HandAvatar: Embodying Non-Humanoid Virtual Avatars through Hands

Yu Jiang*
Tsinghua University
Beijing, China
jiangyu0927@gmail.com

Zhipeng Li*
Tsinghua University
Beijing, China
lzp20@mails.tsinghua.edu.cn

Mufei He
Tsinghua University
Beijing, China
mufei1211@gmail.com

David Lindlbauer
Carnegie Mellon University
Pittsburgh, Pennsylvania, United States
davidlindlbauer@cmu.edu

Yukang Yan†
Carnegie Mellon University
Pittsburgh, Pennsylvania, United States
yanyukanglwy@gmail.com

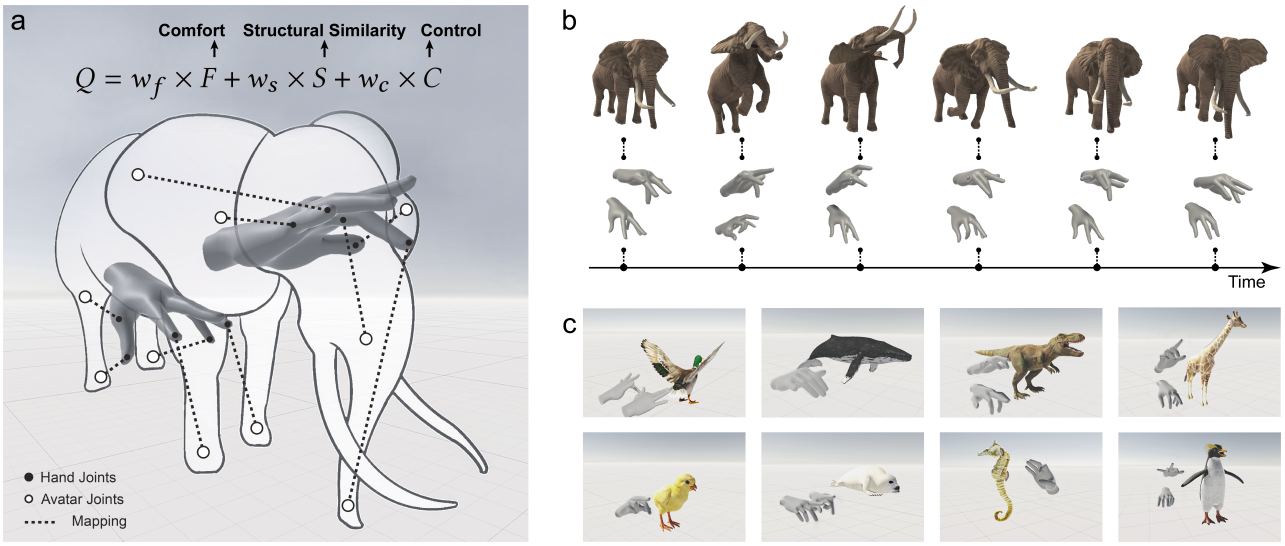


Figure 1: HandAvatar (a) optimizes hand-to-avatar joint-to-joint mappings based on user-prioritized factors to (b) enable embodied, precise, and flexible control over non-humanoid avatars in real-time. (c) The optimization is extendable to a wide range of non-humanoid avatars and thereby unlocks new user-as-non-humanoids interaction experiences.

ABSTRACT

We propose HandAvatar to enable users to embody non-humanoid avatars using their hands. HandAvatar leverages the high dexterity and coordination of users' hands to control virtual avatars, enabled through our novel approach for automatically-generated joint-to-joint mappings. We contribute an observation study to understand users' preferences on hand-to-avatar mappings on eight avatars. Leveraging insights from the study, we present an automated approach that generates mappings between users' hands and arbitrary virtual avatars by jointly optimizing control precision, structural similarity, and comfort. We evaluated HandAvatar on static posing, dynamic animation, and creative exploration tasks. Results indicate that HandAvatar enables more precise control, requires less physical effort, and brings comparable embodiment compared to a state-of-the-art body-to-avatar control method. We demonstrate HandAvatar's potential with applications including non-humanoid avatar based social interaction in VR, 3D animation composition, and VR scene design with physical proxies. We believe that HandAvatar unlocks new interaction opportunities, especially for usage in Virtual Reality, by letting users become the avatar in applications including virtual social interaction, animation, gaming, or education.

*Both authors contributed equally to this research.
†Corresponding author.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581027

CCS CONCEPTS

• **Human-centered computing** → *Gestural input*; **Mixed / augmented reality**.

KEYWORDS

virtual avatar, embodiment, Mixed Reality, gestural interaction

ACM Reference Format:
Yu Jiang, Zhipeng Li, Mufei He, David Lindlbauer, and Yukang Yan. 2023. HandAvatar: Embodying Non-Humanoid Virtual Avatars through Hands. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3544548.3581027

1 INTRODUCTION

Virtual avatars are used for a wide range of applications such as games, instructional videos or animations. Typically, avatars are animated using keyframing software on 2D desktops by trained experts and animators. To enable precise and natural control over virtual avatars with less training, prior approaches leveraged users' body motion in Virtual Reality (VR) settings to create a one-to-one mapping between users and virtual avatars.

For humanoid avatars, this approach works well as users' body parts can be naturally mapped to the corresponding parts on the avatars since they are topologically and structurally close. However, in various VR applications (e. g., gaming), there also exists a need for users to embody non-humanoid avatars (e. g., a monster spider) to perform movements and tasks (e. g., fighting). Past research has proposed mapping the user's full body to non-humanoid avatars [5] or using hand-held controllers to manipulate the key joints of avatars [7]. However, due to the structural differences and limited degrees of freedom, neither of them allows for precise real-time control of non-humanoid avatars, especially on those with complex structures. More specifically, Kin$\hat{E}$tre [5] requires users to engage their full body to animate a character, which might not be desirable for creating fine-grained animations or in constraint environments. Creature Teacher [7] enables users to record movements but does not create a direct mapping. We believe that a direct mapping technique that enables users to flexibly control non-humanoid avatars without having to engage their whole body would be preferable in a variety of scenarios.

To address this challenge, we propose HandAvatar, which enables users to control and embody non-humanoid avatars with their hands. We leverage the high dexterity and coordination of users' hands to provide flexible control and exploit the conceptual similarity between our approach and hand puppetry to simplify learning. The core contribution of HandAvatar is a novel optimization-based method that automatically creates mappings between users' hand joints and the avatar key joints by maximizing the precision, perceived intuitiveness, and comfort of controlling the avatars.

We first conduct a user study to understand the key factors in designing a hand-to-avatar joint-joint mapping. We invited twelve participants to design their own hand-to-avatar mappings for eight animal avatars varying in degree of freedom (DoF) and structural complexity. Eight participants were then re-invited to vote on their three favorite designs. Results show that control, intuition, and comfort are the most prioritized factors; participants also have consistent top-voted mappings for most of the avatars which outperform their own mappings on some or all of the criteria. These results highlight the need and feasibility for the automatic creation of optimal mappings, as users had a consistent preference for good mappings and could not design their favorite mappings on their own. We leverage the insights from this study to design the proposed optimization algorithm as prior constraints (e. g., balancing the usage of left and right hand joints).

Based on the results, we contribute an optimization-based algorithm for the generation of hand-to-avatar mappings. The algorithm takes as input the prior knowledge of hand anatomy and biomechanics, and the designer-defined avatar parameters and interaction requirements (e.g. uni- or bi-manual control). It measures control precision, structural similarity as a proxy for "intuitiveness", and comfort with range of motion overlaps, relative joint distances, and Mayer and Katzakis [26]'s established method which measures the deviation from range of rest postures respectively. The mapping optimization is formulated as an integer program problem of assigning hand joints to avatar joints. By comparing HandAvatar-optimized, user-designed, and user-voted mappings, we verify HandAvatar's ability to generate overall better mappings and thus the validity of the objective function of the optimization.

We evaluated the control precision and user experience of HandAvatar with static poses and an animation-imitation task on four avatars, and creativity-themed animation performances on two avatars respectively. We compare our method to a baseline approach similar to Kin$\hat{E}$tre [5], which leverages the user's full body motion for controlling non-humanoid avatars, as is extensively studied in previous works [37, 49]. We further evaluated the effectiveness of the optimization algorithm by comparing the mappings generated by HandAvatar with the user-voted mappings for four avatars in the creativity-themed animation task. Results showed that compared to the body-based approach, HandAvatar allows significantly more accurate control by reducing 40% joint deviations for static poses and 25% joint deviations for animated motions. HandAvatar also provides higher comfort, less physical load, comparable embodiment, and produces more satisfying animations compared to the baselines. The algorithm-optimized mappings were preferred (27 vs. 21) by the participants compared to user-voted mappings. To demonstrate the feasibility and scalability of the proposed method, we implemented three applications including non-humanoid avatar based social interaction in VR, 3D animation creation, and 3D VR scene design with physical proxies. We believe that our method enables users more fine-grained control over complex non-humanoid avatars while requiring less interaction space by leveraging the dexterity of users' hands. In summary, we make the following contributions:

- We present HandAvatar, which enables users to control and embody non-humanoid avatars with their hands, with hand-to-avatar joint mappings optimized for control precision, intuition, and comfort.
- We reveal users' main considerations while designing hand-to-avatar joint-joint mappings through an observation study, which informs the implementation of HandAvatar.
- We present an optimization algorithm which uses hand biomechanics as prior knowledge and takes avatar parameters and interaction requirements as input to create optimized joint-to-joint mappings.

- We evaluate HandAvatar with a body-based baseline method on static poses, animation imitation, and creative animation performance tasks. Results show that HandAvatar provides significantly higher control precision and overall better user experience. We verify the effectiveness of the optimization algorithm by comparing the mappings generated by HandAvatar with user-voted mappings.

2 RELATED WORK

We built on previous works in embodiment in VR and AR, digital puppetry, and hand-based interaction.

2.1 Embodiment in VR and AR

The sense of embodiment (SoE) in a virtual environment refers to a person's ownership over a virtual body[25]. One common approach to elicit SoE is to use virtual avatars to mirror the users' body movement [39] which has been proven effective[24, 33]. Various factors, including the point of view, motion synchronization, and avatar appearance, affect the level of users' SoE. Specifically, users gain stronger SoE in the first point of view than in the third point of view [35]. SoE can be enhanced through a high correlation of multi-sensory inputs between users' physical body and the virtual avatar [30, 35, 41]. The level of motion synchronization, which was tested to be highly relevant to the level of sense of control [13, 18, 45], massively increases the Rubber Hand Illusion (RHI) in which the rubber hands are regarded as one's own hands [3]. SoE is also influenced by the morphological similarity between the avatar and the user body [45]. Mirroring users' motion onto a humanoid avatar with identical body structures typically helps gain SoE [13]. Beyond human-like virtual avatars, SoE can also be gained via mirroring the motions onto non-humanoid avatars [22].

While technologies like full-body tracking [40] and inverse kinematic methods [51] are developed to improve SoE in controlling humanoid avatars, there is still a lack of research on embodying non-humanoid avatars. We believe that animating non-humanoid avatars would free people's imagination, help them better understand the behaviors of certain creatures, and explore a novel zoological field in VR application market [5, 14, 20]. Therefore, we propose to let users embody a virtual non-humanoid avatar in our work echoing the rising interest in controlling non-humanoid avatar.

