BodyAvatar: Creating Freeform 3D Avatars using First-Person Body Gestures

Yupeng Zhang[1,2], Teng Han[1,3], Zhimin Ren[1,4], Nobuyuki Umetani[1,5], Xin Tong[1], Yang Liu[1], Takaaki Shiratori[1], Xiang Cao[1,6]

[1]Microsoft Research Asia, [2]University of Science and Technology of China, [3]University of Bristol, [4]University of North Carolina, [5]The University of Tokyo, [6]Lenovo Research & Technology
yupengzhang@outlook.com; hanteng1021@gmail.com; zren@cs.unc.edu; n.umetani@gmail.com; {xtong, yangliu, takaakis}@microsoft.com; xiangcao@acm.org

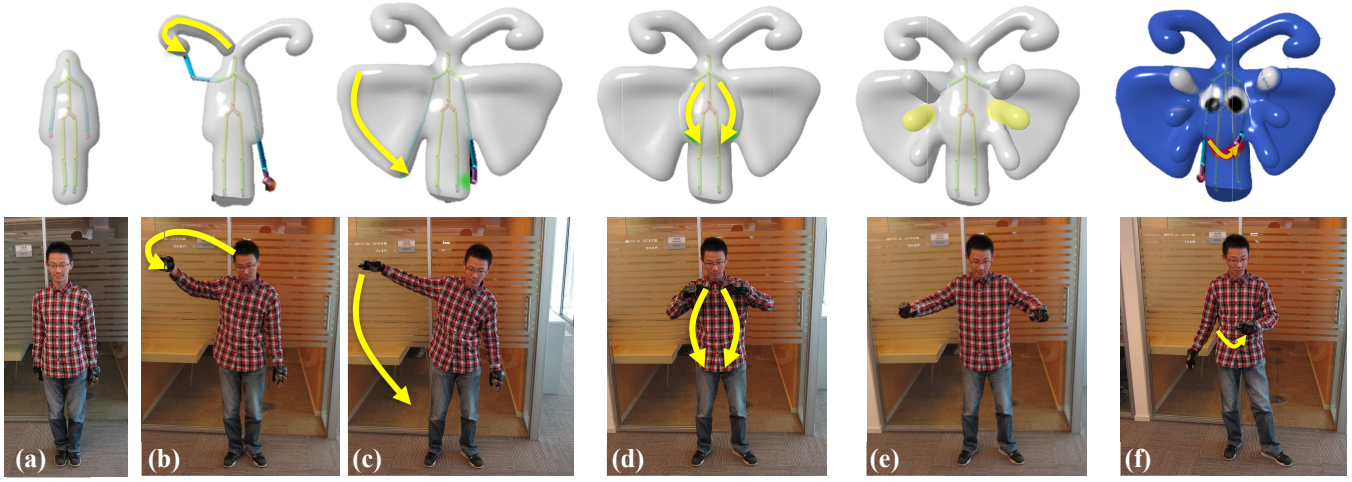

Figure 1. Creating a 3D butterfly using BodyAvatar.
(a) Scan the initial shape. (b) Drag antennas. (c) Sweep wings. (d) Sculpt a big belly. (e) Grow legs. (f) Paint color.

ABSTRACT

BodyAvatar is a Kinect-based interactive system that allows users without professional skills to create freeform 3D avatars using body gestures. Unlike existing gesture-based 3D modeling tools, BodyAvatar centers around a first-person "you're the avatar" metaphor, where the user treats their own body as a physical proxy of the virtual avatar. Based on an intuitive body-centric mapping, the user performs gestures to their own body as if wanting to modify it, which in turn results in corresponding modifications to the avatar. BodyAvatar provides an intuitive, immersive, and playful creation experience for the user. We present a formative study that leads to the design of BodyAvatar, the system's interactions and underlying algorithms, and results from initial user trials.

Author Keywords

3D avatar; body gesture; first-person; creativity.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

3D avatars, i.e. onscreen virtual characters representing players/users, are common in video games and online virtual worlds. Especially, with the popularity of Kinect™, these avatars can now directly mimic players' body movement, making the experience ever more immersive. These 3D avatars are usually designed by professional game artists, with only limited player customization available such as by selecting from a collection of predesigned body parts, colors, and accessories. However, if one likes to be more creative and build an imaginary 3D character of non-predefined forms as their avatar, they usually have to master complex 3D modeling software, much beyond the skills of typical game players.

Motivated to fill this gap and exemplified in a typical Kinect gaming setting, we present BodyAvatar, a system to allow any user to easily create their freeform 3D avatar out of imagination, using full-body gestures as the input language like they do when playing Kinect games. These avatars may also be animated by the user's body similar to in Kinect games. Given that avatars typically take forms of living creatures, BodyAvatar focuses on creation of organic-looking 3D shapes but without structural constraints.

