The Impact of Avatar Retargeting on Pointing and Conversational Communication

Simbarashe Nyatsanga, Doug Roble, and Michael Neff

Fig. 1: Frames from experiment stimuli. The left two panes are from an experiment on pointing perception (Exp. 1) and show the IK and SemanticIK conditions, respectively. The middle two panes are from an experiment on object size estimations from gesture (Exp. 2) and show the elongated skeleton with the IK reconstruction and the base avatar. The right two panes are from an experiment on social signals in speech (Exp. 3) and show the conditions ThinLongFK and BulkyLongIK, respectively.

Abstract—One of the pleasures of interacting using avatars in VR is being able to play a character very different to yourself. As the scale of characters change relative to a user, there is a need to retarget user motions onto the character, generally maintaining either the user's pose or the position of their wrists and ankles. This retargeting can impact both the functional and social information conveyed by the avatar. Focused on 3rd-person (observed) avatars, this paper presents three studies on these varied aspects of communication. It establishes a baseline for near-field avatar pointing, showing an accuracy of about 5cm. This can be maintained using positional hand constraints, but increases if the user's pose is directly transferred to the character. It is possible to maintain this accuracy with a Semantic Inverse Kinematics formulation that brings the avatar closer to the user's actual pose, but compensates by adjusting the finger pointing direction. Similar results are shown for conveying spatial information, namely object size. The choice of pose or position based retargeting leads to a small change in the perception of avatar personality, indicating an impact on social communication. This effect was not observed in a task where the users' cognitive load was otherwise high, so may be task dependent. It could also become more pronounced for more extreme proportion changes.

Index Terms—Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality

1 Introduction

This paper explores the potential impacts of retargeting choices on avatar-based, conversational interaction in VR. During conversations, people obtain both functional and social information from the motion of their interlocutor. Functional information might include what item is pointed to in order to create a reference or a gestural size indication. Social information can give clues about a character's personality, mood and emotions. In VR settings, people may choose to occupy avatars with very different proportions than their own. A popular contemporary example is VRChat, which allows users to interact with others using user-created 3D avatars with very different proportions [26]. Recent research efforts show potential applications in animal embodiment [28], psychotherapy [13], and co-embodiment [52]. A retargeting process must map people's motion to this differently proportioned avatar. It is important to understand how the design of this mapping impacts avatar communication.

The two classical approaches for retargeting are forward kinematics (FK) and inverse kinematics (IK). FK applies the joint angles of the person directly to the avatar. This maintains the person's pose on the new character, but the hands and feet may be in different positions relative to the environment if the proportions of the character differ from the user. Conversely, inverse kinematics takes constraints on world space positions, often of the wrists and ankles, and solves for poses that satisfy these. This maintains the position of these end effectors relative to the environment, but changes the overall pose of the character. For instance, when retargeting to a larger character, the character will be constrained to use the smaller reach space of the source. As discussed in Section 2, there is reason to expect that pose changes, e.g. gesture size and spatial extent, resulting from IK may impact the social communication of the character. Conversely, changes in the end effectors may impact the functional communication. While more complicated retargeting solutions exist (Sec. 2.5), they essentially all function by trading off between some mix of pose (FK) and world-space (IK) objectives, so focusing on FK and IK provides an appropriate basis for an evaluation study in the domain, as results on these approaches can be extrapolated to a range of other systems.

The paper presents three experiments that focus on different aspects of communication, and examine the perception of 3rd-person avatars. The first focuses on pointing, which is frequently used to indicate or create reference, a key aspect of communication [10, 43]. The second focuses on iconicity, or the use of gestures to help describe objects [43], with a particular focus on depictions of size. The third uses a political speech that involves little functional information, but provides social cues. In all cases, we evaluate both an avatar that matches the user's proportions and an avatar that is the same height, but with exaggerated, superhero arm and collarbone lengths. This serves a test exemplar that is a reasonable point on the stylization spectrum:

- *Simbarashe Nyatsanga is with University of California, Davis. E-mail: simnyatsanga@ucdavis.edu*
- *Doug Roble is with Meta Reality Labs. E-mail: droble@meta.com.*
- *Michael Neff is with Meta Reality Labs and University of California, Davis. E-mail: mneff@meta.com.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

clearly exaggerated, but still roughly human scale. Both FK and IK reconstructions are compared. For pointing, we also introduce Semantic IK, an approach that offers a pose closer to that of FK while adjusting the aim direction of the index finger to maintain pointing accuracy. We explore both social cues, here measuring personality as one aspect of social communication, and functional cues. We also look at the impact of people's perception of social cues in a high cognitive load task in Experiment 2 compared with a lower load task in Experiment 3.

The paper offers the following contributions:

- A study of near field pointing that establishes accuracy levels and demonstrates that IK reconstructions of an enlarged character can match baseline accuracy.
- The introduction of Semantic IK as an approach to balance information trade offs and demonstration that this provides comparable pointing accuracy.
- An investigation of how size cues are read across retargeting approaches.
- Evidence that social signals vary between IK and FK retargeting approaches. Whether people notice the signals depends on other task factors. The changes are small on the avatar scales tested here, but may be more significant on more extreme avatars or more muted on smaller proportion changes.

2 Related Work

Given our interest in studying the impact of varying limb length on both functional and social communication, we highlight previous work that tackles different aspects of the problem. First, we discuss studies that investigate two types of communicative motion: gesture and pointing. Then we discuss varied ways motion is adapted to changing character proportions.

2.1 Gesture

Co-verbal gestures are hand movements that accompany speech and can vary in both form and function. Following McNeill [43], they are commonly categorized as beats, deictics, iconics, and metaphorics, and may manifest the properties of multiple categories simultaneously. Beats are simple rhythmic movements that match the rhythm of the speech. Deictics are pointing gestures that often use the index finger to create references. Iconics create concrete representations, while metaphorics represent abstract concepts through the use of metaphor. Of all these, deictics are likely to lose their original concrete reference, while iconics and metaphorics are likely to have their concrete and abstract concepts distorted due to changes in limb length. Clark [10] argues that communication provides three main functions: demonstrating, indicating and describing, all of which are strongly influenced by the use of deictic and iconic gestures. Indeed, in a study of embodied VR where participants discussed a floorplan [60], about 60% of gestures were making reference, over 25% created representations and about 20% provided spatial or distance information (a single gesture may do more than one of these things). This suggests that meaning-carrying gestures, which are likely to be affected by varying limb length, are common when discussing physical artifacts like maps or floorplans.

2.1.1 Gestures and Social Communication

Gesture plays an important role in social communication, particularly for affect and attitude, which are often not conveyed verbally [2, 7, 11, 18, 32, 64]. In this paper, we focus on personality as an exemplar social signal where there is good evidence that the kinds of spatial variations in gesture form caused by retargeting are likely to cause a shift in perceived personality. For example, the size [33, 36, 54] and spatial extent [17, 36] of gestures are both related to the perception of extraversion. Work on animation models has manipulated the perception of extraversion by varying movement parameters including stroke scale, position, duration and the swivel angle positioning the elbow [46]. Neff et al. [45] manipulated Emotional Stability through edits that included stroke scale. Smith and Neff [59] found that positional changes impacted extraversion, stroke scale affected extraversion and openness to experience, arm swivel influenced agreeableness and emotional stability, and average velocity affected agreeableness, emotional stability, and extraversion. For emotion, Castillo and Neff [8] found that gesture height and stroke length impacted arousal. Taking all these factors together, it is reasonable to expect that retargeting that varies spatial parameters of gestures may impact the social perception of a character.

For this work, we adopt the five factor model of personality [44, 47, 49], also referred to as the OCEAN model, which has strong support within social psychology. It consists of the traits Extraversion, Agreeableness, Emotional Stability, Conscientiousness and Openness to experience. Each exists on a bipolar scale. We employ the Ten Item Personality Inventory to rate personality in these studies [20]. This is a compact, validated scale that only requires ten ratings and has been commonly used in computational settings to rate communication [38, 45, 46].

2.2 Pointing

Pointing is the most richly studied aspect of gesture for VR interaction. It establishes reference without the need for a verbal description of the referent, thus leading to efficient communication. It is also an important interaction modality in VR, where people may point at world-space menus, widgets, etc. This has led to significant research on the accuracy of pointing, perceptions of pointing, and modifications to pointing behavior.