2.2 Digital puppetry

Digital puppetry [17, 44] is a convenient and intuitive avatar control method leveraged in animation, storytelling, rehabilitation, and improvisation [6, 16, 20]. Compared to traditional mouse and keyboard based avatar control methods, digital puppetry introduces a direct interaction experience between the user and the target without an intermediary interface [9, 38]. A physical object or human body can serve as the puppet to drive virtual avatars. The mapping between the user and the avatar is the key factor to the control, the expressiveness of the manipulation, and the level of directness or SoE over the avatar [17].

Three main types of puppets have been leveraged in digital puppetry. Hand-held controller based manipulation [7, 31, 46] is easy to learn with the mapping typically determined at runtime by users manually selecting an avatar or its specific body parts that they wish to control. Due to controllers' limited degrees of freedom, rough control over an avatar's locomotion or fine control over a few body parts at one time is supported, limiting the SoE gained from the avatars. Full-body control [6, 10, 32, 42] is another common approach for controlling humanoid avatars. Due to the avatar and the user's high structural similarity, the corresponding joints are intuitively mapped to provide SoE. The users' movement is either directly mapped onto the avatar [6, 42] or retargeted to augment the users' capabilities in VR [1, 19, 48]. A number of studies [5, 37, 49] explore full-body based puppetry on non-humanoid avatars. Among these, Yamane et al. [49] and Seol et al. [37] transform user motion into avatar animations based on mappings synthesized from a set of user-defined pairs of user and avatar poses. Leite and Orvalho [16] creates arbitrary and indirect joint-to-joint mappings to facilitate expressive control. Kin$\hat{E}$tre [5] allows users to specify joint-to-joint mappings by metaphorically "attaching" the overlapped body and avatar joints. However, these approaches require users to perform large-scale movements which may lead to high physical load and result in social awkwardness. The difference in body structures between the human body and the non-humanoid avatars also leads to unsatisfactory mappings that only enable partial or unintuitive control over the avatars.

In HandAvatar, we thus built on works that used hands as the puppet for their high dexterity and coordination to control morphologically diverse non-humanoid avatars.

2.3 Hand-based puppetry

The human hand is a dexterous and complex biomechanical device containing a maximum of 29 degrees of freedoms [4, 15, 43], making it a capable input device for different media such as VR and AR [34, 50]. The hands have been leveraged for digital puppetry with different approaches, namely gestural interaction, marionette, and glove puppetry.

Gestural interaction approaches rely on a pre-defined gesture-to-motion library and trigger an avatar motion upon recognizing a hand gesture. Such control method has been used in sand animation [12], creative storytelling for children [20], and game-based wrist rehabilitation [6]. Despite its usability in a wide range of applications, gestural interaction approaches lack control flexibility and freedom. Approaches that use the marionette metaphor (string puppets) map the position of two paired points and are commonly used in controlling 2D avatars. 2D images are animated by dragging control handles in the avatar's mesh either on a tablet [10, 28] or midair [29] through user-designed finger-to-handle mappings. Lockwood and Singh [21] tracked the touch points of the fingers and transformed them into a human's walking behavior. Marionette approaches require uses to mentally transform the desired animation into the expected destination position to control the avatars. This results in indirect control which hurts SoE.

In glove puppetry methods, the hands are embedded within the avatar and directly drive the avatar by mirroring hand motions onto different avatar body parts' movements. Way et al. [47] and Luo et al. [23] leveraged glove-puppetry based method to control only the arms of the humanoid avatars with self-defined mappings. Oshita et al. [32] provided finer control and used 10 fingers to control a humanoid's body parts. LEite and Orvalho [17] went

further to optimize expressive control and the fingers' range of motion when assigning fingers to control an avatar's facial expressions. By directly mirroring hand motion to avatar movement, glove puppetry-based approaches enable direct control. However, current glove puppetry methods have designer-designed or arbitrary hand-to-avatar mappings that take no or very limited factors into consideration. As a result, the proposed methods are not extendable to other avatars and are sub-optimal in terms of control and user experience. We thus set out to investigate crucial factors of precise and embodied hand-to-avatar control mappings and present an algorithm that automates the process of defining the optimal joint-to-joint mappings.

3 OBSERVATION STUDY

We first conducted an observation study to understand the main influencing factors for users when they create hand-to-avatar joint mappings. We invited twelve participants to design their own hand-to-avatar mappings for eight non-humanoid avatars, and nine of them to vote for their top three favorite mappings. From the results, we distill a taxonomy of user-designed mappings and provide insights into users' considerations when designing hand mappings. The findings also guided the development of our proposed mapping optimization algorithm.

3.1 Hand-to-avatar mapping

A hand-to-avatar mapping is defined by one-to-one paired joints on the hands and the avatar, and a default avatar pose and its corresponding hand pose. The mapping specifies how the avatar is controlled by hands. Each one-to-one joint pair consists of a **control joint** on the avatar which would move together with a mapped hand joint by sharing the same local joint rotation. The rest of the joints on the avatar that are not controlled would be **rigid joints** whose connecting limbs cannot perform relative rotation. The default avatar and hand poses specify the hand poses to control the avatar with the mapping. HandAvatar aims to let users embody the non-humanoid avatar in which the users observing the avatar's movement and associating it with their hand movement is crucial. As a result, we decide to place the avatar in front of the user below their eye level for the convenience of observation.

Table 1: Habitats, topological structures, and motion methods covered by the eight avatars tested in the study.

Avatar	Habitat	Topological structures	Motion
Toucan	airborne	1 mouth, 2 wings, 2 legs	walk, fly
Dragonfly	airborne	4 wings, 6 legs	walk, fly
Salmon	aquatic	1 mouth, 2 fins, 1 tail	swim
Frog	aquatic	1 mouth, 4 legs	swim, jump
Elephant	land	1 mouth, 1 nose, 2 ears, 4 legs, 1 tail	walk
Penguin	land	1 mouth, 2 wings, 2 legs	walk, swim
Spider	land	2 tentacles, 8 legs	walk
Snake	land	1 mouth, 1 trunk	slither

3.2 Design

The observation study included a **design** phase and a **vote** phase to understand how participants design their own hand-to-avatar mappings and what are participants' criteria for good mappings.

In the **design** phase, we asked the participants to design hand-to-avatar mappings for eight avatars. The animal avatars were chosen such that they cover different habitats, body structures, and motion methods (detailed features summarized in Table 1). The tested avatars, as well as other avatars used later in the paper, were sourced online from the Unity Asset Store [1]. Given an avatar at its initial pose, the task was to perform a hand pose that best matches the avatar pose and then map the key joints on the avatar to corresponding hand joints. The instruction for the design goal is to optimize the control. After designing a mapping, we asked the participants to experience the design with 4 additional avatar poses as the test cases. The avatar test poses were randomly captured from the avatar's built-in animations. The avatars were tested in a randomized order for each user. We thus collected 8 avatars ×12 participants = 96 mappings.

In the **vote** phase, we invited the same group of participants who participated in the **design** phase to vote for the top three hand-to-avatar mappings designed by themselves and others. We split the participants into 2 groups to reduce the required time and effort. For each avatar, the participants observed 6 mappings designed by different participants (including their own) in their group. Each participant then voted top 3 different mappings for that avatar. Each participant thus observed 8 avatars × 6 participants = 48 mappings. For each mapping, the participants observed the joint-to-joint mapping and the 4 hand poses recorded for the avatar test poses.

3.3 Participants

We recruited 12 participants (5 males, 7 females) from a local university with an average age of 21.75 years ($SD = 0.75$) for the **design** phase. All participants were right-handed. The self-reported familiarity with VR and with puppetry averaged at 4.63 ($SD = 1.51$) and 3.08 ($SD = 1.24$), respectively, on a 7-point Likert-type scale from 1 (not at all familiar) to 7 (very familiar). The hand width and length averaged at 9.52 cm ($SD = 0.77$) and 17.85 cm ($SD = 1.96$), respectively. For the **vote** phase, we re-invited 8 out of 12 participants who took part in the **design** phase and were available. The 8 participants included 2 males and 6 females with an average age of 21.63 ($SD = 0.92$).

3.4 Procedure

The VR platforms developed for both phases are shown in Figure 2a.

Design phase: After introducing the purpose of the study and collecting demographic information, the participants sat down and wore the VR headset. A walk-through of the VR platform for the study was provided. For each avatar, the participants were first shown clips of the avatar's animation to observe its body structures and to understand its movement patterns. Then an initial static avatar pose is shown for which the user needs to perform the corresponding controlling hand pose to be recorded. Based on the initial avatar pose and the recorded hand pose, the participants

[1] https://assetstore.unity.com/

used the controller to select **control joints** on the avatar and map them to the corresponding hand joints.

After the participants confirmed the designed mapping, they were shown four static avatar test poses. Based on their mappings, they were asked to give a hand pose that could realize each of the given test poses. The participants could experience and evaluate their own mappings during this process. After all the hand poses are recorded, the participants were asked to rate their own mappings on controlling the tested avatar in terms of comfort, embodiment, control, and realness. The metrics were rated based on a 7-point Likert scale (1: strongly disagree, 4: neutral, 7: strongly agree). After all the avatars are tested, the participants were asked about their categorical strategies and design concepts for different types of avatars, as well as the challenges that they met.

Vote phase: The same group of participants was invited to the vote phase after two weeks. For each avatar, participants were presented with 6 mapping designs represented by the initial and test avatar poses and their corresponding hand poses. They were asked to imitate the design's hand pose to understand the mapping. Based on their observation, the participants voted for the top three mappings. After the voting task for each avatar, the participants were asked about the reasons and criteria for making those choices, and their strategies and reflections. The interviews were audio-recorded for future analysis with the participants' consent.

The **design** phase lasted two and a half hours and the **vote** phase lasted approximately one and a half hours to complete. Participants took 5-minute breaks every 40 minutes to avoid fatigue. Each participant was compensated with 50 US dollars for the **design** phase and 25 US dollars for the **vote** phase.

3.5 Apparatus

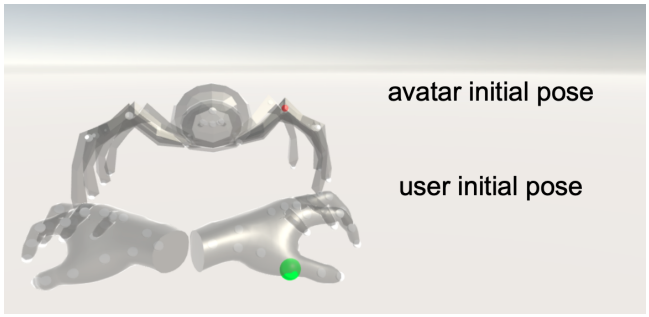
The VR study platforms for both phases were developed in Unity 2021 for the Meta Quest 2 headset, running on a gaming computer (Intel Core i7 CPU, NVIDIA GeForce RTX 3080 GPU). The participants performed the study in a seated position in a quiet lab environment.

3.6 Results

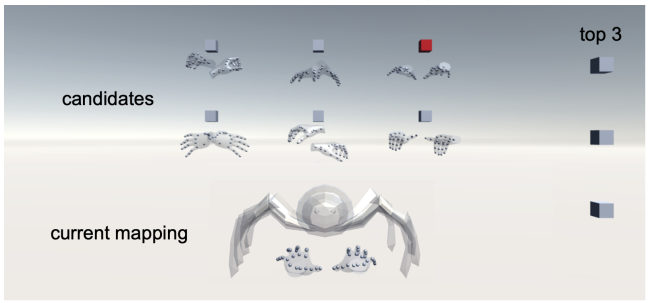
For each mapping designed by one participant for an avatar, we obtained the subjective **comfort**, **embodiment**, **control**, **realness** scores. We transcribed and qualitatively coded all the interview recordings collected. Three researchers worked separately on the open coding process and reached a consensus through discussions. Based on the above data, we present the findings on the taxonomy of participant-designed mappings, the participant preferences on mappings, and the participant preferences on avatars.