BodyAvatar is unique from other gesture-based 3D modeling systems [e.g. 13, 15, 17], in that it centers around a first-person "you're the avatar" interaction metaphor. This metaphor is directly based on the fact that the user is creating a virtual representation of themselves. Instead of treating the 3D model as a separate passive object, the user considers their own body as a physical representation, or proxy, of the avatar being created. Based on an intuitive body-centric mapping, the user performs gestures to their own body as if wanting to modify it, which in turn results in corresponding modifications to the avatar.

UIST'13, October 8–11, 2013, St. Andrews, United Kingdom.

http://dx.doi.org/10.1145/2501988.2502015

The typical workflow of BodyAvatar starts with the user posing their body to set the initial shape of the avatar (e.g. a simple stick or a four-legged animal) (Figure 1a). The avatar can then be "attached" to the user's body and continuously animated by body movement. Under the attached status, the user performs various gestures to their own body to edit the avatar progressively, e.g., dragging from their head to give the avatar an antenna, or gesturing around their stomach to grow a fat belly for the avatar (Figure 1b, d). In addition, two users may create an avatar collaboratively in a similar fashion.

BodyAvatar enables free 3D avatar creation without requiring professional skills. It aims at an intuitive, immersive, and playful experience for the user - its first-person metaphor provides both an *intuitive* frame of reference for gesture operations and an *immersive* "you're the avatar" feeling, and its game-like interaction style offers a *playful* atmosphere.

RELATED WORK

Creating models of digital 3D objects is a common task in engineering, design, and film. It is however typically done using complex software based on conventional GUI, thus remains the privilege of trained professionals. As such, much research has been conducted to make this process accessible to novices, mostly falling into two categories:

Sketch-based: a common approach is to interactively interpret the user's 2D sketches into 3D shapes. For example, SKETCH [20] supports constructing 3D scenes by sketching geometric primitives. In a more "organic" style, Teddy [7] lets the user create rotund 3D objects by drawing their 2D silhouettes, as well as supporting operations like extrusion, cutting, and bending. In a style similar to Teddy, ShapeShop [16] supports creation of more complex and detailed 3D solid models through sketch. ILoveSketch [1] instead maintains the original sketch strokes and lets the user draw models consisted of 3D curves. Sketch-based tools leverages people's natural drawing abilities, hence is more intuitive for novice users than using mouse and keyboard to create shapes indirectly. However, expressing 3D shapes via 2D drawing still requires the mental skills of spatial projection and rotation, especially when the process involves frequent perspective changes. In addition, sketching normally requires pen input thus is mainly suitable for desktop settings.

Gesture-based: other systems aim to allow the user to construct and manipulate 3D shapes using hand gestures performed in 3D space, often based on a virtual sculpting metaphor. For example, Nishino et al. [13] use two-handed spatial and pictographic gestures, and Surface Drawing [15] uses repeated marking of the hand to construct 3D shapes. Such systems are more direct than sketch-based systems in that they do not require conversion between 2D and 3D. However, gesturing in the air without physical references or constraints pose challenges for the user to align their gestures to the virtual 3D object. To address this, McDonnell et al. [11] use a PHANTOM device to simulate haptic feedback from the virtual object, and Rossignac et al. [14] propose using a shape-changing physical surface to output the 3D shape being edited. Instead of generating actuated active feedback, Sheng et al. [17] use an elastic sponge as a passive physical proxy of the 3D model, which provides a frame of reference as well as allows gestures to be directly performed on it with passive kinesthetic and tactile feedback. Going a step further, KidCAD [5] projects a 2.5D model on malleable gel, which also serves as the input device for using tangible tools and gestures to edit the model. However, this setup cannot support full 3D. BodyAvatar is inspired by these works in that the user's body also serves as a physical proxy for the 3D avatar to receive gesture operations. In our case the proxy is neither actively actuated by the system [14] nor passively manipulated by the user [17] – it *is* the user.

Also related to gesture-based modeling is Data Miming [6], which allows using freehand gestures to describe an existing 3D shape in order to *retrieve* it from a database. Especially, Data Miming was designed based on an observation study of how people describe man-made objects and primitive shapes. BodyAvatar was designed following a similar user-centered process, with a complementary focus on *creating* organic avatars from imagination.

