Blanke is founding director of the Center for Neuroprosthetics and holds the Bertarelli Foundation Chair in Cognitive Neuroprosthetics at the Ecole Polytechnique Fédérale de Lausanne (EPFL). He directs the Laboratory of Cognitive Neuroscience at EPFL and is Professor of Neurology at the University Hospital of Geneva.

Bruno Herbelin is senior researcher in virtual reality and cognitive neuroscience in the laboratory of Prof. O. Blanke at the École Polytechnique Fédérale de Lausanne (EPFL, Switzerland). From 2012 to 2019, he was deputy director of the EPFL Center for Neuroprosthetics. From 2005 to 2009, he was Assistant Professor at Medialogy Department of Aalborg University, Denmark. He obtained his PhD from EPFL School of Computer and Communications in 2005 for his research work on virtual reality exposure therapy under the supervision of Prof. D. Thalmann (Virtual Reality Laboratory).

Ronan Boulic CS PhD University of Rennes (1986), CS Habilitation University of Grenoble (1995). He joined EPFL LIG-VRLAB in december 1989 after a postdoc in University of Montreal. He is leading the Immersive Interaction research Group EPFL-IIG since 2011. His research interests include Virtual Reality, Embodied 3D interactions, motion capture, modeling, and synthesis for virtual humans and robots.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/