Pointing can be divided into two categories: one is variously called distal [50], distant [40,65], or mid-air pointing [57]. It involves pointing at targets a few meters from the subject, normally with a fully extended arm and either an index finger for specific targets or an open hand or thumb when the referent is not the primary focus or topic of discourse [57]. Proximal pointing, conversely, involves referents roughly within arm's reach. Distal pointing has been much more extensively studied, likely because of its use as an interaction paradigm and the limited variation in how it is performed. To understand how retargeting might impact communication conveyed by pointing, we discuss previous studies that investigated distal and proximal pointing accuracy.

2.2.1 Distal pointing

Systematic error has been observed in how people perform distal points, both in real life [16] and with avatars in VR [57]. Observers tend to interpret the pointing gesture as a ray extending either from the index finger or the forearm-to-finger complex, while the pointer generally perceives the point as a ray coming from the dominant eye through the tip of the index finger [57]. This has implications for how to perform embodied ray casting and how to create gestures that communicate effectively.

Wong and Gutwin [65] conducted a pair of empirical studies on distal pointing, examining the accuracy of both the pointer and the observer. The first was an informal, observational study where subjects pointed at distant objects through a window. Results suggested participants used an extended arm and index finger, circling when indicating an area; pointing was clear for isolated targets, but greater precision was required for grouped targets; and observers tend to look at the referenced location, not the person pointing. A second, controlled study was done in a collaborative virtual environment (CVE) – a 2D display of a low-fidelity, 3D cylinder-based avatar controlled by a mouse. They found the pointer was more accurate than the observer, although both exhibited errors. Remarkably, although there was greater error in interpreting the pointing direction in the CVE than in the real world, the difference was relatively small, even with such a primitive CVE implementation.

A body of work has explored correcting errors made by distal pointers so that the intention can be more accurately interpreted. Mayer et al. [42] investigated the accuracy of three ray-casting techniques and found a general preference for an Eye Finger Ray Cast (EFRC), with the ray from between the two eyes to the tip of the index finger. They could reduce the error by fitting a polynomial function to the errror data. Mayer et al. [41] replicated this study in both VR and real life, where EFRC showed significantly lower error overall.

Schwind et al. [57] explored the impact of self-avatar embodiment on distal pointing accuracy by embodying participants in a range of avatars, including realistic human, robot, cartoon, and abstract representations, and having them point at a set of targets. The robot and abstract representations performed better, perhaps due to clearer finger geometry, while the cartoon performed worse, likely due to difficulty in perceiving depth cues because of its cel shader. Benda and Ragan [5] used a desktop viewing setup to investigate how avatar visualization impacted the perception of pointing. Using avatars with varying body part visibility, ranging from head and hands to full body, of both realistic and humanoid avatars, they found only small and often unexpected effects; for instance, the head and hands configuration performed better on some measures. The realistic avatar generally performed better than the humanoid one. The greatest impact on accuracy was the distance of the observer from the pointer, with accuracy decreasing as the distance increased.

Several alternative interfaces have also been proposed to improve pointing accuracy [9, 48, 55, 58, 62, 66].

2.2.2 Proximal pointing

Proximal pointing for avatars has received much less research attention. For the related problem of target acquisition, there has been a great deal of research around Fitts' law that relates target acquisition as a function of target distance and size [15, 37]. For pointing, early in-person work by Foley and Held [16] studied near-field pointing where subjects pointed at point lights or thumbtacks with their hand obscured by a horizontal board. They found systematic overreach, which was greater with point lights, indicating the potential impact of depth cues. Pfeiffer et al. [50] studied pointing accuracy across a large table that included both proximal and distal targets. They found that gaze-to-finger (GFP) ray casting was more accurate than index finger ray casting (IFP) for all but the closest row of objects. IFP performed better when additional cues, such as verbal, could be used for disambiguation. The dearth of previous research on proximal pointing accuracy presents a great opportunity, given that conversational agents are likely to perform more gestures within their proximal personal area, with pointing being an elemental part of their vocabulary. Therefore, a pillar of our functional communication study is to investigate proximal pointing accuracy and motion adaptation techniques, in the form of retargeting, to mitigate pointing errors.

2.3 Motion adaptation

Several works have explored modifying avatar motion for more effective communication with observers, which is closely related to our goal of maintaining semantic consistency across varied avatar proportions. Sousa et al. [61] improved pointing accuracy by leveraging the observation that pointers use a vector from the eye-to-fingertip to aim, while observers use a vector following the orientation of the index finger to interpret the point. They fit a Bayesian model to the vertical error which could be used to improve the vertical accuracy for distant pointing. Mayer et al. [40] extended this work, applying both a horizontal and vertical correction by adjusting the shoulder angle of the avatar's straight arm.

Other works have looked at adjusting avatar behavior to provide an informative view of the scene by combining a first-person view of the objects of interest with a view that shows the interlocutor's gaze, gestures, and posture. Hoppe et al. [22] provided users with the same viewpoint on a scene, even when they were standing across a table from each other, in a table and pointer-based interaction without avatars. This approach improved several user ratings—mental demand, temporal demand, effort—and performance, although the response time was longer. Hoppe et al. [23] extended this approach to an avatar system in which each avatar had a first-person view of the objects of interest, but moved to the side in other people's views so that they could read the non-verbal communication. Fidalgo et al. [14] proposed MAGIC, a similar avatar system featuring a first-person view of the scene, but with each avatar on opposite sides of the table. This opposite orientation enabled the application of mirror symmetry to the animation. A user evaluation showed that this approach exceeded face-to-face VR in terms of the correlation between the intended and perceived reference area.

Another interesting option is to shrink the avatar to a much smaller size, as was done in Mini-me [51], where the avatar was reduced to between 5 and 50% of life-size. This is particularly suitable for AR displays with limited fields of view. Gaze and gestures were redirected to match the original users. The user could click to point at a target, and IK would be applied to the miniaturized avatar to point at the same target, with an added raycast laser to improve clarity.

2.4 Skeleton change

Research on varying skeleton proportions has generally focused on the self-avatar. Dewez et al. [12] explored the impact of combined visualizations where arms were shown at different lengths with varied speed control. They found that people generally preferred a single representation, varied arm mappings did not lead to a significant difference in performance, and results for subjective factors like embodiment were limited. Kammerlander et al. [30] studied actors portraying scenes that involved a larger monster and a person, either in VR where the size differences were accurately represented, or on a motion capture stage where both were human sized. They found numerous advantages and disadvantages of both, some related to gaze cues. Our work is focused on changing the relative proportions rather than the actual height of the avatar, so avoids the gaze issues.

2.5 Novel inverse kinematics

Our goal with Semantic IK is to maintain the meaning of gestures across varied embodiments, particularly with changes in arm length. Although there are no previous attempts to solve this exact problem, as far as we know, there has been considerable research into approaches for retargeting motion onto characters with different proportions. Gleicher [19] used a numerical solver to apply spacetime constraints across a motion sequence, modifying the original trajectory with a displacement map to maintain the integrity of critical motion aspects while adapting to a new character. Kulpa et al. [34] subdivides the skeleton into subparts like arms and the trunk, combining Cartesian and angular data without complex inverse kinematics. Their system stores only essential joint angles and normalized Cartesian data to facilitate easy scaling across different character proportions. Hecker [21] proposed an authoring tool that enabled the recording of motion in a morphology-independent format that retains its structural and stylistic elements, later adapted to specific and radically different characters, providing pose goals for an efficient inverse kinematics solver.

Deep learning-based retargeting approaches have been proposed such as novel differentiable operators that can map different skeletons represented as homemorphic graphs in a common latent space, facilitating retargeting by transforming and adapting motion data to and from this unified representation [1]; predicting full-body poses from head and hand motions by using a Transformer encoder to extract features and decouple motion dynamics, then refining arm joint positions through inverse kinematics to match motion capture quality [27]; translating a user's deictic motion into the virtual avatar's corresponding deictic motion, using a network that can map joint angular states into latent representations and adapt them to the avatar's pose based on the user's scale [31]; or synthesizing high-quality continuous motion from six tracking devices by learning a motion manifold using a convolutional autoencoder and employing a learned IK component to adjust the hands and feet toward the corresponding trackers [53].

Other interesting and relevant approaches involving environmental constraints include a motion planning framework for coordinating whole-body actions and navigating obstacles for demonstration tasks [25], and viewpoint dependent animation warping through the specification of visual motion features such as visibility, or spatial extent [29]. These share our goal of preserving semantics, but focus on a different aspect of the problem.