Another relevant research area is animating an arbitrarily shaped 3D model using human body movements. Traditionally this has aimed at mapping professional motion capture data to 3D meshes [2] or articulated characters [18]. Most recently, KinÊtre [4] allows novice users to use Kinect and their own body to animate a diversity of 3D mesh models, including those scanned from real-world objects. Note that animation has almost always been treated as a separate process from modeling, and often supported by different tools. In contrast, BodyAvatar unifies modeling and animating under the same first-person metaphor, and fuses them into one seamless activity.

FORMATIVE STUDY

BodyAvatar was designed through a user-centered process. Our exploration started from an abstract concept: to let a user create the shape of a 3D avatar using their full body in ways most intuitive to them. To concretize what those ways may be, we learned from our prospective users through a formative study.

Procedure

The goal of the study is to identify common patterns in how people intuitively express 3D shapes using their body. To do so, we adopted a Wizard-of-Oz [8] style method. The "system" was simulated by a researcher drawing 2D sketches of 3D shapes on a whiteboard in front of the participant, representing the avatar being created (Figure 2a). The participant used any body actions they felt appropriate to generate and modify the avatar until satisfied, while thinking aloud to explain the anticipated effect of each action. The researcher drew and modified the sketch according to the participant's actions and explanations.

Nine volunteers (2 female) participated. We included both young adults and children (aged 7 to 25) who are likely (but not the only) prospective user groups of BodyAvatar. Only one had experience with 3D modeling software. The participant was asked to "create" 5 avatars from a blank canvas using their body. For the first 4, the participant was asked to reproduce 4 example cartoon characters shown to them as 2D images (Figure 2b), and for the last one they created an avatar out of their own imagination. We refrained from giving any concrete instruction on how they should achieve the tasks, and only gave high-level hints when they were out of ideas, e.g. "you may pretend yourself to be the avatar", "try using your arm or leg to do something to change it". Each study session lasted 50-60 minutes. We observed and recorded both the workflow and the individual actions they took to create the avatars.

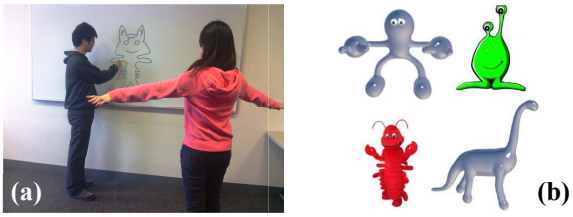

Figure 2. Formative study. (a) Setup. (b) Example characters.

Observations

The majority (seven) of the participants were dominated by a first-person mentality, i.e. they fantasized themselves as the avatar they were creating, and performed actions on or around their own body. One participant instead adopted a third-person style, i.e. imagining the avatar being in front of him, and performed actions in the air where he imagined the avatar to be. Another participant combined both styles.

Almost all participants treated the avatar as a single entity to be continuously built upon, hence they naturally followed a general "generating basic shape → adding features/details" workflow. The only exception was the one participant with 3D modeling experience, who combined this workflow together with cases where he first created subparts and then assembled them to the main entity.

For generating the basic shape, two strategies were observed: posing their full body to mimic the intended shape (8 participants, Figure 3a); and sketching the 2D silhouette in the air using their finger (2 participants). One participant used both strategies.

Several common actions were observed for adding features and details to the basic shape:

Growing a new limb was fairly common. Many participants expressed this by extending their arms or legs outwards, mimicking the shape of the expected limb (Figure 3b).

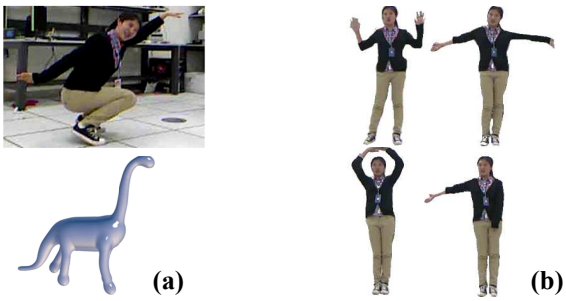

Figure 3. Participant actions. (a) Posing the body to mimic a dinosaur. (b) Extending arms to grow new limbs.

For adding thinner features, such as an antenna on the head, many participants used a *pinching* gesture, starting from the respective part of their own body and moving away, as if pulling the feature out. Unlike growing limbs, here participants expected the shape of the new feature to follow the path of their hand movement. This also allowed the participant to create more complex geometric features than they can directly mimic using arms or legs.

Tracing an imaginary shape to express it was frequently observed. Participants tended to use a finger to trace a curve, and use palms to trace a surface (either a free hanging surface, or the surface surrounding a volume traced using both hands, such as a big belly).

Another interesting action we observed was semantically *pointing* to one's own body features, e.g. pointing to their eyes with a finger to indicate a pair of eyes should be added to the avatar. Such actions were frequently used by participants to add human facial features such as eyes, mouth, nose, and ears.