3.6.1 Taxonomy of participant-designed mappings. The taxonomy is generated based on interview recordings from the **design** phase. The mappings are categorized into four dimensions, which were extracted based on the repeated examination of the codebook and discussions among the researchers about participants' mapped gestures and their spoken intents. We detail the considerations of participants in each dimension.

Nature. Participants' motivations in designing the mappings are mostly *form-driven* and *functionality-driven*. *Form-driven* mappings imitate the structure of the avatars with the hand poses. The spatial relationship between different avatar body parts is projected on the corresponding joints on the hands. *Functionality-driven* mappings focus on functionalities or motions that the avatar can perform and take less count of the overall spatial structure. This involves mapping the key joints needed to perform a motion to the hand joints that provide the best control for the motion. In comparison, *form-driven* mappings provide intuitive control ("*P6: Avatar-alike mappings don't require mentally translating avatar motion to hand motion for different functionalities.*"); while *functionality-driven* mappings provide optimized control tailored to each task by "*tearing the avatars apart*" (P8).



(a) Design phase: users perform a hand pose to match the avatar initial pose and select joint-to-joint mappings.



(b) Vote phase: users select the red cube to view one of the candidate mappings, shown by hand and avatar pose pairs, and vote their 3 favorite mappings.

Figure 2: The VR platforms for the observation study.

Control. The control dimension explains the level of control over the avatars, which is reflected by the proportions of control joints on the avatar labeled by the participants. *Full control* mappings control over 80% of the avatar joints, leaving only a few insignificant body parts behind. *Partial control* mappings control less than 80% of the avatar joints. Though the participants were told to design for the finest control, they came up with different reasons for adopting *partial control* designs. It becomes necessary when the number or the spatial layout of the avatar key points is beyond the control of two hands (e.g., for the snake avatar). *Partial control* also occurs when participants hope to prioritize the control on main body parts and intentionally leave out peripheral body parts (e.g., fins of a salmon).

Balance. *Balanced* mappings assign both hands with roughly the same number of avatar joints to control, with both hands having more than $\frac{1}{3}$ of the mapped hand joints. *Left/Right hand dominated*

Dimension		Classification		Description
Mapping	Nature	Form-driven		*Hand(s) imitate the the avatar's shape*
		Functionality-driven		*Hand parts separately realize the avatar's functions*
	Control	Full control		$\geq$ 80% *avatar joints are control joints*
		Partial control		$<$ 80% *avatar joints are control joints*
	Balance	Balanced		*Both hands have* $\geq \frac{1}{3}$ *of all mapped hand joints*
		Left hand dominated		*The left hand have* $\geq \frac{2}{3}$ *of all mapped hand joints*
		Right hand dominated		*The right hand have* $\geq \frac{2}{3}$ *of all mapped hand joints*
Gesture	Hand	Uni-manual		*Control an avatar with one hand*
		Left-right	symmetric	*Hands' relative position is left-right with the same gesture*
			asymmetric	*Hands' relative position is left-right with different gestures*
		Up-down		*Hands' relative position is up-down*
		Front-back		*Hands' relative position is front-back*

Table 2: Dimensions, classification, and criteria in the taxonomy of user-designed avatar-to-hand joint mappings.

mappings rely on either the left or the right hand for major control while the other hand handles the peripheral parts. Participants commented that *balanced* mappings are helpful in reducing mental load during manipulation as the control tasks are equally distributed between the hands and *left/right hand dominated* mappings are used mostly on avatars with simple structures.

Hand. *Uni-manual* manipulation controls the avatar through one hand, which was reported by some participants to be "*simpler and require less hand movement.*" (P12) *Left-right*, *up-down*, and *front-back* describes the spatial relationship between two hands if bi-manual manipulation methods are designed. In the *left-right* category, the *symmetry* sub-dimension describes whether the two hands have the same gestures. Participants commented that "*symmetric gestures makes me comfortable, and think less while imitating the animals' poses*".(P1)

Figure 3 shows the breakdown of 96 hand-to-avatar joint mappings. Within the four dimension, the most common classifications of the mappings were *form-driven*, *full control*, *balanced*, *left-right symmetric*.

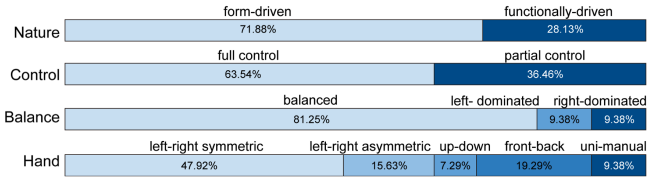


Figure 3: Breakdown of user-designed hand-to-avatar joint mappings on the dimensions nature of design, degree of control, balance of joints between hands, and hand gestures used.

3.6.2 Participants' preferences on mappings. 8 out of 12 available participants took part in the **vote** phase. We thus collected 8 avatars $\times$ 8 participants $\times$ 3 votes = 192 votes in total, based on which we summarize the participants' preferences on mappings. This also informs us of the mapping optimization goals.

Participants did not always prefer their own mapping designs. Participants only voted 13/64 times for their own designs as the top 1 and 38/192 times as the top 3. Participants explained that they preferred others' designs when "*they have clear and intuitive mapping logic and really easy to learn*"(P8), or they are "*quite inspiring because they solved the problems I was struggling on.*"(P7)

Participants evaluated a mapping based on its control, intuition, and comfort with varying priorities. Most participants favored a balance between control, intuition, and comfort. For **control**, participants preferred *full control* as shown in the voting results: 126 out of 192 voted top 3 mappings and 41 out of 64 voted top 1 mappings were in the *full control* category. Control-focused participants were willing to sacrifice intuition to optimize control as "*some avatars are too complicated for mimicking, and I would rather choose to separate tasks into two hands over the comfortable or form-alike mappings*". (P12)

Participants gained better **intuition** from *form-driven* mappings as they can intuitively map the control joints to its closet avatar joint and thus "*get the idea from the first sight.*" (P8) 149 out of 192 voted top 3 designs for 8 avatars are form-driven for their high legibility and learnability.

Participants (P6, P8, P11) found **comfort** important for long-term uses and they determined comfort based on the relative hand position, wrist rotation, and the physical effort required for performing motions with the mappings.

Participants have consensus on good mappings. The top-voted mappings gained 5 votes on average (SD = 0.76). After clustering mappings that are almost identical with minor differences, the voting results show that participants have one or two consistent mapping preference(s) for all avatars except for penguin. Elephant and salmon each has two equally top-voted designs which are different in at least one of the taxonomy dimensions but are equally satisfying for the participants. Control prioritized and intuition/comfort prioritized participants favored a task-driven and a form-driven method respectively. Penguin, however, didn't show a unified design due to its "structural simplicity" (P3), enabling participants to satisfy control, intuition, and comfort with different approaches.

3.6.3 User preference for avatars. We performed Friedman tests with pairwise Wilcoxon signed-rank post hoc tests on participants' subjective scores. The Friedman test showed statistical significance on **comfort** ($\chi^2(7,\ N = 12) = 18.95,\ p < 0.05$), **control**

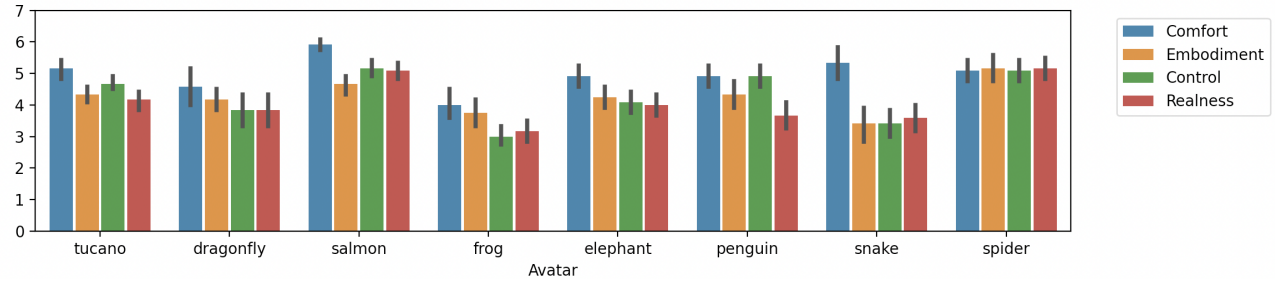


Figure 4: Subjective ratings on comfort, embodiment, control, and realness with regard to eight avatars to control and embody. The error bars represent the standard errors.

($\chi^2(7,\ N = 12) = 33.25,\ p < .001$), and **realness** ($\chi^2(7,\ N = 12) = 24.33,\ p < 0.05$).

We summarize and discuss our findings below which inform us what avatars are suitable for hand manipulation. Please refer to the full statistical analysis result in the supplementary materials.

Participants favor avatars that are structurally simple or similar to the hands. Salmon is favored for its structural simplicity. Salmon has a significantly high **comfort** score ($M = 5.92,\ SD = 0.67$) and **subjective realness** score ($M = 5.08,\ SD = 0.99$) than most of the other avatars (all $p < 0.05$) except snake and frog. Spider is favored for its structural similarity with human hands, as indicated by the participants. (P2, P9)

Participants found avatars with abstract body shapes or overly complex body structures hard to manipulate. Participants mentioned that they had to use a functionality-driven method to design a mapping for snake because it is "abstract in shape and structurally different from human hands". Its unique slithering motion patterns made it hard to control which is also indicated in the subjective result: snake has a lower **control** score ($M = 3.42,\ SD = 0.02$) than salmon, spider, toucan, and penguin ($all\ p < .05$). Since human hands are incapable of mimicking the slithering pattern, snake's **embodiment** score ($M = 3.42,\ SD = 1.88$) is significantly lower than salmon ($M = 4.67,\ SD = 0.99,\ p < 0.05$) and spider ($M = 5.17,\ SD = 1.53,\ p < 0.05$).

Table 3: Description of the parameters that the optimization algorithm takes as input. Parameters input by application designers are denoted by *.

Hand parameters	
Parameter	Description
$J^h = \{J^l, J^r\}$	Set of left and right hand joints.
$J^l = \{j_1^l, \ldots, j_m^l\}$	Left hand joints.
$J^r = \{j_1^r, \ldots, j_m^r\}$	Right hand joints.
$G = \{g_1, \ldots, g_m\}$	Set of hand poses.
$R^h_{i,k} \in \{0, 1\}$	Whether joint i is senior than joint k .
RoM_i^h	The range of motion of hand joint i
* $U \in \{0, 1\}$	If uni-manual control is used.
* $H^l,\ H^r \in \mathbb{R}^3$	Positions of the hand wrists.

*** Avatar parameters**	
Parameter	Description
$P = \{p_1, \ldots, p_n\}$	Set of poses of avatar a.
$R^a_{i,k} \in \{0, 1\}$	Whether joint i is senior than joint k.
RoM_i^a	The range of motion of avatar a's joint i.
$J^a = \{j_1^a, \ldots, j_n^a\}$	Set of avatar a's control joints.
N	Number of control joints.