Some notable commercial systems attempt to make animations perform well across a diverse set of characters. In the Sims [3] generic motion, e.g. "gender agnostic" acting, is adapted to avatars with significantly different shapes where mesh penetration can be a problem.

They developed a relative IK approach called *slotting*, which places markers at various points on the character's surface, called slot joints. Animations are then defined relative to these slots; for instance, the hand can be parented to a slot such that as the character's mesh becomes larger or thinner, the attached slot moves accordingly, automatically adjusting the arm motion.

Ubisoft developed a flexible system to retarget motion across a range of characters [6]. The approach converts animation into an IK proxy format; for instance, a leg can be represented by the position of the hip, foot, and the direction of the knee. These controls can then be mapped to a new character and adjusted as needed.

Inspired by this work, our work seeks to understand what is required to maintain meaning across varied arm lengths. For functional communication, we focus on maintaining accuracy in proximal pointing using the proposed Semantic IK. For social communication, our experiments aim to understand the impact of varied character proportions on personality characteristics, using FK and IK retargeting.

3 Experimental Methods

Three experiments were run to explore the impact of changing character proportions on different aspects of communication: deictic accuracy, perception of size cues and social cues. The three experiments were part of a single VR session with a brief break in between each. Particulars are described below and in the following sections. Before undertaking the research, ethics approval was obtained from Advarra IRB. All participants provided consent before participating.

3.1 Experimental Design

Participants attended a single session in which they completed a VR experience that included all three experiments. The experience began by collecting demographic data. Each experiment then consisted of watching a character perform a series of motions followed by questions. All questions were asked and answered within the VR experience using displayed dialogs. Experiments 1 and 3 used a within subjects design. Experiment 2 was effectively between subjects. Details on each experiment are contained in Sections 4 to 6.

Apparatus

The VR experience was developed in Unity 2022.3.10f1 and presented using a Meta Quest 3 head mounted display. The Quest 3 has a resolution of 2064x2208 pixels per eye, with a 110 degree horizontal and 96 degree vertical resolution. A Touch controller was used to interact with all dialogs by employing a virtual laser pointer. People viewed a single character in a virtual environment that showed a plain room, and included a table between the user and the character in Experiment 1. An X was marked on the floor for participants to stand on to ensure a consistent viewpoint. The X is about 1.6m from where the character stands. Please see the supplemental video for stimuli examples.

Model

The study used a custom designed avatar, in a slightly cartoony style, that could vary upper limb proportions and bulkiness (Figure 2). The base skeleton was matched to the proportions of our motion capture actor. The elongated version increased the collarbones by 40% and the upper and lower arms by 30%. This was the largest change that still felt believable in our tests. The height was kept constant to avoid confounds. Two sculpts of the character were created, one that was very skinny and one that was very muscular. Blending between these sculpts provides a range of bulkiness. Our experiments used blend weights of 0.25 (fairly thin), 0.5 (average) and 1.0 (very muscular).

For the experiments that contain audio (object size and speech), lip syncing was driven from the audio using an Autodesk Maya plugin [1].

3.2 Demographics

Sixty participants were recruited for the experiment and paid for their time. They ranged in age from 18 to 66, with mean 33.9, sd 13.4. Thirty-one were female, 28 male and one Non-binary/third gender. Two were American Indian/Alaska Native, 16 were Asian/Asian American, 1 was Black/African/African American, 4 were Latin/Hispanic, 6 were Multiple Races, 2 were Pacific Islander/Native Hawaiian and 29 were White/Caucasian. In terms of education, 6 had a Masters, PhD or JD; 38 had a Four-year college degree, 5 had a two-year college degree, 10 some college and 1 some high school. Regarding VR experience, 5 have their own hardware, 14 had used VR four or more times, 28 had used VR 1-3 times and 13 had no prior experience. Fifty two participants were right handed and 8 were left.

[1] https://github.com/joaen/maya-auto-lip-sync

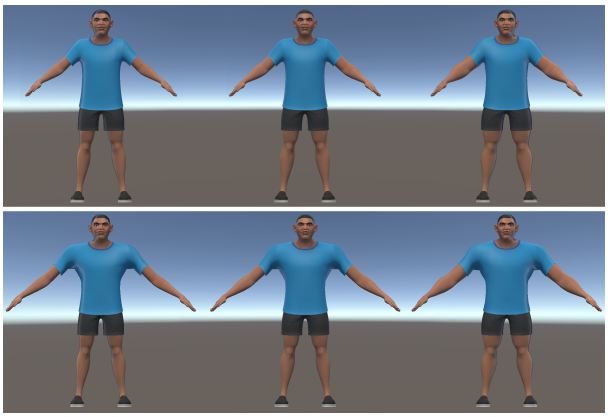

Fig. 2: Avatar used in study: Rows show normal limb length (top) vs. elongated (bottom). The columns show 0.25, 0.5 and 1.0 Bulkiness, from left to right.

4 Exp. 1: Pointing Experiment

The goal of the first experiment was to understand how accurately people can perceive near-range pointing in VR, how different retargeting approaches impact the error in decoding pointing and whether it is possible to define a reconstruction strategy that maintains more natural poses while not degrading the pointing information.

4.1 Stimuli

During the motion capture session, the actor pointed to letters on a regularly spaced, 4x5 grid on a table in front of him. The letter locations were marked by circles placed about 16cm apart in width and depth. The letters were called out in random order to avoid inducing any pattern in the pointing behavior. The actor said "This is [letter]." while pointing to the letter, and then returned to a rest pose. The actor was right handed and all pointing motions were done with his right arm. A careful visual analysis of the avatar motion in the 3D environment, with a visualization of the target locations, showed the actor was very accurate in completing the points. No exact measure of his pointing location was calculated, however.

Five different avatar conditions were used in the experiment, with all using default avatar bulkiness of 0.5. Base uses the skeleton with proportions matched to the actor and all other conditions use the elongated skeleton.

Base: This condition directly uses the joint angles solved from the motion capture session. Our pipeline takes in raw C3D marker data, which we use as constraints to fit a skeleton that matches the actor in size—what we consider the base skeleton—and solve for the subsequent poses. The resulting joint angles are then applied to the base skeleton via FK.

IK: This condition uses the position and orientation of the hand from Base as a constraint and solves for angles in the elbow and shoulder to satisfy this constraint. The position objective function is the Euclidean distance between two 3D points (Eq. 1), and orientation is the F-norm of the element-wise difference between two rotation matrices (Eq. 2). A rotate plane constraint, defined as an element-wise difference between two axes (Eq. 3) is also imposed on the elbow to ensure the shoulder, elbow and hand joints of the elongated skeleton stay in the same plane as in the Base skeleton, preventing unnatural swivel. The rest of the

body is driven by joint angles from the Base skeleton. The resulting IK solution maintains the original hand position and orientation on the elongated skeleton, while applying the rest of the joint angles from Base as is. However, as illustrated in Figure 3b, the IK solution can produce an unnatural over-bending of the wrist.

$$PE(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{3}(q_i - p_i)^2} \tag{1}$$

$$OE(\mathbf{A},\mathbf{B}) = ||(\mathbf{A}-\mathbf{B})||_F \tag{2}$$

$$FAE(\mathbf{a}_{base},\mathbf{a}_{target}) = (\mathbf{a}_{base} - \mathbf{a}_{target}) \tag{3}$$

FK: The joint angles match those from the motion capture recording (Forward Kinematics), but the longer limbs will lead to different hand positions.
FKH: With FK, the longer proportions will cause the hand to penetrate into the table, as illustrated in Figure 3b and c. FKH stays close to the pose of FK, but increases the height of the hands to avoid penetrating the table. It does this by taking the position at the end of the pointing motion with FK as a constraint for inverse kinematics, but increasing the height sufficiently to avoid table penetration. Figure 3c illustrates the FKH pose compared to the FK counterpart.
SemIK: Semantic IK attempts to define a solution that maintains a pose closer to the original (FK), but also maintains the correct semantics of the point. It does this by imposing a position constraint on the wrist at the pointing location that is halfway between the FK and IK position, while also adding an aim constraint on the index finger that forces it to point at the desired target location. The aim constraint is formulated as an element-wise difference between an aim vector (the index finger's local x-axis) and the vector from the finger-tip to the desired target, Eq. 4. The weighted objective function for SemIK is shown in Eq. 5, where α,β,ϕ,γ are adjustable weights set to 0.1. Figure 3b illustrates a SemIK pose compared to the IK and FK counterparts. For the final motion, the FK and SemIK solutions are blended in and out around the semantically salient temporal regions, obtained by manually annotating the temporal points in the Base motion, using spherical linear interpolation, as shown in Figure 4. All the retargeting solutions are implemented using the Momentum library [2].