Other than adding features, some participants also used their hand like a knife to cut unwanted parts. One participant used his hand like a brush on his body in order to paint color on the avatar.

Bimanual actions were commonplace, where participants simultaneously used both hands/arms to perform the same type of actions, mostly in a geometrically symmetric fashion. There were also occasional observations of asymmetric bimanual actions, e.g., two participants used one hand to describe the shape of a feature, and the other hand to point to their body where they would like the feature to be added.

Despite not seeing a real system, most participants already expressed much fondness in the concept of creating avatars using their body during brief interviews after the study. They thought the actions they came up with were quite intuitive. Not surprisingly, the young children among the participants turned out to be more creative (if less predictable) in their body expressions; while the adults demonstrated more stepwise planning in their workflow.

DESIGN

We now present the interaction design of BodyAvatar, which is directly based on principles and elements provided by the formative study. In the next section we will explain the underlying algorithms to enable these interactions.

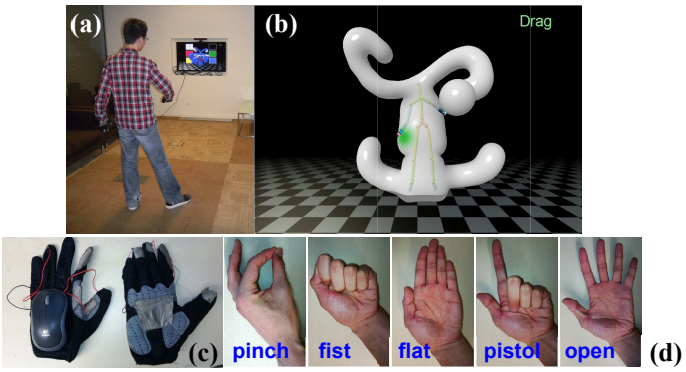

Figure 4. (a) BodyAvatar setup. (b) Software interface. (c) Gesture gloves. (d) Hand poses (shown without glove for clarity).

General Setup

BodyAvatar employs a typical Kinect gaming setup, where the user stands in front of the Kinect sensor and the screen displaying the software interface (Figure 4a). The user can walk freely inside the Kinect sensing range and make all kinds of body movements which are tracked by Kinect. Given that current Kinect SDK does not yet support view-independent recognition of multiple hand poses except open and close hands, we made a pair of very low-cost "gesture gloves" (Figure 4c) that can detect 5 different hand poses (Figure 4d) used for triggering different gesture operations. We took an off-the-shelf pair of gloves, and attached several flexible electrodes (made of conductive cloth) at strategic locations on each glove. These electrodes are connected to the circuit of a wireless mouse which is also attached to the glove. When the user makes certain hand poses, certain

combinations of electrodes contact each other, and in turn change the mouse button states that are wirelessly transferred to the computer. We envision the use of gloves will be replaced by Kinect sensing itself in the near future with technology advances [9]. In addition, we use automatic speech recognition to support a few abstract commands.

The software interface (Figure 4b) shows a virtual stage, on which a 3D body skeleton is displayed to represent the user tracked by Kinect, as if the user is looking into a mirror. Hence in some contexts we will use the phrases "user" and "body skeleton" interchangeably. On the same virtual stage also shows the 3D avatar being created, which may be either moving together with the body skeleton or static, depending on its status (to be detailed later). To help the user judge the spatial relationship between themselves and the avatar, parts of the body skeleton behind the avatar are occluded, and the parts inside the avatar are shown semi-transparently. Textual information tips are shown above the scene at times to indicate current statuses and operations.

Workflow

Consistent with what we observed in the formative study, the BodyAvatar workflow consists of two stages: generating the initial shape; and further editing it to the final result. For the first stage, we adopt the action used by most participants, i.e. posing their body to mimic the shape they want. We call this stage *Scan*.

Figure 5. Creating various initial shapes through scan.

Scan

BodyAvatar always starts in *Scan* stage, where the user sees a 3D blobby shape that surrounds their body skeleton and deforms like a viscous fluid as they move (Figure 5). The "fatness" of the shape, i.e. the radius from the skeleton to its surrounding surface, is adjustable by the user by saying "bigger" or "smaller". The user can keep deforming the shape using their body. When satisfied they say "scan" to freeze the shape as the initial avatar and enter *Edit* stage.

Edit

In *Edit* stage, the user performs a variety of gestures (to be detailed later) to edit the avatar. They may also return to *Scan* stage at any point by saying "scan", so that they can regenerate the initial shape.