4 ALGORITHM

4.1 Overview

Given a non-humanoid avatar, HandAvatar generates a one-to-one joint mapping M and a pair of hand pose and avatar pose (g, p) optimizing three main objectives considering the control precision, structural similarity between pose pairs, and comfort. M determines how avatar key joints are mapped to hand joints such that the avatar joints mirror the mapped hand joints' movement. (g, p) fixes the avatar pose to be p when the user's hand pose is g. If and only if when both M and (g, p) are decided, we can calculate the avatar pose given any hand pose as input. Our approach takes inputs from application designers into account, including the avatar model with labeled control joints, motion animations demonstrating the joints' range of motion, and interaction requirements on hands (e.g., single- or dual-handed control). Prior knowledge of the hands' biomechanics and a series of heuristic constraints obtained in Section 3 are also considered. HandAvatar simultaneously optimizes the objectives control precision, structural similarity, and comfort level, which were valued the most by users. By comparing HandAvatar-generated, user-generated, and user-voted mappings, we verify the effectiveness of the proposed algorithm and its objective function calculation.

4.2 Inputs and outputs

The algorithm takes as input prior knowledge of hand motion, interaction requirements, and avatar parameters. Then it outputs the joint-to-joint mapping M and the pair of hand pose and avatar pose (g, p).

4.2.1 Hand parameters. Pre-processed parameters factorizing the anatomy and the biomechanics of the hands and designers' hand interaction requirements are included as input. All the joints on

the hands $J^{l/r}$, with 24 joints on each hand are considered as candidates that an avatar control point can map to. Based on the hand joints' hierarchical relationship extracted from the Oculus hand hierarchy [2], we define a seniority variable $R^h_{i,k}$ which checks if joint j^h_i is more senior than joint j^h_k in the hierarchy (i.e., j^h_i is more senior if j^h_i is j^h_k's parent/grandparent). The range of motion of each hand joint j^h_i, $RoM^h_i = (\phi_{min}, \phi_{max}, \theta_{min}, \theta_{max})$, was recorded from [8]. Since it is nearly impossible to cover the extremely large space of hand poses, we sampled a set of 10000 hand poses G with $j^h_{i\in\{1,\,m\}} = random(RoM^j_i)$ for the joints on both hands, which form the set of starting hand pose candidates.

For interaction requirements, the algorithm allows the designer to suggest whether one or two hand(s) are used for control U, and the positions of the hand wrists H^l, H^r which indicate whether the relative spatial relationship between the two hands is left-right, up-down, or front-back.

4.2.2 Avatar parameters. In a wide range of applications (e.g., animation creation), not all the structural joints of the virtual avatar needs to be controlled by users, instead they could be taken care of by automated algorithms (e.g., IK [2]). HandAvatar allows application designers to specify the key joints on the avatar that they wish users to control, which is denoted as J^a. The algorithm also takes as input the structural hierarchy of the avatar from which we define the seniority variable $R^a_{i,k}$. Motion animations of the avatar are also required. From the input animations, we segment a set of possible static poses P, specified by the joint positions in each pose, for the avatar a. We also interpret the range of motion of all joints from the animation in spherical coordinates such that for a joint i, its range of motion RoM^a_i records the minimum and maximum angles joint i reaches in the animation.

4.2.3 Outputs. The algorithm outputs a mapping M which is a set of joint-to-joint pairs $\{(j^h_i, j^a_k)\}$. Each pair (j^h_i, j^a_k) indicates that the movement of avatar key joint j^h_k mirrors the movement of hand joint j^h_i, or, hand joint j^h_i controls avatar joint j^h_k. As M only determines how hand joints and avatar key joints correlate in relative movements, we still need to decide on a pair of reference hand pose and avatar pose. Thus the algorithm also outputs a pair of hand pose and avatar pose (g, p) as the reference point on which the mapping M can be applied. With the mapping and the pose pair, given a hand pose, we can find its corresponding avatar pose by first calculating the change in relative rotation on each joint compared to the reference hand pose, then applying the same change on the mapped key joints on the reference avatar pose. The corresponding hand pose with a given avatar pose can be similarly calculated. The reference pair also informs users of the hand poses to start with to embody the avatar.

4.3 Constraints

We describe a hand-to-avatar joint-to-joint mapping with

$$u_{j^a_i j^h_k} = \begin{cases} 1 & j^a_i \text{ is mapped to } j^h_k \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

[2] https://developer.oculus.com/documentation/unity/unity-handtracking/

The algorithm considers a set of constraints when generating a joint mapping, which was obtained based on the findings in Section 3.

A *one-to-one* joint mapping, in which every avatar control joint is mapped to one hand joint and every hand joint is used for at most 1 time, is required. We thus have that for every avatar joint j^a_i, and for every hand joint j^h_i:

$$\sum_{j^h_k \in J^h} u_{j^a_i j^h_k} = 1 \tag{2}$$

$$\sum_{j^a_k \in J^a} u_{j^h_i j^a_k} \leq 1 \tag{3}$$

We also define a *hierarchy constraint* to keep the control ergonomically intuitive. The hierarchy in the avatar should at least be loosely maintained on the hands' mapped joints - that is if j^a_i is mapped to j^h_k, any children of j^a_i should be mapped to children of j^h_k:

$$\sum_{R^h_{k,p}=0} \sum_{R^a_{i,q}=1} u_{j^a_i j^h_k} \times u_{j^a_q j^h_p} = 0 \tag{4}$$

The taxonomy retrieved from the observation study (Section 3.6.1 indicates that the majority of the user-designed mappings are balanced, with both hands having $\geq \frac{1}{3}$ of the mapped hand joints. We thus require that the difference in the number of the mapped joints on both hands should not exceed $\frac{1}{3}$ of the control joint numbers:

$$|\sum_{j^h_i \in J^l} \sum_{j^a_k \in J^a} u_{j^a_k j^h_i} - \sum_{j^h_p \in J^r} \sum_{j^a_q \in J^a} u_{j^h_p j^a_q}| \leq \frac{N}{3} \tag{5}$$

4.4 Optimization Scheme

The core decision for the algorithm to make is how to map a avatar control joint to a hand joint and how to select a pair of starting hand pose and avatar pose that maximize the objective function. Through the observation study (Section 3.6.2), we found that users prioritized control precision, intuition gained from structural similarity, and comfort level with varying personal preferences. We thus optimize them as sub-objectives in determining the joint-to-joint mappings. **Control(C)** indicates how precisely the avatar can be controlled with a mapping and is crucial to realize HandAvatar's flexible manipulation. We optimized the **structural similarity(S)** between the mapped hand poses and the avatar poses since the majority of the mapping designs in the observation study were *form-driven* and shape similarity was reported to be helpful to the embodiment. **Comfort(F)** describes how comfortable the hand poses associated with the mapping are when controlling an avatar. Participants favored gestures that are less physically demanding in the observation study. Comfortable mapping gestures are also fundamental to reducing physical workload in long-term uses of HandAvatar.

The control score C is calculated by how much the mapped hand joint range of motion overlaps with the control joint. If j^a_i is mapped to j^h_k, their pair-wise control score is calculated as

$$C_{j^a_i j^h_k} = |RoM^a_{j^a_n} \cap RoM^h_{j^h_m}| \tag{6}$$

We describe a hand pose or an avatar pose with each joint's rotation relative to its parent in the hierarchy. We use $t_{j^a_x}$ as the

angle of j_x^a related to its parent. The similarity score S between an avatar pose $p_i = \{t_{j_1^a}, \ldots, t_{j_m^a}\}$ and a hand pose $g_k = \{t_{j_1^h}, \ldots, t_{j_k^h}\}$ is calculated as the angular distance between a control joint and its mapped joint. The score between two joints can be found by:

$$S^{p_i g_k}_{j_x^a j_y^h} = |t_{j_x^a} - t_{j_y^h}| \tag{7}$$

For comfort score F of a hand pose g_i we adopt Mayer and Katzakis [26]'s method, which calculates a gesture's deviation from the rest posture, to evaluate hand pose comfort and rapid upper limb assessment (RULA) [27] to evaluate wrist comfort. The inter-finger angle score (IFA) evaluates the co-activation between the fingers. The hyperextension score (HE) measures the discomfort when the fingers are hyperextended. The finger abduction score (FA) measures the stress on the joints and the abduction muscles when fingers abduct. The RULA score takes the wrist's location and twist into account. Please refer to [26] and [27] for detailed calculation.

$$F_{g_i} = RULA_{left,\, right} + IFA(g_i) + HE(g_i) + FA(g_i) \tag{8}$$

Then the objective function of a mapping represented by $u_{j_p^h j_q^a}$ on a gesture-pose pair (g, p) is given by Equation 9. We empirically determined the weights to be $w_s = -5$, $w_c = 0.8$, $w_f = -0.01$ by considering the relative balance between these three metrics and fine-tuning the weights through trials. Designers can adjust the weights for different sub-objectives accordingly to generate mappings that suit their specific purposes better.

$$Q = w_f \times F_{g_i} + \sum_{j_p^h \in p_k} \sum_{j_q^a \in g_i} u_{j_p^h j_q^a} \times (w_s \times S^{g_i p_k}_{j_p^h j_q^a} + w_c \times C_{j_p^h j_q^a}) \tag{9}$$

Since the mapping is used for performing consecutive animations, the ideal optimization goal is to find the mapping M that maximizes the objective function Q considering all pose pairs (g_i, p_k) that may appear in the animations. This avoids over-fitting to a single static pose pair. However, as discussed in Section 4.2.3, the possible pairs are determined by the joint-to-joint mapping M and the reference pose pair (g, p). It is computationally extremely difficult to exhaust all possible combinations of the mappings and pose pairs. We thus decide to first optimize the joint-to-joint mapping with an estimation of the set of possible animation pose pairs $\{(g_i, p_k)\}$ with combinatorial optimization. With the obtained mapping, we then optimize the pose pair in a sampled set. For mapping optimization, we obtain two reasonable sampled sets of hand poses (G) and avatar poses (P), all of which are within the appropriate range of motion as discussed in Section 4.2. By enumerating the combinations of hand poses $(g_i \in G)$ and avatar poses $(p_k \in P)$, we obtain the estimation of the animation pose pairs. We then formulate finding the best mapping as an optimization problem.

$$M = \arg\max\Big(\sum_{g_i \in G} \sum_{p_k \in P} Q\Big) \tag{10}$$

With the obtained mapping M, we then calculate the optimal reference pose pair out of all possible combinations of hand pose from G and avatar pose from P. For each candidate reference hand pose and avatar pose pair (g, p), we calculate the hand poses for all possible avatar poses in P under the mapping M. As the result, we obtain a set of animation pose pairs $\{(g', p')\}$ as (G_g, P_p) for each candidate pose pair. We then calculate the reference pose pair that maximizes the overall score Q_M for the set of animation pose pairs (G_g, P_p).

$$\begin{aligned}(g, p) = \arg\max_{g \in G,\, p \in P}(Q_M) = \arg\max_{g \in G,\, p \in P}\Big(\sum_{(g',\, p') \in (G_g,\, P_p)} (w_f \times F_{g'} \\ + \sum_{(j_r^h,\, j_q^a) \in M} (w_s \times S^{g' p'}_{j_r^h j_q^a} + w_c \times C_{j_p^r j_q^a}))\Big)\end{aligned} \tag{11}$$

4.5 Optimization Evaluation

We used Gurobi [3] through a Python 3.7 interface to solve the formulated integer program problem. It takes about one hour to generate the mapping, with the exact time depending on the number of control joints. To evaluate the algorithm's performance in generating user-preferred, optimal mappings, we used the algorithm to generate optimal mappings for all avatars tested in the observation study.