$$AE(\mathbf{v}_{aim},\mathbf{v}_{direction}) = (\mathbf{v}_{aim} - \mathbf{v}_{direction}) \tag{4}$$

$$\theta = \arg\min_{\Theta}\left[\alpha\cdot PE + \beta\cdot OE + \phi\cdot FAE + \gamma AE\right] \tag{5}$$

4.2 Experiment Details

Each avatar condition was presented one at a time, randomized across participants. For each condition, all 20 pointing animations were shown, again in random order. In each animation, the avatar pointed at a location on a plain gray table without the reference points that were available to the actor. After viewing each pointing motion, the participant used a laser pointer to indicate the position on the desk they believed the character was pointing to. They were then given the option to confirm or redo their selection in case they misplaced the marker. The animation would not replay if they redid the placement. Their final location was logged. The VR controller was disabled while the animations played to ensure that the participants did not try to follow the characters motion with the laser pointer. After all the animations played for a given avatar condition, participants were asked to rate their agreement with the statement "The motion in the previous clips appeared natural." on a seven-point Likert scale, ranging from Strongly Disagree to Strongly Agree.

[2] https://facebookincubator.github.io/momentum/

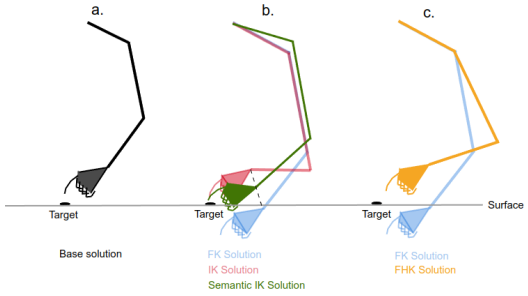

Fig. 3: Illustration of retargeting solutions used in our experiments. **Base** and **FK** directly use joint angles from the motion capture session, however the latter penetrates the surface. **FKH** alleviates surface penetration but introduces lateral error. **IK** maintains hand position and orientation, and pointing accuracy, but results in an unnatural wrist bend. **SemanticIK** maintains a more natural overall pose, like **FK**, while preserving pointing accuracy, like **IK**.

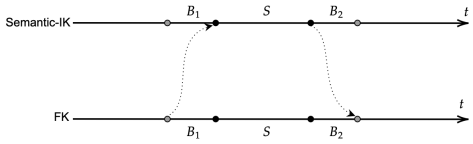

Fig. 4: Illustration of how the blending of FK and SemanticIK solves for the final SemanticIK motion. S represents the semantically salient region where pointing occurs. B_1 and B_2 are the buffer regions used to ease in and out of the SemanticIK solution, and t denotes time.

4.3 Results

A scatterplot of participants' perception of the pointing location for each pointing target is shown in Figure 5. The average error levels for each avatar condition are summarized in Figure 6. Mean error by condition was: Base $\mu = 0.0496m, SD = 0.0354$, FK $\mu = 0.129m, SD = 0.0337$, FKH $\mu = 0.135m, SD = 0.0581$, IK $\mu = 0.0486m, SD = 0.0352$, SemIK$\mu = 0.0416m, SD = 0.0297$. This shows that, as expected, FK reconstructions on large skeletons can have high error, large enough in this particular case to likely cause incorrect or unclear references in many practical scenarios. To provide a simpler view of the data, Figure 7 shows the distance from the actual target to the mean of the estimates. The larger errors for FK and FKH reconstructions are clear. There is also an interesting shift in the orientation of the FKH error, discussed below.

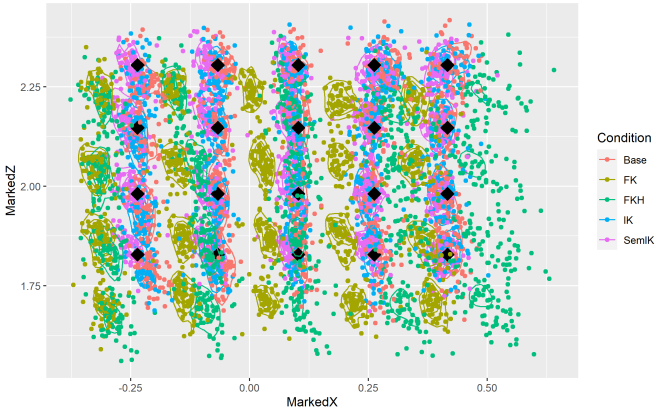

Fig. 5: Exp. 1: A scatter plot of the estimated point locations around each point target. The image represents the horizontal plane of the table. The actor stood at the top of the figure and the participant at the bottom, both centered, just outside the table.

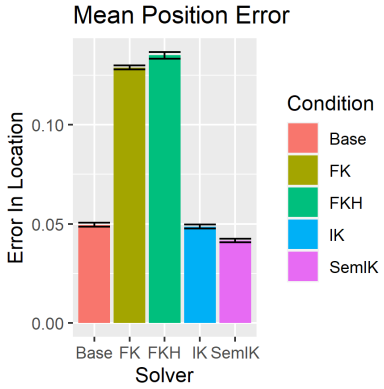

Fig. 6: Exp. 1: The mean position error for each avatar condition. All conditions are significantly different, except for Base and IK.

Condition	Max Mean Error	Min Mean Error
Base	0.0620	0.0379
FK	0.152	0.107
FKH	0.171	0.0915
IK	0.0598	0.0358
SemIK	0.0544	0.0352

Table 1: Exp 1: Min and max average errors in meters.

An ANOVA shows that the error performance was significantly different across conditions $F_4 = 315.4, p < .0001$. Post-hoc analysis was done using paired t-tests with Bonferroni correction. All conditions were significantly different (p=.0048 for FK vs. FKH; p<.0001 for all others), except Base and IK (p=1.000). The lowest error was for SemIK, followed by {Base, IK}, then FK and finally FKH.

Pointing error was remarkably consistent across the different target locations, as can be seen in the figures. Min and max average errors are in Table 1

A linear mixed effects model was fit to the data with factors Condition x TargetLocation. Using expected means for post-hoc analysis with Tukey correction and grouping by condition showed that the only condition where the error varied for different letters with any consistency was FKH. In all other cases, performance was statistically similar for all (or almost all) locations.

Naturalness ratings for each avatar condition are shown in Figure 8.

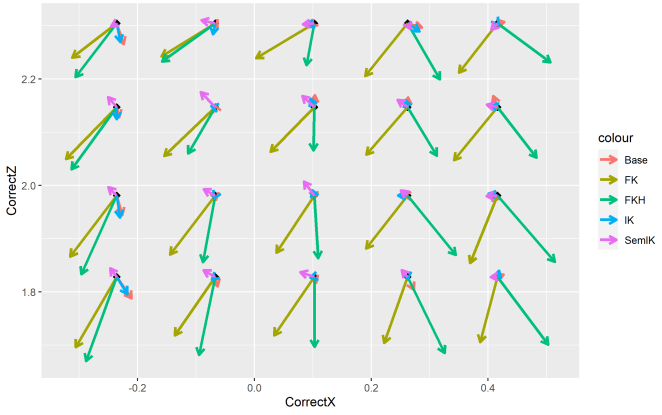

Fig. 7: Exp. 1: Arrows indicate the distance from the actual pointing target to the center of participants estimate of the point location.

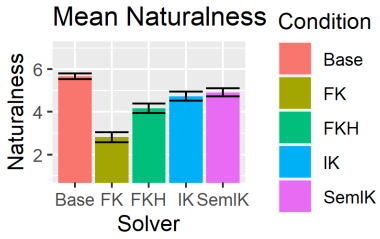

Fig. 8: Exp. 1: Naturalness ratings for the different avatar conditions.