Metaphors

BodyAvatar includes interactions under two metaphors: *first-person* and *third-person*. The first-person metaphor characterizes BodyAvatar from other systems, and is used predominantly; the third-person metaphor is included as a supplement to support necessary operations that are less convenient to perform as first-person. Besides *Scan* stage which is always first-person, in *Edit* stage the avatar may have two statuses to switch between: *attached* or *detached* to the user's body (or equivalently the user may think of their body attached or detached to the avatar), corresponding to the first-person and third-person metaphors respectively.

First-Person

Under the first-person metaphor, the onscreen avatar and the user's body are considered embodiments of each other. In addition to *Scan* stage where the onscreen shape directly mirrors the user's body pose, this metaphor is further reflected in the attached status in *Edit* stage. In this status, the avatar is continuously animated by the user's body motion in a style similar to KinÊtre [4], i.e. the body skeleton is embedded inside the avatar, and different body parts are attached to different surrounding sections of the avatar in a user-defined manner (we use the word "section" to refer to any subset of the avatar shape that can be animated individually, to differentiate from "body parts" that are parts of the user's body). For example, the avatar walks, jumps, bends, and turns around as the user does so, enforcing the perception that "you're the avatar". Note this does not assume geometric or structural similarity between the body and the avatar.

The user's two arms are treated specially, in that each arm (or both arms) can be individually detached from the avatar by a quick arm swing as if to "fling off" the attached avatar section, so that the arm is "freed up" to perform editing gestures to the rest of the body that remains attached. To attach the arm again, the user keeps it inside an avatar section for 3 seconds to "stick to it", then the arm becomes attached to the avatar section and animates it as usual.

In the attached status, the user's body serves as a physical proxy for the avatar to receive editing gestures. The user performs a hand gesture at a position on or around their body, and the editing effect is applied to the corresponding position on the avatar, e.g. dragging from their head to create a horn from the top section of the avatar (Figure 6a), based on a *body-centric mapping* in relation to the body part. This is true even if the body part is in motion itself, e.g. the user may use a detached hand to drag out a tentacle from the other (attached) hand when both hands are moving. Since the user's gestural movement is visualized by the body skeleton inside the avatar, it also creates an impression that the avatar is editing itself.

By doing so, the user's body provides an intuitive frame of reference that is both physical and semantic, making alignment of the editing gestures much easier than doing it in open 3D space. By looking at how the avatar moves following their body, the user can easily derive the mapping between the two, and use it to guide their actions. For avatar sections not directly attached to a body part, e.g. additional limbs, the user may use nearby attached body parts as a reference and extrapolate the mapping into the neighboring 3D space. For example, if the avatar has a horn that sticks forward from a limb attached to the user's arm, the user can put the other hand in front of that arm to reach the tip of the horn. Note that the user does not need to look at their body or hand to operate. Instead they can quickly put their hand near a body part through proprioception, and the visualization of their body skeleton and the avatar provides continuous feedback to further guide them to reach the exact target location. To further assist alignment, a green highlight (see

Figure 4b) is shown on the avatar's surface when the user's hand is near the surface, indicating the target location.

One potential challenge for first-person editing is when the avatar becomes too large in one or more dimensions, so that the user cannot reach its boundary. When this happens, the system automatically scales up the body skeleton's arm length so that the hand can always reach outside the avatar by a small distance when fully extended. The scaling factor is adjusted based on the elbow angle, so that when the user straightens the arm they can achieve maximal extended reach, but when bending the arm back to touch their body the action remains truthful to the physical arm pose.

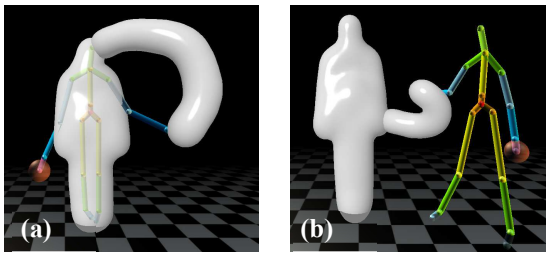

Figure 6. (a) First-person editing. (b) Third-person editing.

Third-Person

The detached status of the avatar symbolizes the third-person metaphor, where the avatar remains still as a separate passive object to be operated on, and the user acts as the operator that can freely move around or through the avatar. Provided as a supplement to first-person interactions, this status allows the user to do certain operations that are less convenient to perform in the first-person status, such as scaling and 3D rotation of the avatar model, and editing regions of the avatar that are harder to reach under the body-centric mapping due to human anatomy constraints, e.g. its backside.