We then calculated the control(C), structural similarity(S), and comfort(F) sub-objective scores for the 8 HandAvatar optimized mappings and all the 8×12 user-designed mappings in the observation study. Figure 5 shows the sub-objectives' scores of in total $8 + 96 = 104$ mappings. The HandAvatar optimized mappings receive higher scores in terms of comfort, similarity, and control compared to the user-generated mappings. This verifies that HandAvatar is able to generate overall better hand-to-avatar joint mappings than users. The distribution of user-voted mappings indicates that the calculation of sub-objectives is effective, as the voted mappings outperform most of the user-designed mappings in at least one of the metrics. This is also confirmed by the overall objective scores, which is 4.49 ($SD = 0.22$) for user-designed mappings, 4.63 ($SD = 0.19$) for user-voted mappings, and 5.07 ($SD = 0.18$) for HandAvatar-optimized mappings.

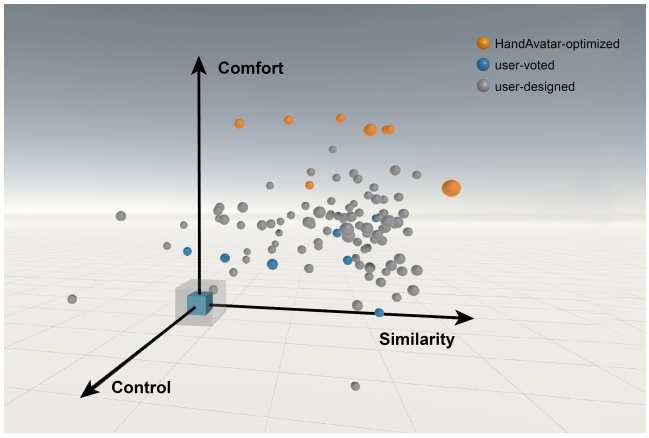


Figure 5: The distribution of the user-designed, user-voted, and HandAvatar-optimized mappings coordinated by their control, structural similarity, and comfort scores.

[3] https://www.gurobi.com/

To test the extensibility of the algorithm, we generated hand-to-avatar mappings for 25 different non-humanoid avatars. The mappings are illustrated by Figure 6. The averaged objective score of the generated mappings is 5.15 ($SD = 0.23$), which is consistent with the previous results.

5 EVALUATION

To evaluate the performance of HandAvatar, we chose three avatar manipulation tasks which include a target avatar imitation task to investigate the control precision, a creative exploration task to test the user experience, and a mapping comparison task to verify the effectiveness of the optimization algorithm. We invited 44 participants in different groups to complete the control precision evaluation ($N = 16$), the user experience evaluation ($N = 16$), and the mapping comparison evaluation ($N = 12$). We split the experiment into those three parts, since a single session would have taken too long for individual participants. We used a human body based mapping approach [5] and a hand based approach with user-voted mappings as the baselines for comparison.

5.1 Baseline approach

We select the human body based approach to control non-humanoid avatars (which we call BodyAvatar) as the baseline to compare with HandAvatar. BodyAvatar is implemented with a Kinect depth camera which provides the 3D coordinates of the user body key joints similar to the state-of-the-art method (Kin$\hat{E}$tre [5]). The control joints of the avatar mirror the movement of the corresponding body key joints based on the mapping. In Kin$\hat{E}$tre [5], the mapping between the user body and avatar joints are defined by the user through overlapping their joints and the corresponding avatar joints. Yet to keep the body-to-avatar mapping consistent which is a control factor in the evaluation, we invited 3 designers with expertise in animation creation to together design body-to-avatar joint mappings for all the tested avatars for BodyAvatar. For HandAvatar, we generated all the hand-to-avatar joint mappings with the proposed algorithm in Section 4. For BodyAvatar, the participants stand and use their full body to control the avatar; for HandAvatar, the participants are seated and use only their hands to control the avatar.

To further validate the effectiveness of the optimization algorithm for HandAvatar, we include an additional baseline approach in the mapping comparison task which involves the top-voted mappings obtained from the observation study. In the observation study's vote phase, we were able to re-invite one group of six participants to vote their favorite hand-to-avatar mappings within the group. We take the resulting top-voted mappings as the baseline as they are optimized within a group of six participants and can approximate high-quality mappings.

5.2 Task design

Control precision evaluation: To evaluate the control precision of HandAvatar, we choose two relatively controlled tasks, which are static posing and animating with a target. For the posing task, participants were asked to manipulate the avatar with either HandAvatar or BodyAvatar to be as similar to the target avatar pose as possible. We consider this task as the most fundamental avatar manipulation task which requires high control precision. We calculate the **pose deviation** - the average deviation of the control joints in the static poses compared to the target poses, to measure the control precision. For the animating task, participants were asked to mimic a given target animation clip to create a piece of animation. This task requires participants to carefully control the continuous movement in the animation clip and requires even higher control precision compared to the static poses. Participants are asked to create at least 3 animation clips. We calculate the **animation deviation** - the per frame deviation of control joints in the created animation compared to the target animation, as the metric of precision. We applied a dynamic time warping algorithm to align the frames of the test animation and the user-performed animation. The tested animations were taken from the avatar's built-in animators and the tested poses were randomly selected from animation frames. Four avatars were tested in the two tasks, spider and elephant from the observation study, and swan and sea horse as new avatars. The avatars were selected such that they covered different motion patterns (walking, flying, and swimming) and body structures, and their generated mappings were distinct. The order of the control method and the avatars is counter-balanced using Latin-squares. In total, we collected 16 participants ×4 avatars ×2 control methods ×3 static poses = 384 static poses and at least 16 participants ×4 avatars ×2 control methods ×3 clips = 384 animation recordings.

User experience evaluation: We evaluate user experience of HandAvatar with free exploration tasks where participants create animation clips without any constraints. In this task, we compare HandAvatar with BodyAvatar with two new avatars (crab and duck). Instead of completing a standard action, we expect this task to allow participants to fully experience the creative animating process which is closer to HandAvatar's real usage scenario. To help participants get into the state of creation, we ask them to come up with a theme with the animation they create, and iterate on the creations at will and need. To measure the user experience, we ask participants to rate embodiment, degree of control, comfort, physical workload, and satisfaction with the produced animation on a 7-point Likert scale (1: strongly disagree, 7: strongly agree).

Mapping comparison evaluation: We apply the same task as user experience evaluation while comparing HandAvatar to the baseline of user voted mappings. We ask the participants to focus on the designs of joint-to-joint mappings generated by two different methods. The same dimensions of user experience (embodiment, degree of control, comfort, physical workload, and satisfaction) as user experience evaluation were evaluated by the participants.

5.3 Participants

We recruited 16 participants (6 males, 10 females) with an average age of 22.69 ($SD = 3.34$) for the control precision evaluation. For the user experience evaluation, we recruited 16 new different participants (7 males, 9 females) with an average age of 23.63 ($SD = 3.50$). For the mapping comparison evaluation, we recruited 12 new participants (6 males, 6 females) with an average age of 22.08 ($SD = 1.78$). None of them participated in the observation study. The self-reported familiarity with VR and with puppetry averaged 3.69 ($SD = 1.08$) and 3.06 ($SD = 1.12$) for control precision evaluation, averaged 3.50 ($SD = 1.10$) and 2.50 ($SD = 1.32$) for

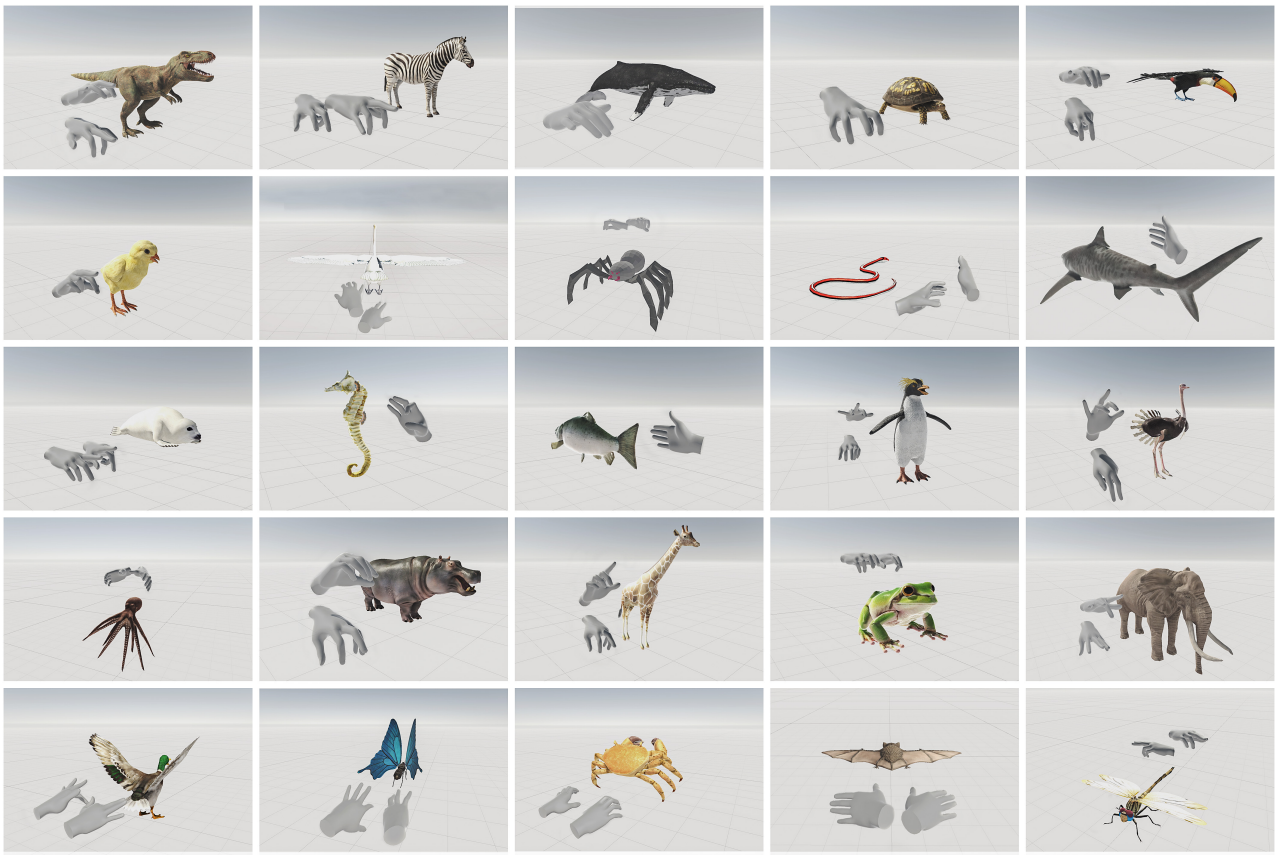


Figure 6: Hand-to-avatar mappings for 25 animal avatars optimized by HandAvatar algorithm.

user experience evaluation, and averaged 4.29 (SD = 1.60) and 2.86 (SD = 1.57) for the mapping comparison evaluation on a 7-point Likert scale (1: strongly agree, 7: strongly disagree).

5.4 Procedure

Control precision evaluation: Participants first completed demographics surveys. For both HandAvatar and BodyAvatar, participants were given 5 minutes to learn and familiarise themselves with the help of the researchers. Then for each of the four avatars, participants first try their best to pose according to 3 static test poses. They then perform to mimic a 10-second animation clip for at least 3 times and can have more trials until they are satisfied with the performances. The entire study lasted around 1 hour. Each participant was compensated 15 U.S. dollars for their time and effort.

User experience evaluation and mapping comparison evaluation: Participants first completed demographic surveys. For the two control methods, participants were first guided to learn and familiarise themselves. Then for each avatar, participants were given 5 minutes to freely explore the control experience during which they needed to perform a short animation clip with their own theme. The recorded animation is replayed afterward and can be recorded again until the participants are satisfied. After each trial, the participants provided subjective ratings on user experience. After completing all the recordings, the participants were interviewed about their experiences and comments. The entire study lasted about 40 minutes and each participant was compensated 15 U.S. dollars.