An ANOVA shows that there are significant differences between the ratings ($F_4 = 40.45, p < .0001$). Post-hoc analysis was done with paired t-tests with Bonferroni correction. Base outperformed all other conditions (FK, $p < .0001$; FKH, $p < .0001$; IK, $p < .001$; SemIK, $p = .0023$). SemIK and IK were next and not signficantly different from each other. Both outperformed FK ($p < .0001$ for both). SemIK was significantly better than FKH ($p = .021$), but IK was not ($p = .15$). FKH was seen as significantly more natural than FK ($p < .0001$).

4.4 Discussion

This experiment shows that people can read pointing quite accurately, to an average precision of about 5 cm (2 inches), with no visualized reference points beyond a featureless grey table. The error is consistent for motion matched to the actor, with the same skeleton proportions, and also for IK reconstructions on much larger skeletons that only match the hand pose. It is also possible to create Semantic IK that moves away from the recorded hand position, but maintains the same accuracy. Accuracy was consistent for these measures across the tested reference area. FK reconstructions on exaggerated skeletons produce noticeable error (over 12 cm on average for our height matched, but long limbed skeleton).

Since IK and Base yielded statistically similar results, it appears that the pointing information is fully contained in the orientation of the hand, which is identical for each. Specifically, it seems that the index finger is read as a vector by observers, so its orientation is key. This seems to explain the changing location of the FKH estimates. As the right handed character changes from pointing targets that are in front of his arm to those that increasingly require him to reach across his body, his index finger moves from being almost perpendicular to his chest to parallel (from the sagittal to the coronal plane). Since the hand is moved up to avoid the table, this will rotate the finger up, creating the appearance that it is referencing a point further to the avatar's left.

The Semantic IK condition demonstrates that it is possible to build a retargeting approach that maintains body poses that are closer to the user's true pose while also providing very clear references. It likely outperformed the Base condition because it had the positions of the physical target centers as input and the actor may have made some small errors in his motion. The challenge with a semantic approach is knowing the information that needs to be conveyed, preferably ahead of time so that the motion can be smooth. For non-player characters, this is likely feasible. For avatars driven in realtime, a prediction module would be required.

The fact that Base was rated more natural than all other conditions is likely related to people finding the very long limbed model used in the remaining conditions less natural. The FK condition has good quality motion, but the hand penetrates the table, likely explaining the low Naturalness. The current implementation of FKH blends into an IK solve for the final pointing pose. The timing of these blends seems slightly off, which leads to some unnatural accelerations and likely the drop in Naturalness.

5 Exp. 2: Object Size Experiment

Iconic gestures use the hands to create an imagistic representation of a concrete thing, such as an object. They can convey objective information, such as object size. This experiment explores how that

objective information is read across different avatar retargeting approaches. Specifically, we show people gesture sequences in which the size of an object is indicated only using gesture and ask them to then estimate the object width. As a secondary task, we also ask them to estimate the personality of the character for each avatar condition. Given that the size rating task has fairly high cognitive load, and is emphasized in the instructions, we anticipate people will be less sensitive to the personality signal in this experiment than Experiment 3. If this holds, it will provide guidance on when care must be taken to preserve social cues.

5.1 Stimuli

We had the actor improvise eight different stories in which he discussed three different items he found while antique shopping. Each item had a different size, specified to the actor ahead of time – small, medium or large – and the order of these sizes was varied. While motion captured, the actor described each item and was instructed to use gesture to provide an indication of its size. He was told to pick items that could be any size, for example, a picture frame or a box, to avoid providing a size indication by the item type. He was also instructed not to verbally mention the size. Based on a review of video of the session, the researchers picked three sequences that best achieved these goals and used the motion capture from them to create stimuli. To be clear, no objects were ever shown. There were only gestures that indicated their size.

This experiment used three of the avatar conditions from Exp. 1, **Base**, **FK** and **IK**. **SemIK** and **FKH** solutions depend on manually annotated timestamps of relatively simple pointing animations. Due to the complexity of annotating iconic gestures depicting size and shape, we excluded these solutions from Experiment 2.

5.2 Experiment Details

Since there are three different stories, and three different avatar conditions, there are six possible combinations of avatar and story for which participants hear each story once. We counterbalanced to sample these evenly. The presentation order was randomized. Each story was broken into three clips, each describing an object. These clips were shown in order to maintain the original story.

After each clip describing one object, the participant was asked to estimate the size of the object by moving a pointer on a slider. The slider was scaled accurately and had markings at every foot and half foot, indicated in inches. After seeing all three stories in a condition, the participant completed the Ten Item Personality Inventory (TIPI) [20], a validated instrument for measuring the five factor personality model.

In summary, every participant saw all three stories, but they only saw one retargeting condition per story. They provided nine size estimates, one for each object described, and answered three TIPI surveys, one per story.

5.3 Results

The size estimates are shown in Figure 9. The original goal for this study was for the actor to present equal sized, small, medium and large items in each story. This proved impractical, however. Our actor provided a high quality, naturalistic performance that kept to the general size categories, but there was some variation in the precise size. Therefore, rather than averaging over object size as a factor when doing the analysis, we opted to analyze each of the nine objects separately. This data is between subjects and there are twenty samples per object-retargeting strategy combination for each of the nine objects. This is relatively low power for a between subjects design. Nonetheless, ANOVAs fit to the data for each object, followed by pairwise comparisons, show that for all the Large objects (main diagonal), the FK motion on a larger skeleton was perceived as showing a significantly larger item than either the base motion or IK retarget. This was also the case for one of the medium items. For another medium item, there was no significant difference between the FK and IK retargets on the large skeleton and for the last medium item, this differences only tended toward significance (p=.09). For the small items, the FK retarget of

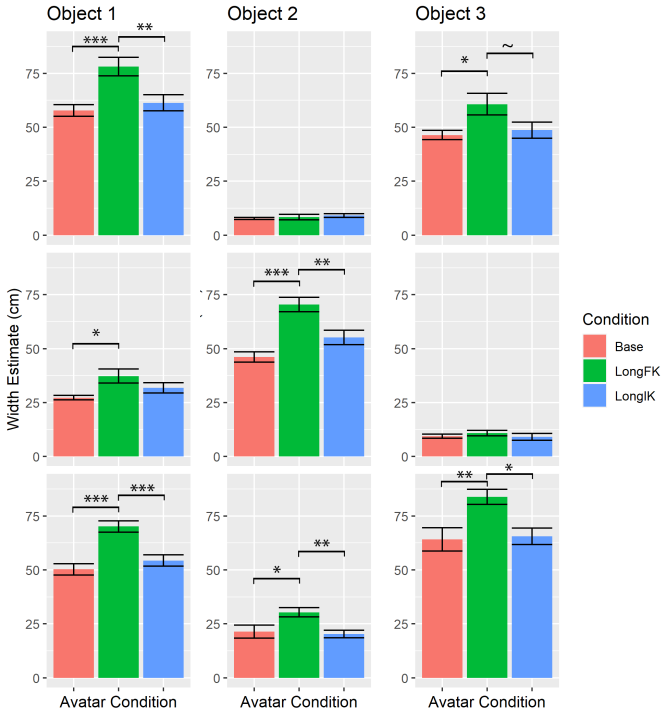

Fig. 9: Participants' estimates of object size, across conditions. The rows correspond to the three stories, columns correspond to the order each item was described in. The actor was instructed to use the following baseline object sizes: soft-ball size for small, 30 cm for medium and shoulder width for large.

the long skeleton led to significantly larger estimates in one case, but there were no significant differences in the other two. These results will be discussed below. The significance levels are shown graphically in Figure 9.

ANOVAs were used to test each personality measure. In no case did they show a significant difference, so people did not read different personality cues based on the avatar conditions during this task.

5.4 Discussion

People do read gestural size indications and, as anticipated, the FK reconstruction on an elongated character led to larger size estimates. Size information varies based on retargeting approach. Considering the large and medium object sizes, estimates were about 40% larger, which is similar to the scale up in the collarbones.

The two small objects that showed no significant difference for reconstruction were a (mimed) ball and pocket mirror. The actor almost entirely indicated these with one hand, so the retargeting change had no impact on item size. The small item that did show variation was a small pillow. Here, the actor used a combination of one and two handed gestures and also placed it between his head and neck. There are thus size indications that involve both hands (or a hand and body) that can be influenced by retargeting and those that are provided by the fingers and are not impacted if hands are not varied. We could have disallowed one handed gestures, which likely would have led to significant variation in all size categories, but preferred to allow the actor to use the indication that felt most natural. We think the one handed results provide an interesting insight on the complexity of size indications.