For editing gestures in this detached status, an absolute *spatial mapping* is used, in that the location of the editing effect is the same as that of the user's hand, regardless of its spatial relationship to the body or the avatar (Figure 6b). Editing under this mapping is similar to most existing freehand gesture-based 3D modeling systems (e.g. [13, 15]) and may suffer similar challenges in aligning the gesture to the 3D avatar, thus is meant to be used sparingly.

When the avatar is detached, the user can also *manipulate* it as a whole. This is done in a style similar to the handle bar metaphor [19], where two hands moving in 3D space are used to rotate, scale, and translate a virtual 3D object as if it is skewered on a bimanual handle bar. To trigger manipulation, the user makes a *fist* pose in both hands and moves the hands in 3D, and ends the manipulation by opening the hands. To reduce mental complexity, manipulation is constrained to one dimension at a time (scaling; translation along x, y, or z axis; rotation around x, y, or z axis) which is determined by the type of movement the user makes in the beginning.

Switching between Metaphors

For switching between attached (first-person) and detached (third-person) statuses, we adopt an intuitive physical analogy. While attached, the user can "jump out" of the avatar by making a large hop to the side to *detach* from it; and while detached, the user can "walk into" the avatar and stay inside for 3 seconds to *attach* to it again. During the attaching step, the user can freely define the mapping between their body and the avatar similarly to KinÊtre [4], by posing their body and aligning certain body parts to certain avatar sections. In the beginning of *Edit* stage, the initial avatar generated by *Scan* also starts in a detached status, so that the user can freely define the initial mapping by attaching to it. As mentioned before, the user can also individually attach and detach their arms to various avatar sections using a similar physical analogy.

Detaching and attaching again (and optionally manipulating the avatar between the two steps) also provides the user an easy process to modify the body-to-avatar mapping on the fly for both animating and editing purposes. To make it easier to attach to the avatar in a semantically sensible manner, during manipulation the avatar is constrained to be always grounded on or above the virtual stage, and the user can snap its scale to be of the same height of their body.

Editing Gestures

Directly inspired by actions observed in the formative study, BodyAvatar offers the following gestural operations to edit the avatar (Figure 7). Each gesture is triggered by a specific hand pose, and ended by opening the hand(s). The editing effect is continuously previewed during the operation.

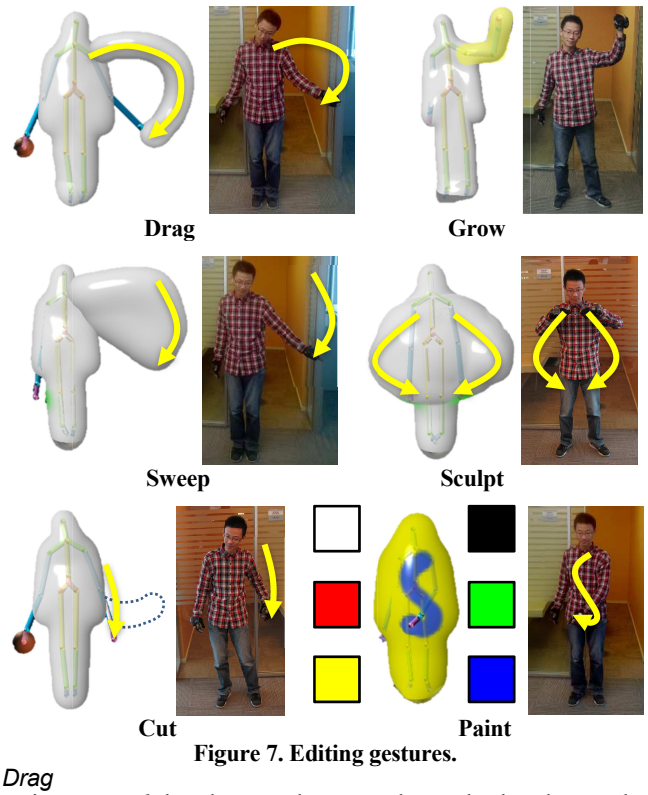

Figure 7. Editing gestures.

Drag

Using a *pinch* hand pose, the user places the hand near the surface of the avatar and then moves the hand in an arbitrary 3D trajectory, as if dragging out a thread. This operation adds a curved tube shape that follows the hand trajectory, e.g. an antenna. *Drag* can be performed in both first-person and third-person fashions. It unifies pinching and finger tracing actions observed in our formative study. Incidentally, *Drag* may also be appropriated as a more general "3D drawing tool" to create other 3D shapes by filling the volume with 3D "strokes" (i.e. tubes).