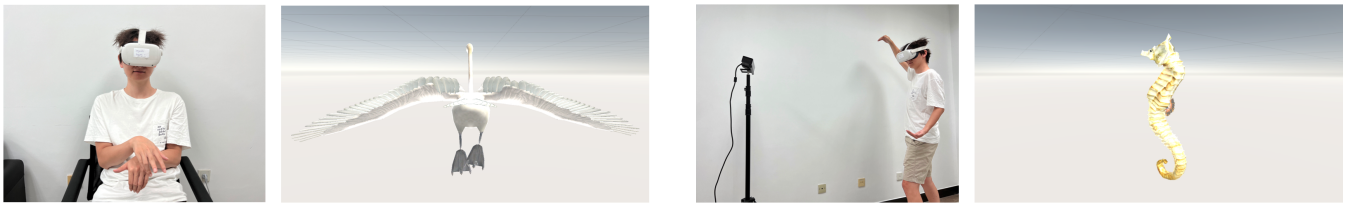
5.5 Apparatus

The experimental platform was developed in Unity 2021 for the Meta Quest 2 headset powered by an Intel Core i7 CPU and an NVIDIA GeForce RTX 3080 GPU. The participants animate each avatar in a default background. The study setups and VR platforms for both HandAvatar and BodyAvatar are shown in Figure 7. All statistical analyses were conducted with SPSS 26.0.

5.6 Results

5.6.1 Control precision. Though both HandAvatar and BodyAvatar were optimized for fine control through our algorithm and by the designers respectively, avatars controlled with HandAvatar (M = 21.67, SD = 9.57) have more control joints than with BodyAvatar (M = 9.67, SD = 1.37). This indicates that because of the hands' dexterity, they offer a more elaborate control compared to human bodies.

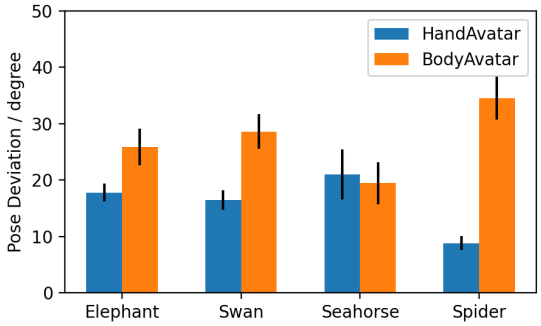
HandAvatar allows more accurate control on static poses compared to BodyAvatar. For pose deviation, we conducted Mauchly's test of sphericity on avatars (W = 0.79, p = 0.66)



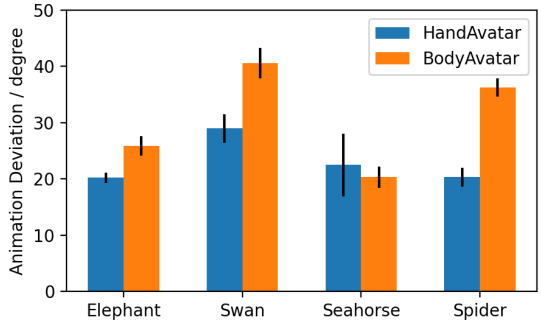
(a) HandAvatar: participants control avatars with hands.

(b) BodyAvatar: participants control avatars with their body.

Figure 7: Study setup for HandAvatar and BodyAvatar in the evaluation.



(a) Average and standard deviation for static poses.



(b) Average and standard deviation for animations.

Figure 8: (a) Averaged pose deviations for static poses and (b) animation deviations for animation imitation tasks, as the metrics for control precision. The error bars stand for standard deviations.

and avatars × control methods ($W = 0.75$, $p = 0.55$) and the results indicated no sphericity violation. We thus conducted repeated measures ANOVA tests on pose deviation and used Bonferroni-corrected post hoc T-tests for pairwise comparisons if any significant influence exists ($p < 0.05$). RM-ANOVA tests showed that both avatars ($F_{3,45} = 3.43$, $p < 0.05$) and control methods ($F_{1,15} = 286.18$, $p < 0.001$) affected pose deviation significantly and the interaction effect between avatars and control methods were also significant ($F_{3,45} = 137.05$, $p < 0.001$). The pose deviation of HandAvatar ($M = 16.01$, $SD = 4.48$) is significantly lower than BodyAvatar ($M = 27.11$, $SD = 5.44$). Pairwise comparisons indicated no significant difference between HandAvatar ($M = 21.00$, $SD = 4.56$) and BodyAvatar ($M = 19.43$, $SD = 3.89$) only on seahorse ($t(15) = 1.19$, $p = 0.25$). This is due to the seahorse's relatively simple body structure with few moving body parts, making the seahorse easy to control with both control methods.

Animating avatars with HandAvatar is also more accurate than with BodyAvatar. For animation deviation, we found sphericity violation on avatars × control methods ($W = 0.21$, $p = 0.001$) but not on avatars ($W = 0.56$, $p = 0.161$) with Mauchly's test of sphericity. Therefore, we used the Greenhouse-Geisser correction to report the repeated measures ANOVA results. RM-ANOVA tests showed that both avatars ($F_{3,45} = 129.82$, $p < 0.001$) and control methods ($F_{1,15} = 202.30$, $p < 0.001$) affected animation deviation significantly and the interaction effect between avatars and control methods were also significant ($F_{3,23.94} = 56.40$, $p < 0.001$). The animation deviation of HandAvatar ($M = 23.00$, $SD = 3.57$) is significantly lower than BodyAvatar ($M = 30.76$, $SD = 8.04$). Similar to pose deviation, pairwise comparisons indicated no significant difference between HandAvatar ($M = 22.50$, $SD = 5.75$) and BodyAvatar ($M = 20.34$, $SD = 1.95$) only on seahorse ($t(15) = 1.27$, $p = 0.19$) due to its structural simplicity.

5.6.2 User experience evaluation. We performed Wilcoxon signed-rank tests on participants' subjective scores. The tests revealed no significance on embodiment ($Z = -1.08$, $p = 0.28$) between HandAvatar ($M = 4.50$, $SD = 1.34$) and BodyAvatar ($M = 4.78$, $SD = 1.43$), indicating that HandAvatar is comparable to BodyAvatar in embodiment. HandAvatar ($M = 5.38$, $SD = 1.39$) brings better sense of control ($Z = -3.12$, $p < 0.01$) compared to BodyAvatar ($M = 4.28$, $SD = 1.35$). Participants indicated that HandAvatar brings finer control of more body parts of the avatars than BodyAvatar (P8). HandAvatar ($M = 6.03$, $SD = 1.20$) requires significantly less physical load ($Z = -2.69$, $p < 0.01$) than BodyAvatar ($M = 5.03$, $SD = 1.62$) which requires the users to constantly perform large scale body movements (e.g. waving arms at a high frequency when animating the duck). Similarly, participants have significantly more comfortable experience ($Z = -2.00$, $p < 0.05$) animating with HandAvatar ($M = 5.47$, $SD = 1.17$) than BodyAvatar ($M = 4.66$, $SD = 1.41$) which at times require twisting and hard-to-maintain body poses like standing on one leg or leaning forward. In general, participants were more satisfied ($Z = -2.65$, $p < 0.01$) with their own animations created with HandAvatar ($M = 5.13$, $SD = 1.24$) than HandAvatar ($M = 4.38$, $SD = 1.04$). This indicates that **HandAvatar outperforms BodyAvatar on control, comfort, physical load, and satisfaction, and is comparable to BodyAvatar on providing embodiment.**

Interview results show that 11 out of 16 participants prefer HandAvatar over BodyAvatar, 3 participants have avatar-dependent preferences, and 2 participants like BodyAvatar more. Those with

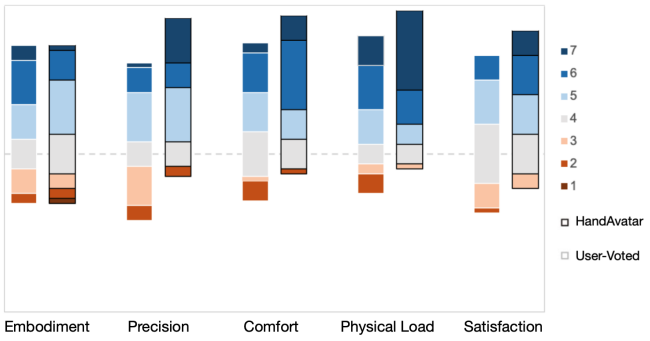


Figure 9: Subjective ratings comparing HandAvatar and BodyAvatar in the dimensions of sense of embodiment, control, comfort, physical load, and satisfaction towards the performance.

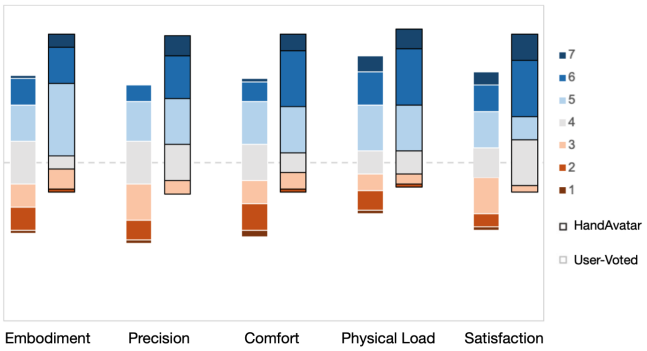


Figure 10: Subjective ratings comparing HandAvatar and user-voted mappings in the dimensions of sense of embodiment, control, comfort, physical load, and satisfaction towards the performance.

avatar-dependent preferences indicated that HandAvatar is a better choice when it comes to *"hand-alike avatar with more than 4 limbs like spiders and crabs"* (P16) while BodyAvatar suits better for *"human-alike avatars like penguin and duck"* (P8).

The participants were generally optimistic toward HandAvatar, because it requires *"less physical efforts"* (N = 5), is *"direct and intuitive to control"* (N = 4), and can *"control more joints and complete more complex movements"* (N = 7). However, the participants also pointed out and wished to overcome the fingers' co-activation constraint when using HandAvatar (N = 9).

BodyAvatar was less preferred due to its high physical demands (N = 6), limited controlled joints (N = 3), and the social awkwardness brought by the exaggerated body movements (N = 3).

The created themed animations show that with both HandAvatar and BodyAvatar, participants came up with creative and especially anthropomorphic themes, including conducting an orchestra (P1, 4, 15), parachuting (P14), boxing (P6, 15), dancing (P3, 6, 13, 14, 16), playing piano (P16), etc. This demonstrates that with interactive avatar manipulation approaches, users tend to animate the avatars for creative and expressive content that exceed these avatars' real-world activities. This thus provides the motivation for boosting the precision and embodiment of such interactive avatar manipulation methods to further enhance the user experience.

5.6.3 Mapping comparison evaluation. We performed Wilcoxon signed-rank tests on participants' subjective scores. The tests revealed significance on all metrics except for embodiment ($Z = -1.92$, $p = 0.06$) between HandAvatar ($M = 5.00$, $SD = 0.74$) and user-voted mappings ($M = 4.13$, $SD = 1.08$), indicating that they are comparable in terms of embodiment. HandAvatar ($M = 5.13$, $SD = 0.83$) outperforms user-voted mappings ($M = 3.92$, $SD = 0.97$) on the sense of control ($Z = -2.41$, $p = 0.016$). Participants felt more comfortable ($Z = -2.52$, $p = 0.012$) using HandAvatar ($M = 5.17$, $SD = 0.94$) to control avatars than user-voted mappings ($M = 3.98$, $SD = 1.08$). HandAvatar ($M = 5.27$, $SD = 0.86$) also required less physical load ($Z = -2.54$, $p = 0.011$) than user-voted mappings ($M = 4.60$, $SD = 1.38$). Overall, the participants were more satisfied ($Z = -2.15$, $p = 0.031$) with their animations created using HandAvatar ($M = 5.31$, $SD = 0.79$) than those created using user-voted mappings ($M = 4.35$, $SD = 1.18$).