The IK reconstruction on the longer body generally has an average estimate slightly larger than the Base motion. This difference was never significant here, but studies with higher power might show that people slightly increase their size estimates based on the overall size of the character.

We see that no social cues, in terms of personality at least, were

	BulkyBase	BulkyLong	ThinBase	ThinLong
FK	✓	✓	✓	✓
IK		✓		✓

Table 2: The six avatar conditions for the Experiment 3.

perceived in this experiment based on retargeting. This is likely due to both the strong prior provided by actors consistent audio and the high cognitive load involved in needing to remember object sizes to complete the survey. It may also be that the type of gestures employed led participants to focus on the avatar's hands rather than the overall pose. Social cues may be less of a concern in situations like this, unless the social cues are designed to directly impact the experience.

Finally, this study introduces another type of information that would need to be maintained in a general approach to Semantic IK. Size indications often involve the palms, but a switch could be made to using the finger tips, which would allow the wrist positions to be expanded out to something that would be closer to what you would get with the FK retarget.

6 Exp 3: Speech Experiment

This experiment is focused on social communication and seeks to understand to what degree retargeting choices may impact how the social qualities of interlocutors in VR are read. In stylized character design, both the proportions and bulk of characters are often changed, for example, to create very muscular superhero characters. As both factors may impact social communication, both are modulated here.

6.1 Stimuli

For this experiment, the actor performed Marc Antony's "Friends, Romans, Countryman" speech from the play Julius Ceasar by William Shakespeare. Rather than a traditional performance that might vary for dramatic purpose over the speech, the actor was directed to maintain a consistent tone throughout so that all parts of the performance had a similar quality. The speech was then divided into six approximately equal length segments.

Six avatar conditions were run as specified in Table 2. *Bulky* uses the model blend weight of 1.0 to give a very muscular appearance and *Thin* uses 0.25 to give a thin appearance. *Base* uses the skeleton proportions that match the actor, and *Long* uses the elongated proportions. *FK* uses the joint angles from motion capture and *IK* maintains the captured positions of the wrists, recalculating the arm angles to do so.

6.2 Experiment Details

Participants saw each segment of the speech in the correct order, displayed on one of the six avatar conditions. The ordering of the avatar conditions was randomized.

After each clip, participants would complete the TIPI to rate their perceptions of the personality, as in Exp. 2. They also completed two additional ratings, stating their agreement with "The character is dominant." and "The character is submissive." on seven point Likert scales. These measures are taken from [35]. The ratings are used to compute a Dominance rating by reversing the second measure and averaging the two.

6.3 Results

There are three underlying factors: skeleton length, mesh bulk and retargeting method. Each have two levels, but these could not be fully sampled as there is only one retargeting method for the base skeleton. This allows two sets of comparisons to evaluate the impact of each factor. To compare the impact of the retargeting method and bulk, we can drop the data for the two base skeleton cases. To compare skeleton length and bulk, we drop the IK retargeting. An ANOVA was calculated for each of these for every personality trait and the dominance rating using ezANOVA in R.

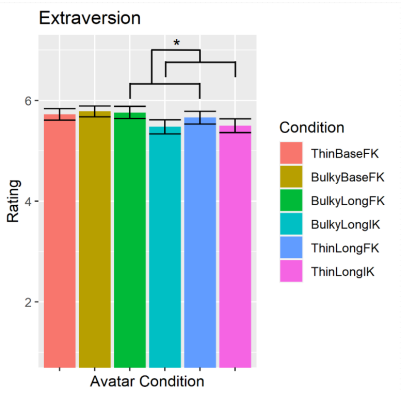

Fig. 10: Exp 3: Mean ratings of Extraversion on the speech task. The last four bars were used in a retargeting method x bulkiness comparison that showed Extraversion was rated higher with FK.

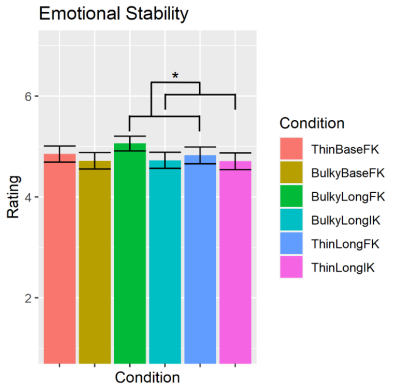

Fig. 11: Exp 3: Mean ratings of Emotional Stability on the speech task. The last four bars were used in a retargeting method x bulkiness comparison that showed Emotional Stability was rated higher with FK.

Three significant differences were found. Extraversion was rated higher with the FK retargeting on the long skeleton vs. IK retargeting ($F_{1,1} = 5.874, p = .018$; $\mu_{FK} = 5.71$, $SD_{FK} = 0.963$; $\mu_{IK} = 5.49$, $SD_{IK} = 1.06$). Neither the body bulk factor nor interaction were significant. The full means are shown in Figure 10. Emotional Stability was also rated higher with the FK retargeting on the long skeleton vs. IK retargeting ($F_{1,1} = 4.208, p = .045$; $\mu_{FK} = 4.94$, $SD_{FK} = 1.22$; $\mu_{IK} = 4.72$, $SD_{IK} = 1.25$). Again, the body bulk factor and interaction were not significant. See Figure 11. Finally, Bulkiness decreased ratings on Openness to Experience ($F_{1,1} = 4.208, p = .032$; $\mu_{Bulky} = 4.59$, $SD_{Bulky} = 0.835$; $\mu_{Thin} = 4.74$, $SD_{Thin} = 0.898$), with no significant impact from skeleton proportion or the interaction of the two factors for the set of FK motions (Figure 12). In all three cases, the effect size was small.

6.4 Discussion

The impacts on Extraversion and Emotional Stability are consistent with the literature. Previous work has shown that larger gestures and more extended body poses are perceived as more extraverted [46, 59]. The IK reconstruction on a long skeleton induces an elbow bend which may look less relaxed or more nervous, which could explain the change in Emotional Stability. The effect on Openness to Experience was not anticipated and is harder to explain. Openness to Experience relates to intellectual curiosity, so one potential explanation is that people are influenced by a "dumb jock" stereotype, which lowers Openness ratings for the very muscular character. There is no direct evidence of this, so it is appropriate to be cautious with the interpretation. Further investigation would be useful.

Overall, it is notable that impact on personality ratings was quite

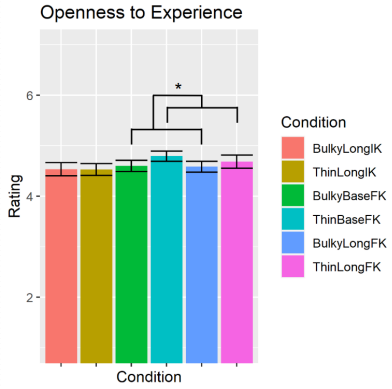

Fig. 12: Exp 3: Mean ratings of Openness to Experience on the speech task. The last four bars were used in a skeleton length x bulkiness comparison that showed Openness to Experience was rated higher with a Thin body model.

muted. We also anticipated that Dominance would be impacted in a similar way as Extraversion, but it was not. Going through the experience, one clear reason for the limited variation is the strong impact of the actor's voice. Some participants commented that they paid attention mostly to the voice. His delivery was clear and very calm throughout, so it may have been difficult to project other personalities onto that strong signal. It may also be that some people are less sensitive to motion cues in general.

The limited impact on social cues may mean that VR experience designers do not need to worry too much about mitigating the issue. That would be positive news. Before concluding this, however, it would be interesting to look at cases where there is a mismatch between clearly emotional vocal tracks and character motion. For example, it might look unnatural for a very large character to yell something out in a very excited voice, but make only a limited range gesture. In tasks that involve character observation without audio, people will rely more on motion cues. The direction of the impact should be similar to what is seen here and we anticipate that the effect will be larger. No doubt, designers will need to balance the motion requirements to the particular application they are building.

The results here are specific to the skeleton variations studied. For example, if a shorter rather than longer skeleton was used, IK might increase the perception of extraversion by leading to larger gestures. Results may also be stronger for much larger skeletons.