Grow

Using a *fist* hand pose, the user places their arm (or part of the arm) outside the avatar in any pose, as if pushing out from a shell. This operation adds a multi-segment cylindrical shape based on the pose of the arm segments outside, e.g. to make an additional limb. Unlike *Drag* which follows the hand trajectory, *Grow* only reflects the current arm pose, thus can be seen as a partial *Scan* only of the arm. *Grow* is derived from participant actions shown in Figure 3b, and is restricted as a first-person operation since it assumes the user to reside inside the avatar.

Sweep

Using a *flat* hand pose, the user sweeps their arm in space to traverse a 3D surface, which adds the surface (with a given thickness) to the avatar, e.g. to create a wing.

Sculpt

Using also the *flat* hand pose but with both hands simultaneously, the user traces around the outer side of an imaginary rotund volume rooted on their body, as if shaping a piece of clay. The system infers the volume and adds it to the avatar. For example, one can sculpt around their stomach to give the avatar a big belly. *Sculpt* acts as an incremental operation in that the volume is always added upon the existing avatar surface regardless of the starting points of the gesture, so that the user can sculpt repeatedly to create a volume beyond normal hand reach.

Both *Sweep* and *Sculpt* operations are inspired by how participants used their palms to trace surfaces (free hanging or surrounding a volume). Although conceptually such actions could be also performed in free space in a third-person manner, in practice this would cause considerable ambiguity in interpreting the boundary of the surface or volume. Instead, under first-person metaphor, the user's body naturally serves as a boundary to close the under-defined surface or volume. Therefore, we keep *Sweep* and *Sculpt* as first-person only operations.

Cut

Using a *pistol* hand pose, the user moves the hand across any region of the avatar, as if cutting through it. Any volume in the cutting path is removed, as well as sections that become disconnected from the avatar's main body afterwards. *Cut* can be performed in both first-person and third-person fashions.

Paint

Saying "Paint" opens a color palette shown on both sides of the screen, using which the user can add colors to the avatar. Placing their hand inside a color block picks the color, then the user either makes a *fist* hand pose to fill the entire avatar with the color at once, or uses the hand like a paintbrush to paint on the avatar with a *pistol* hand pose. Saying "Edit" closes the color palette and then the user can make other editing gestures as usual.

Except for *Sweep* and *Sculpt* that are differentiated by the number of hands used, the user may freely perform bimanual operations combining the same or different types of editing gestures. Similarly to *Scan*, the user can say "bigger/smaller" to adjust the size parameter used for certain operations, including diameter of the tube/cylinder in *Drag*/*Grow*, thickness of the surface in *Sweep*, and diameter of the paintbrush in *Paint*. When not performing gestures, a sphere is shown on each hand of the user, the size of which indicates this size parameter. Saying "Cancel" triggers undo of the most recent operation.

These editing gestures provide a complete set of operations that allow the user to create curves, surfaces, and volumes on the avatar, as well as trim and color it, resulting in a diverse range of avatars that can be created (Figure 8). Compared to existing gesture-based 3D modeling systems, some of our gestures are unique to the first-person metaphor, e.g. *Grow*; others take on a new meaning (e.g. *Sculpt*) under the first-person metaphor. There are also other interesting actions from the formative study that are less straightforward to support but we would like to explore in the future, such as semantic pointing and using one's legs to edit.

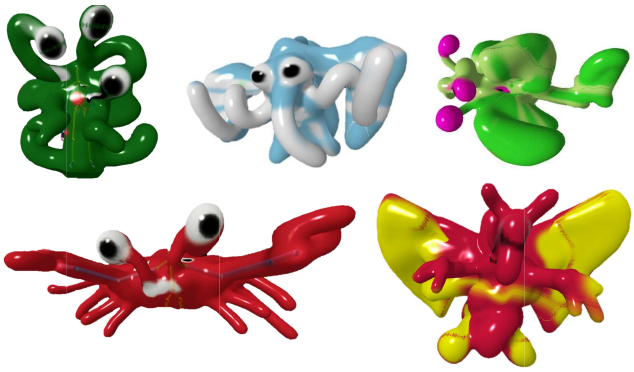

Figure 8. Avatars created using BodyAvatar.

Two-Person Creation

Based on the same functionalities described above, BodyAvatar also allows two people to collaboratively create an avatar (Figure 9). In *Scan* stage, two users can join their bodies to scan into a more complex initial shape than a single user can possibly pose. In *Edit* stage, the avatar is attached to one user who can animate it and perform first-person editing, while at the same time the second user may perform third-person editing to the avatar but using the first user's body as a proxy. This interaction style combines merits from both first-person and third-person metaphors – the first user can animate the avatar to position and orient it for the convenience of the second user, and the second user can make edits that are inconvenient for the first user to perform, yet still have the benefit of a physical frame of reference.