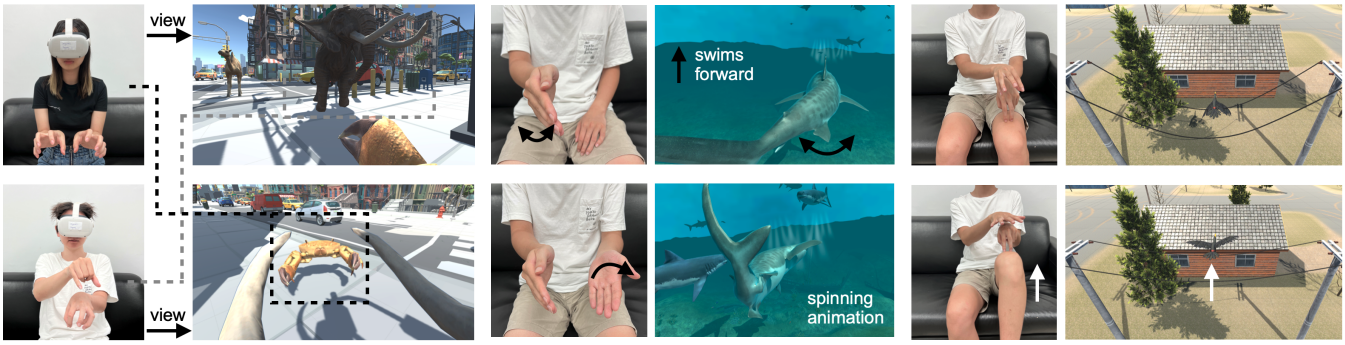
The algorithm-optimized mappings were preferred 27 out of 48 times by the participants. Participants mentioned the lack of gestural comfort, especially due to the hands' co-activation constraint (N = 6), and the high physical workload (N = 2) as reasons they disliked the user-voted mappings. They were also dissatisfied with the smaller range of motion on some finger joints compared to the avatar joints they were mapped to (N = 5). However, user-voted mappings were sometimes preferred over the algorithm-optimized ones for their simplicity (N = 3) and accustomed puppetry conventions like using crossed hands to imitate wings (N = 4). On the other hand, the participants liked algorithm-optimized mappings' precise control (N = 4), comfort (N = 3), and larger range of motion on some joints (N = 2). They also commented that the gestures used were structurally similar to the avatar and thus intuitive to learn (N = 4). The results indicate that the optimized mappings have equal or better performance on all avatars than user-voted mappings, which already leveraged six users' knowledge, and a single user might not be able to design. This implies that the algorithm can help individual, non-expert users to produce optimal hand-to-avatar mappings for controlling the avatars. The results also suggest further inputs or constraints the algorithm can involve, e.g., puppetry conventions to which the users are accustomed.

6 APPLICATIONS

Based on HandAvatar's proposed control method and optimization algorithm, we extended our implementation to demonstrate three potential applications to showcase future interaction scenarios that would benefit from users embodying non-humanoid avatars. The applications include non-humanoid avatar based social interaction in VR, embodied 3D animation creation, and 3D VR scene design with physical proxies.

6.1 Non-humanoid Avatar based Social Interaction in VR

Interaction in a virtual workspace and social scenarios is increasingly popular, especially in the post-COVID era. Yet in most of the available online interaction applications, users are represented and interact with each other as humanoid avatars to bring a strong



(a) Users socially interact as non-humanoid avatars in the first point of view as a crab and an elephant respectively. They can see each other's body movements and respond in real-time.

(b) User creates a 3D animation of the shark by swaying the tail at different frequencies to change the swimming speed and using a rolling hand gesture to trigger a spinning animation.

(c) User animates a toucan and uses a leg to provide active feedback of a cable that the toucan stands on. The up-down movement of the leg corresponds to the oscillation of the cable.

Figure 11: Three applications implemented to demonstrate HandAvatar's potentials.

sense of embodiment in a virtual scene. HandAvatar contributes to the less explored non-humanoid avatar based virtual social interaction. We envision that representing users as non-humanoid avatars brings diversity to virtual social interaction as users can carry out different activities with changing non-humanoid avatar identities. With an optimized and customizable mapping, HandAvatar enables users to control arbitrary non-humanoid avatars precisely and thus expressively in real-time, allowing users to manipulate the non-humanoid avatars to express emotions essential in social interactions.

The demonstrate this potential, we developed an application that enables multiple users to interact with each other as animal avatars from the first point of view. The animal avatars can freely meet, greet, and fight with other users with their unique features (e.g., crab's claw, elephant's nose). HandAvatar enables the users to both experience as an animal and perceive other users in the form of animals. During the bi-directional communication and interaction, we envision this as a unique and exciting future social interaction scenario in VR, where users can gain a strong sense of embodiment as non-humanoid animals.

6.2 3D animation composition with HandAvatar

One beneficial application of HandAvatar is simple, quick, and embodied 3D animation composition. Compared to the traditional keyframing method which requires expertise in animation and hours for inputting key poses, rendering, etc., HandAvatar allows novice users to compose 3D animations with a much simpler workflow. With the intuitive control provided by the optimized mapping, the users can directly manipulate arbitrary avatars with their hands in real-time and adjust their performances quickly. We hope to motivate everyday animation creation with the simplified animation workflow and to spark creativity in animation through the embodiment brought by HandAvatar. Additional design and editing tools can be developed to help users create 3D animations with HandAvatar, such as timeline-based editing tools that allow the creation of layered and more complex animations.

To demonstrate one such example, we developed an application to facilitate real-time 3D animation compositions by extending our implementation of HandAvatar in the evaluation study (Section 5) with two extra functionalities suggested by the participants. The two functionalities are locomotion which enables the avatar to move to other places in a VR scene, and trigger-to-motion which triggers a frequently used animation clip upon a pre-define hand gesture. To demonstrate this we developed an underwater VR scene in which the user uses the right hand to control a shark based on an optimized mapping. The frequency of the body swaying movement determines the swimming speed and the direction of the wrist determines the swimming direction; On the left hand, a rolling hand gesture triggers a spinning motion of the shark, and pinching the thumb with the other four fingers triggers an attacking animation.

6.3 3D VR scene design with physical proxies

HandAvatar enables controlling non-humanoid avatars with only the hands. This frees up other body parts which we imagine can be used to simulate the VR environment that the avatar is currently in to provide passive or active haptic feedback. This analogously lets users embody the entire avatar-included VR scenario with their whole body, extending HandAvatar from hand-to-avatar to body-to-scene control. The other body parts are suitable candidates to provide haptic feedback for HandAvatar due to their active motion capability, good coordination with the hands, and human proprioception. This further enhances HandAvatar's storytelling capabilities by allowing users to control more and enabling body parts to interact as different agents. We developed a VR demonstration in which the legs are used to simulate hanging cables to provide active environmental feedback to a toucan controlled through HandAvatar. The toucan is playing and resting on the cable which starts to oscillate. The user's leg mimics the oscillation motion by moving up and down, thereby providing haptic feedback to the hands in real-time. This application also echoes our participants' suggestions to place their hands on the table or on their laps which provides passive force to support their hands to reduce the physical load in the observation study (P2, P11).

7 DISCUSSION

We proposed HandAvatar to enable users to control and embody non-humanoid avatar through hands. We contribute an optimization algorithm to automatically create hand-to-avatar joint-to-joint mappings that consider control, structural similarity, and comfort. Evaluation results indicate that HandAvatar is better in both control precision and user experience than a body-based control method. In the following, we discuss possible extensions of HandAvatar as well as limitations and future work.

7.1 Extensions of the optimization algorithm

Informed by the observation study (Section 3), the current implementation optimizes the control precision, structural similarity, and comfort level of the hand-to-avatar joint mappings (Section 4). The optimization can be further extended by incorporating other factors. For instance, to allow flexibility in hardware choices when implementing HandAvatar, the level of tracking requirements (e.g., proportion of hands being occluded) can be considered as a sub-objective in optimization so that severe hand occlusions are avoided and light occlusion can be traded for promotion of other sub-objectives. For tracking sensitive scenarios, tracking requirements can also be framed as hard constraints that the algorithm must not violate. This allows HandAvatar to be extendable and applicable to a wider range of use cases and different devices. The current implementation applies the empirical weights for three sub-objectives which proved to be consistent with participants' preferences. We expect that weights can be customized for specific applications with extra knowledge of the priority of different sub-objectives.

7.2 Augmentations of the one-to-one joint mappings

In HandAvatar, only one-to-one mappings are considered - each mapped hand joint can control one and only one avatar control joint - to avoid the discrepancy between avatar and hand joint movement affecting the sense of embodiment. Many other mapping mechanisms applicable to different scenarios can be considered. However, their respective effects on user experience, including control precision, the sense of embodiment and more, need further investigation.

For instance, a degree of freedom(DoF)-to-DoF mapping can be considered, which can map the rotations of different axes on two joints. This may help further overcome the finger joints' limited range of motion on some axes. One-to-many mappings can also be considered in which multiple avatar control joints are mapped to the same hand joint, e.g. all of the spider's legs move synchronously with one hand finger. This would enable users to have more control joints than the number of hand joints or to reduce the physical workload by mirroring one hand joint's motion to multiple control joints. However, the mental workload might increase as users need to coordinate between multiple control joints and the specific effect will need to be investigated.

Another useful extension of the current algorithm is to allow generating a mapping for not one, but multiple non-humanoid avatars in an avatar group. In the observation study, participants mentioned a storytelling scenario in which the user may want to switch frequently between different avatar characters. We believe it would be beneficial to enable users to avoid switching between different hand-to-avatar mappings by using the same generalized mapping to control multiple different avatars through inputting multiple avatars into the algorithm.

7.3 Limitation and future work

HandAvatar is implemented such that the users control the avatars from the third point of view. This ensures that the users can have a clear view of the avatars' movement when controlling them in tasks like animation. Yet past work has shown that the first point of view brings stronger SoE in VR. In the future, we will further investigate how point of view affects SoE, control, and other user experience metrics when controlling non-humanoid avatars with HandAvatar.

HandAvatar's hand tracking is currently realized by Meta Quest 2's built-in cameras and hand tracking modules which track the hands with satisfactory accuracy when both hands are not occluded. It requires no extra tracking equipment, providing users a free-hand experience and ensuring HandAvatar to be applicable to head-mounted displays (HMDs) with similar hand-tracking modules. However, there were cases when the participants' hands were inaccurately tracked when their hands were crossed or lost when they were far out of the field of view. We envision that the combination usage of HMD and other advanced tracking approaches [11, 36] help alleviate this issue.

In evaluation study, we chose BodyAvatar [5] as the baseline to compare with HandAvatar. The consideration is that body-based avatar control provides higher degrees of freedom compared to controllers and other input devices, is the most commonly adopted interactive avatar control method on today's VR headsets, and users are thus familiar with the approach. We acknowledge that the body-to-avatar joint mappings were not computationally optimized for BodyAvatar but were manually designed by three designers. This is because there is no optimization directly applicable to body-to-avatar mappings to our knowledge, and our optimization algorithm (for hand-to-avatar mappings) did not work well with the human body as we tested in pilots. We regard it as an important future work to adapt the proposed method to the generation of optimal mappings between other inputs (e.g., whole body) and virtual avatars.