7 General Discussion, Limitations and Conclusion

Several insights can be gathered from the studies presented here. Pointing accuracy can be read to a mean error of about 5cm and this is consistent with both avatars matched to the user and larger avatars that perform an IK based reconstruction. Using an FK reconstruction leads to error increases. It appears that the orientation of the index finger is the critical cue in providing near field point accuracy. It is possible to build a semantic IK that moves the hands further away with no loss in accuracy as long as this finger orientation is maintained. Simple approaches, such as keeping a longer limbed avatar's hands above a table, while improving naturalness, will not maintain pointing accuracy.

Size indications can be performed by one or two hands. The perceived size for two handed indications will vary based on the retargeting method, but one handed size indications remain consistent if the finger pose is not changed.

There is some impact on social communication from using IK vs. FK, but this depends on the task of the user. When the cognitive load was high, as in Experiment 2, there was not a consistent change in personality perceptions, but when people were watching a character deliver a speech in Experiment 3, perceptions did change, albeit in modest ways. The degree of shift in these social cues may be stronger if there is less other information, such as the strong audio track present here. They may also be reduced in cases where the avatar proportions are varied less. The amount of consideration that should be given to maintaining social cues should depend on the nature of the application being developed. It would also be valuable to conduct further studies on the social impact of retargeting on varied gesture forms. We used a speech performance here that relied on metaphoric and deictic gestures. Further research could study the impact on specific forms.

We did not use shadows in our pointing experiment as shadows are highly variable. They can be hard or soft, come from multiple directions, one light or many, etc. Prior work shows how properties of shadows can influence perception; for instance, location can affect distance estimates for objects in VR [24], sharpness can affect shape matching of objects in ray-traced images [63], pattern can affect light source distance estimates [56], and shape can improve automatic human pose and shape estimation [4] as well as spatial arrangement recovery [39]. The impact of shadows on pointing warrants a separate study. Our shadowless setting provides baseline data.

While the proportion changes seen here may seem large on a realistic human scale, they are small compared to the kinds of changes people may want for their avatars in VR, e.g. cf. VR Chat [26]. There is no reason a person would not want to play a ten foot monster. As we move to such more extreme avatars, the issues identified here will also become more extreme. There will be a degradation of functional and social communication, in the directions identified. For example, it is reasonable to expect a stronger shift in social cues if a much larger character is limited to the same gesture space. It will be important to balance these social and functional needs, and a flexible, semantic IK with prediction may be one way forward.

Lessons can be drawn from the studies here on how to design a general form of Semantic IK. The needs will vary based on the type of communication required and the scale of the retargeting and involve tradeoffs between IK and FK style reconstructions. Pointing and object size information are tied to the users actual hand positions. For pointing, it was possible to move the wrist constraints 50% of the way back to their FK position with no loss of pointing accuracy as long as the index finger was aimed appropriately. It may be possible to move further with this technique. Object indications often involved palm positions, so are harder to move out without changing the size perception, as was seen in the FK condition. It may be possible to use the finger tips to allow the wrists to be moved while maintaining the same size information. When size information is indicated with one hand, the wrists can be freely moved. Overall, there seems to be a need to fade between more IK and more FK solutions depending on the situation and the need to maintain either social or functional information. While doing this, object penetrations should be avoided as these lead to a drop in naturalness.

There are many more questions to explore in order to understand the communication impact of retargeting. One important area is to look at more extreme character mappings where people are playing avatars much larger or smaller. Another important issue is to look at mismatches between audio and motion. Here, the actors performance was steady and consistent. Social motion cues may be more important in extreme moments where the vocal quality is exaggerated, for example, in a moment of anger. Here limited motion range relative to the character proportions may look more unnatural. Finally, there are other social cues beyond personality that warrant investigation, such as emotion and attitude.

In conclusion, we presented three experiments investigating how retargeting choices affect the perception and communicative efficacy of 3rd-person avatars. Our first experiment on near-field pointing demonstrated that IK reconstructions for an enlarged character could match baseline accuracy, proposing Semantic IK as a potential method to balance information trade-offs while maintaining pointing accuracy. Subsequent studies examined how object size cues are interpreted and revealed variations in social signals between IK and FK retargeting approaches, with the impact of these signals influenced by other task factors. Changes were minor at the avatar scales tested, but might be more pronounced with more extreme avatars or less noticeable with smaller proportion changes.

References

[1] K. Aberman, P. Li, D. Lischinski, O. Sorkine-Hornung, D. Cohen-Or, and B. Chen. Skeleton-aware networks for deep motion retargeting. *ACM Transactions on Graphics (TOG)*, 39(4):62–1, 2020. doi: 10.1145/3386569.3392462 3

[2] L. Alem and J. Li. A study of gestures in a video-mediated collaborative assembly task. *Advances in Human-Computer Interaction*, 2011(1):987830, 2011. doi: 10.1155/2011/987830 2

[3] E. Arts. The Sims Video Games. 3

[4] A. O. Balan, M. J. Black, H. Haussecker, and L. Sigal. Shining a light on human pose: On shadows, shading and the estimation of pose and shape. In *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8. IEEE, 2007. doi: 10.1109/ICCV.2007.4409005 9

[5] B. Benda and E. D. Ragan. The effects of virtual avatar visibility on pointing interpretation by observers in 3d environments. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 50–59. IEEE, 2021. doi: 10.1109/ISMAR52148.2021.00019 3

[6] A. Bereznyak. Ik rig: Moving forward. In *Game Developers Conference (GDC)*. GDC, 2016. 4

[7] S. A. Bly. A use of drawing surfaces in different collaborative settings. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pp. 250–256, 1988. doi: 10.1145/62266.62286 2

[8] G. Castillo and M. Neff. What do we express without knowing?: Emotion in gesture. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 702–710. International Foundation for Autonomous Agents and Multiagent Systems, 2019. doi: doi/10.5555/3306127.3331759 2

[9] J. W. Chastine, K. Nagel, Y. Zhu, and L. Yearsovich. Understanding the design space of referencing in collaborative augmented reality environments. In *Proceedings of graphics interface 2007*, pp. 207–214, 2007. doi: 10.1145/1268517.1268552 3

[10] H. H. Clark. *Using language*. Cambridge University Press, 1996. 1, 2

[11] H. H. Clark and M. A. Krych. Speaking while monitoring addressees for understanding. *Journal of memory and language*, 50(1):62–81, 2004. doi: doi/10.1016/j.jml.2003.08.004 2

[12] D. Dewez, L. Hoyet, A. Lécuyer, and F. Argelaguet. Do you need another hand? investigating dual body representations during anisomorphic 3d manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2047–2057, 2022. doi: 10.1109/TVCG.2022.3150501 3

[13] N. Döllinger, J. Topel, M. Botsch, C. Wienrich, M. E. Latoschik, and J.-L. Lugrin. Exploring agent-user personality similarity and dissimilarity for virtual reality psychotherapy. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 424–427. IEEE, 2024. doi: 10.1109/VRW62533.2024.00082 1

[14] C. G. Fidalgo, M. Sousa, D. Mendes, R. K. Dos Anjos, D. Medeiros, K. Singh, and J. Jorge. Magic: Manipulating avatars and gestures to improve remote collaboration. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 438–448. IEEE, 2023. doi: 10.1109/VR55154.2023.00059 3

[15] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954. doi: 10.1037/h0055392 3

[16] J. M. Foley and R. Held. Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics*, 12:263–268, 1972. doi: 10.3758/BF03207201 2, 3

[17] K. Frank. Posture & perception in the context of the tonic function model of structural integration: An introduction. *IASI Yearbook*, 2007:27–35, 2007. 2

[18] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309, 2004. doi: 10.1207/s15327051hci1903_3 2

[19] M. Gleicher. Retargetting Motion to New Characters. In *Proceedings of the 25th Annual Conference on Computer graphics and Interactive techniques*, pp. 33–42, 1998. doi: 10.1145/280814.280820 3

[20] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003. doi: 10.1016/S0092-6566(03)00046-1 2, 7

[21] C. Hecker, B. Raabe, R. W. Enslow, J. DeWeese, J. Maynard, and K. Van Prooijen. Real-time motion retargeting to highly varied user-created morphologies. *ACM Transactions on Graphics (TOG)*, 27(3):1–11, 2008. doi: 10.1145/1399504.1360626 3

[22] A. H. Hoppe, F. van de Camp, and R. Stiefelhagen. Personal perspective: Using modified world views to overcome real-life limitations in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 577–578. IEEE, 2018. doi: 10.1109/VR.2018.8446311 3