In the observation study and evaluation study, we tested 12 avatars in total, and we showed the generated mappings of 25 avatars in algorithm evaluation. We acknowledge that although these avatars cover a variety of important dimensions including structural complexity and motion methods, more non-humanoid avatars need to be tested to further enhance the algorithm's scalability to different avatars. Depending on the specific avatar, in some cases HandAvatar's performance may be similar or less satisfying than BodyAvatar when the avatar is human-like (e.g. chimpanzee) or controller-based approach when its structure is extremely simple.

8 CONCLUSION

We present HandAvatar, an enabling technology that allows users to embody non-humanoid avatars in VR in a precise, intuitive, and comfortable manner. HandAvatar is built on the knowledge of user's preferences on hand-to-avatar mappings from an observation study. The core contribution is an automatic mapping algorithm, which considers knowledge of the hand's biomechanics, takes avatar parameters and interaction requirements as input, and simultaneously optimizes for control precision, intuitiveness and comfort. We evaluated HandAvatar in the tasks of mimicking static poses, animated motions, and free animation creation, compared to a body-based baseline method [5] and user-voted mappings. Results show that HandAvatar outperformed the baselines in control precision, physical load, and comfort level with a comparable sense of embodiment. We also implemented a set of applications to demonstrate the potential of HandAvatar to unlock new user-as-non-humanoids interaction experiences.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China (NSFC) under Grant No.62102221, No.62002198.

REFERENCES

[1] Karan Ahuja, Eyal Ofek, Mar Gonzalez-Franco, Christian Holz, and Andrew D. Wilson. 2021. CoolMoves: User Motion Accentuation in Virtual Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 52 (jun 2021), 23 pages. https://doi.org/10.1145/3463499

[2] Andreas Aristidou, Joan Lasenby, Yiorgos Chrysanthou, and Ariel Shamir. 2018. Inverse Kinematics Techniques in Computer Graphics: A Survey. *Computer Graphics Forum* 37 (2018).

[3] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands 'feel'touch that eyes see. *Nature* 391, 6669 (1998), 756–756.

[4] Ian M Bullock, Raymond R Ma, and Aaron M Dollar. 2012. A hand-centric classification of human and robot dexterous manipulation. *IEEE transactions on Haptics* 6, 2 (2012), 129–144.

[5] Jiawen Chen, Shahram Izadi, and Andrew Fitzgibbon. 2012. KinEtre: Animating the World with the Human Body. *UIST'12 - Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, 435–444. https://doi.org/10.1145/2380116.2380171

[6] Farnaz Farahanipad, Harish Ram Nambiappan, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. 2020. HAND-REHA: Dynamic Hand Gesture Recognition for Game-Based Wrist Rehabilitation. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '20)*. Association for Computing Machinery, New York, NY, USA, Article 17, 9 pages.

[7] Andreas Fender, Jörg Müller, and David Lindlbauer. 2015. Creature teacher: A performance-based animation system for creating cyclic movements. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. 113–122.

[8] Mary C. Hume, Harris Gellman, Harry McKellop, and Robert H. Brumfield. 1990. Functional range of motion of the joints of the hand. *The Journal of Hand Surgery* 15, 2 (1990), 240–243. https://doi.org/10.1016/0363-5023(90)90102-W

[9] Edwin Hutchins, James Hollan, and Donald Norman. 1985. Direct Manipulation Interfaces. *Human-computer Interaction* 1 (12 1985), 311–338. https://doi.org/10.1207/s15327051hci0104_2

[10] Takeo Igarashi, Tomer Moscovich, and John F Hughes. 2005. As-rigid-as-possible shape manipulation. *ACM transactions on Graphics (TOG)* 24, 3 (2005), 1134–1141.

[11] Umar Iqbal, Pavlo Molchanov, Thomas Breuel, Juergen Gall, and Jan Kautz. 2018. Hand Pose Estimation via Latent 2.5D Heatmap Regression. In *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 125–143.

[12] Rubaiat Habib Kazi, Kien Chuan Chua, Shengdong Zhao, Richard Davis, and Kok-Lim Low. 2011. SandCanvas: A multi-touch art medium inspired by sand animation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1283–1292.

[13] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387.

[14] Andrey Krekhov, Sebastian Cmentowski, and Jens Krüger. 2019. The illusion of animal body ownership and its potential for virtual reality games. In *2019 IEEE Conference on Games (CoG)*. IEEE, 1–8.

[15] Kyung-Sun Lee and Myung-Chul Jung. 2015. Ergonomic evaluation of biomechanical hand function. *Safety and health at work* 6, 1 (2015), 9–17.

[16] Luís Leite and Veronica Orvalho. 2012. Shape your body: control a virtual silhouette using body motion. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*. 1913–1918.

[17] Luis LEite and Veronica Orvalho. 2017. Mani-pull-action: Hand-based digital puppetry. *Proceedings of the ACM on Human-Computer Interaction* 1, EICS (2017), 1–16.

[18] Zhipeng Li, Yu Jiang, Yihao Zhu, Ruijia Chen, Ruolin Wang, Yuntao Wang, Yukang Yan, and Yuanchun Shi. 2022. Modeling the Noticeability of User-Avatar Movement Inconsistency for Sense of Body Ownership Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 64 (jul 2022), 26 pages. https://doi.org/10.1145/3534590

[19] Zhipeng Li, Yu Jiang, Yihao Zhu, Ruijia Chen, Ruolin Wang, Yuntao Wang, Yukang Yan, and Yuanchun Shi. 2022. Modeling the Noticeability of User-Avatar Movement Inconsistency for Sense of Body Ownership Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 64 (jul 2022), 26 pages. https://doi.org/10.1145/3534590

[20] Hui Liang, Jian Chang, Shujie Deng, Can Chen, Ruofeng Tong, and Jianjun Zhang. 2015. Exploitation of novel multiplayer gesture-based interaction and virtual puppetry for digital storytelling to develop children's narrative skills. In *Proceedings of the 14th ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry*. 63–72.

[21] Noah Lockwood and Karan Singh. 2012. Fingerwalking: motion editing with contact-based hand performance. In *Proceedings of the 11th ACM SIGGRAPH/Eurographics conference on Computer Animation*. 43–52.

[22] Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. 2015. Avatar anthropomorphism and illusion of body ownership in VR. In *2015 IEEE Virtual Reality (VR)*. 229–230. https://doi.org/10.1109/VR.2015.7223379

[23] Zhiqiang Luo, Chih-Chung Lin, I Chen, Song Huat Yeo, Tsai-Yen Li, et al. 2011. Puppet playing: An interactive character animation system with hand motion control. In *Transactions on computational science XII*. Springer, 19–35.

[24] Matteo Martini. 2016. Real, rubber or virtual: The vision of "one's own" body as a means for pain modulation. A narrative review. *Consciousness and cognition* 43 (2016), 143–151.

[25] Marta Matamala-Gomez, Tony Donegan, Sara Bottiroli, Giorgio Sandrini, Maria V Sanchez-Vives, and Cristina Tassorelli. 2019. Immersive virtual reality and virtual embodiment for pain relief. *Frontiers in human neuroscience* 13 (2019), 279.

[26] Jonas Mayer and Nicholas Katzakis. 2016. A Metric for Short-Term Hand Comfort and Discomfort: Exploring Hand Posture Evaluation *(SUI '16)*. Association for Computing Machinery, New York, NY, USA, 131–134. https://doi.org/10.1145/2983310.2985752

[27] L Mcatamney and E Nigel Corlett. 1993. RULA: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics* 24 2 (1993), 91–9.

[28] Ali Momeni and Zachary Rispoli. 2016. Dranimate: Paper Becomes Tablet, Drawing Becomes Animation. 3735–3737. https://doi.org/10.1145/2851581.2890267

[29] Ali Momeni and Zachary Rispoli. 2016. Dranimate: paper becomes tablet, drawing becomes animation. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 3735–3737.

[30] Jean-Marie Normand, Elias Giannopoulos, Bernhard Spanlang, and Mel Slater. 2011. Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PloS one* 6, 1 (2011), e16128.

[31] Sageev Oore, Demetri Terzopoulos, and Geoffrey Hinton. 2002. A desktop input device and interface for interactive 3d character animation. In *Graphics Interface*, Vol. 2. 133–140.

[32] Masaki Oshita, Yuta Senju, and Syun Morishige. 2013. Character motion control interface with hand manipulation inspired by puppet mechanism. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*. 131–138.

[33] Sofia Adelaide Osimo, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater. 2015. Conversations between self and self as Sigmund Freud—A virtual body ownership paradigm for self counselling. *Scientific reports* 5, 1 (2015), 1–14.

[34] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. 2022. Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 429, 16 pages. https://doi.org/10.1145/3491102.3501898

[35] Valeria I Petkova, Mehrnoush Khoshnevis, and H Henrik Ehrsson. 2011. The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in psychology* 2 (2011), 35.

[36] Fernando Quivira, Toshiaki Koike-Akino, Ye Wang, and Deniz Erdogmus. 2018. Translating sEMG signals to continuous hand poses using recurrent neural networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 166–169. https://doi.org/10.1109/BHI.2018.8333395

[37] Yeongho Seol, Carol O'Sullivan, and Jehee Lee. 2013. Creature features: online motion puppetry for non-human characters. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 213–221.

[38] Ben Shneiderman. 1984. The Future of Interactive Systems and the Emergence of Direct Manipulation. In *Proc. of the NYU Symposium on User Interfaces on Human Factors and Interactive Computer Systems* (New York, USA). Ablex Publishing Corp., USA, 1–28.

[39] Mel Slater and Maria V Sanchez-Vives. 2016. Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI* 3 (2016), 74.

[40] Mel Slater, Bernhard Spanlang, and David Corominas. 2010. Simulating Virtual Environments within Virtual Environments as the Basis for a Psychophysics of Presence. *ACM Trans. Graph.* 29, 4, Article 92 (jul 2010), 9 pages. https://doi.org/10.1145/1778765.1778829

[41] William Steptoe, Anthony Steed, and Mel Slater. 2013. Human Tails: Ownership and Control of Extended Humanoid Avatars. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 583–590. https://doi.org/10.1109/TVCG.2013.32

[42] William Steptoe, Anthony Steed, and Mel Slater. 2013. Human tails: ownership and control of extended humanoid avatars. *IEEE transactions on visualization and computer graphics* 19, 4 (2013), 583–590.

[43] David Joel Sturman. 1992. *Whole-hand input.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[44] David J Sturman. 1998. Computer puppetry. *IEEE Computer Graphics and Applications* 18, 1 (1998), 38–45.

[45] Manos Tsakiris. 2010. My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia* 48, 3 (2010), 703–712.

[46] Daniel Vogel, Paul Lubos, and Frank Steinicke. 2018. Animationvr-interactive controller-based animating in virtual reality. In *2018 IEEE 1st Workshop on Animation in Virtual and Augmented Environments (ANIVAE)*. IEEE, 1–6.

[47] Der-Lor Way, Weng-Kei Lau, and Tzu Ying Huang. 2019. Glove puppetry cloud theater through a virtual reality network. In *ACM SIGGRAPH 2019 Posters*. 1–2.

[48] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376687

[49] Katsu Yamane, Yuka Ariki, and Jessica Hodgins. 2010. Animating non-humanoid characters with human motion data. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 169–178.

[50] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173652

[51] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *2010 IEEE Virtual Reality Conference (VR)*. 95–102. https://doi.org/10.1109/VR.2010.5444807