[23] A. H. Hoppe, F. Van De Camp, and R. Stiefelhagen. Shisha: Enabling shared perspective with face-to-face collaboration using redirected avatars in virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–22, 2021. doi: 10.1145/3432950 3

[24] R. L. Hornsey and P. B. Hibbard. Distance mis-estimations can be reduced with specific shadow locations. *Scientific Reports*, 14(1):9566, 2024. doi: 10.1038/s41598-024-58786-1 9

[25] Y. Huang and M. Kallmann. Planning motions and placements for virtual demonstrators. *IEEE transactions on visualization and computer graphics*, 22(5):1568–1579, 2015. doi: 10.1109/TVCG.2015.2446494 3

[26] V. C. Inc. VR Chat. 1, 9

[27] J. Jiang, P. Streli, H. Qiu, A. Fender, L. Laich, P. Snape, and C. Holz. Avatarposer: Articulated full-body pose tracking from sparse motion sensing. In *European conference on computer vision*, pp. 443–460. Springer, 2022. doi: 10.1007/978-3-031-20065-6_26 3

[28] A. Jovane, D. Egan, M. F. Vargas, S. Mythen, and R. McDonnell. Virtual animal embodiment for actor training. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 428–431. IEEE, 2024. doi: 10.1109/VRW62533.2024.00083 1

[29] A. Jovane, P. Raimbaud, K. Zibrek, C. Pacchierotti, M. Christie, L. Hoyet, A.-H. Olivier, and J. Pettré. Warping character animations using visual motion features. *Computers & Graphics*, 110:38–48, 2023. doi: 10.1016/j.cag.2022.11.008 3

[30] R. K. Kammerlander, A. Pereira, and S. Alexanderson. Using virtual reality to support acting in motion capture with differently scaled characters. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 402–410. IEEE, 2021. doi: 10.1109/VR50410.2021.00063 3

[31] J. Kang, D. Yang, T. Kim, Y. Lee, and S.-H. Lee. Real-time retargeting of deictic motion to virtual avatars for augmented reality telepresence. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 885–893. IEEE, 2023. doi: 10.1109/ISMAR59233.2023.00104 3

[32] D. Kirk and D. Stanton Fraser. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1191–1200, 2006. doi: 10.1145/1124772.1124951 2

[33] M. L. Knapp, J. A. Hall, and T. G. Horgan. *Nonverbal communication in human interaction*, vol. 1. Holt, Rinehart and Winston New York, 1978. 2

[34] R. Kulpa, F. Multon, and B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. In *Eurographics*, 2005. doi: 10.1111/j.1467-8659.2005.00859.x 3

[35] B. Lance and S. C. Marsella. The relation between gaze behavior and the attribution of emotion: An empirical study. In *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings 8*, pp. 1–14. Springer, 2008. 8

[36] R. Lippa. The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis. *Journal of Research in Personality*, 32(1):80–107, 1998. doi: 10.1006/jrpe.1997.2189 2

[37] I. S. MacKenzie. Fitts' law. *The wiley handbook of human computer interaction*, 1:347–370, 2018. 3

[38] F. Mairesse and M. A. Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37(3):455–488, 2011. doi: 10.1162/COLI_a_00063 2

[39] P. Mamassian, D. C. Knill, and D. Kersten. The perception of cast shadows. *Trends in cognitive sciences*, 2(8):288–295, 1998. 9

[40] S. Mayer, J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze. Improving humans' ability to interpret deictic gestures in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. doi: 10.1145/3313831.3376340 2, 3

[41] S. Mayer, V. Schwind, R. Schweigert, and N. Henze. The effect of offset correction and cursor on mid-air pointing in real and virtual environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3174227 2

[42] S. Mayer, K. Wolf, S. Schneegass, and N. Henze. Modeling distant pointing for compensating systematic displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4165–4168, 2015. doi: 10.1145/2702123.2702332 2

[43] D. McNeill. *Gesture and thought*. University of Chicago Press, 2005. 1, 2

[44] M. R. Mehl, S. D. Gosling, and J. W. Pennebaker. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, 90(5):862, 2006. doi: 10.1037/0022-3514.90.5.862 2

[45] M. Neff, N. Toothman, R. Bowmani, J. Fox Tree, and M. Walker. Don't scratch! self-adaptors reflect emotional stability. In *Intelligent Virtual Agents*, pp. 398–411. Springer, 2011. doi: doi/10.5555/2041666.2041717 2

[46] M. Neff, Y. Wang, R. Abbott, and M. Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *Intelligent Virtual Agents*, pp. 222–235. Springer, 2010. doi: doi/10.5555/1889075.1889103 2, 8

[47] W. T. Norman. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The journal of abnormal and social psychology*, 66(6):574, 1963. doi: doi/10.1037/h0040291 2

[48] O. Oda and S. Feiner. 3d referencing techniques for physical objects in shared augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 207–215. IEEE, 2012. doi: 10.1109/ISMAR.2012.6402558 3

[49] J. W. Pennebaker and L. A. King. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296, 1999. 2

[50] T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth. Conversational pointing gestures for virtual reality interaction: implications from an empirical study. In *2008 IEEE Virtual Reality Conference*, pp. 281–282. IEEE, 2008. doi: 10.1109/VR.2008.4480801 2, 3

[51] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. doi: 10.1145/3173574.3173620 3

[52] I. Podkosova and H. Brument. Towards full body co-embodiment of human and non-human avatars in virtual reality. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 432–435. IEEE, 2024. doi: 10.1109/VRW62533.2024.00084 1

[53] J. L. Ponton, H. Yun, A. Aristidou, C. Andujar, and N. Pelechano. Sparseposer: Real-time full-body motion reconstruction from sparse data. *ACM Transactions on Graphics*, 43(1):1–14, 2023. doi: 10.1145/3625264 3

[54] R. E. Riggio and H. S. Friedman. Impression formation: The role of expressive behavior. *Journal of personality and social psychology*, 50(2):421, 1986. doi: doi/10.1037/0022-3514.50.2.421 2

[55] J. Schjerlund, K. Hornbæk, and J. Bergström. Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021. doi: 10.1145/3411764.3445759 3

[56] H. H. Schuett, F. Baier, and R. W. Fleming. Perception of light source distance from shading patterns. *Journal of Vision*, 16(3):9–9, 2016. doi: 10.1167/16.3.9 9

[57] V. Schwind, S. Mayer, A. Comeau-Vermeersch, R. Schweigert, and N. Henze. Up to the finger tip: The effect of avatars on mid-air pointing accuracy in virtual reality. In *Proceedings of the 2018 annual symposium on computer-human interaction in play*, pp. 477–488, 2018. doi: 10.1145/3242671.3242675 2, 3

[58] J. Simões, A. Maciel, C. Moreira, and J. Jorge. Sparc: Shared perspective with avatar distortion for remote collaboration in vr. *arXiv preprint arXiv:2406.05209*, 2024. 3

[59] H. J. Smith and M. Neff. Understanding the impact of animated gesture performance on personality perceptions. *ACM Trans. Graph.*, 36(4):49:1–49:12, article no. 49, 12 pages, July 2017. doi: 10.1145/3072959.3073697 2, 8

[60] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2018. doi: 10.1145/3173574.3173863 2

[61] M. Sousa, R. K. dos Anjos, D. Mendes, M. Billinghurst, and J. Jorge. Warping deixis: distorting gestures to enhance collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019. doi: 10.1145/3290605.3300838 3

[62] M. Sousa, D. Mendes, R. K. d. Anjos, D. S. Lopes, and J. Jorge. Negative space: Workspace awareness in 3d face-to-face remote collaboration. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–2, 2019. doi: 10.1145/3359997.3365744 3

[63] L. Wanger. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pp. 39–42, 1992. doi: 10.1145/147156.147161 9

[64] S. Whittaker. Theories and methods in mediated communication: Steve whittaker. In *Handbook of discourse processes*, pp. 246–289. Routledge, 2003. 2

[65] N. Wong and C. Gutwin. Where are you pointing? the accuracy of deictic pointing in cves. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1029–1038, 2010. doi: 10.1145/1753326.1753480 2

[66] F. Zhang, K. Katsuragawa, and E. Lank. Conductor: Intersection-based bimanual pointing in augmented and virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):103–117, 2022. doi: 10.1145/3567713 3