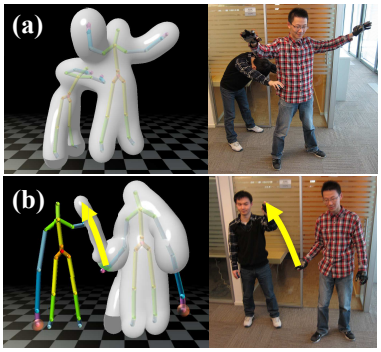

Figure 9. Two-person creation. (a) Scan. (b) Edit.

ALGORITHM

Avatar Model Representation

The avatar's static 3D shape is modeled by an implicit surface constructed from a number of meta-balls in 3D space

[12]; and its kinematic structure is represented by a tree-structured skeleton (referred to as "avatar skeleton" to differentiate from the "body skeleton" of the user). A triangle mesh is used to render the avatar. The mesh is generated from the meta-ball model using the marching-cube algorithm [3] and animated by the avatar skeleton. A texture map is used for coloring. Figure 10 illustrates.

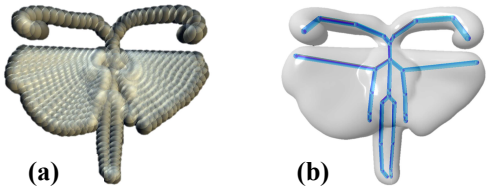

Figure 10. Avatar model representation. (a) Meta-ball model. (b) Rendered avatar mesh (without texture) and skeleton.

A meta-ball is an approximate spherical 3D Gaussian function, and the implicit surface is an iso-surface in the accumulated function of all meta-balls. We choose the meta-ball representation because of its flexibility in supporting ad hoc modification; it also naturally results in a smooth "organic" 3D shape that fits our context of creating avatars.

The avatar skeleton is made of a number of line segments ("bones") connected by joints. Each bone has a fixed length, and dynamically maintains its rotation angles relative to its parent joint.

At the time of its creation, each meta-ball is associated to one of the avatar bones. This allows us to also use meta-ball functions to calculate animation weights for the bones, to be detailed later.

The *Scan* stage generates an initial avatar model based on the user's body pose. The avatar skeleton is generated by replicating the user's body skeleton (both structure and pose), and meta-balls are placed along each bone with equal radius and distance (0.875 times the radius) (Figure 12a). This initial model is updated at every frame, until the user says "scan" to freeze the result. The avatar mesh is also regenerated from the meta-balls at every frame, which gives the avatar a fluid feeling during preview, and its topology may be changed on the fly by joining and separating certain body parts. "Bigger/smaller" speech commands adjust the meta-ball radius and in turn the fatness of the model.

Once the avatar model is frozen and passed to *Edit* stage, we define its pose at this time as the "default" (i.e. before animation) pose, which is used as the standard representation of the model. All further edits to the avatar are to be applied to this default model, after appropriate transform if applicable.

Animation

When the avatar is attached to the user in *Edit* stage, it is continuously animated by the user's body movement. Unlike in *Scan* stage where the meta-balls continuously change positions and regenerate the avatar mesh, here we directly animate the existing mesh because we need to maintain the general shape and structure of the avatar. To do so, we adopt the well-known skinning animation [10] technique, where each mesh vertex is moved by blending motion from one or several avatar bones. To calculate the blending weight of each bone for the vertex, we make use of function values (indicating "influence") of the meta-balls associated with the bones again. Calculated under the default pose, for each bone we sum the function values at the vertex position from all meta-balls associated with the bone, and use this accumulated influence as its weight (after normalization across all bones). By doing so, we obtain smooth animation results directly consistent with the mesh geometry, since it is also determined by the meta-balls.

Then to map the user's body motion to motion of the avatar bones, we need to establish an association between bones of the user's body skeleton ("body bones") to avatar bones. This association is determined by the user's *attach* action when the user poses their body inside the avatar. Each body bone searches for the closest avatar bone (considering both position and orientation) and attaches to it. If no avatar bone is sufficiently close (i.e. < 0.15 meter in distance and < 60° in separation angle), the body bone remains unattached. The root joints of both skeletons (body and avatar) are attached to each other by default. When the user moves, global translation and rotation of their body is applied to the avatar through the root joints, and rotation angles of each body bone is applied to their attached avatar bone if applicable. Unattached avatar bones preserve their original rotation angles.

Body-Centric Mapping

Key to BodyAvatar's first-person editing is a body-centric mapping to transform a position p_w in the world space (where the user and the animated avatar reside) to a position p_m in the model space defined by the default avatar pose (where the edit takes effect), based on current poses of the user and the avatar. This allows the user to always think of their editing gesture to be relative to their body regardless of its pose.

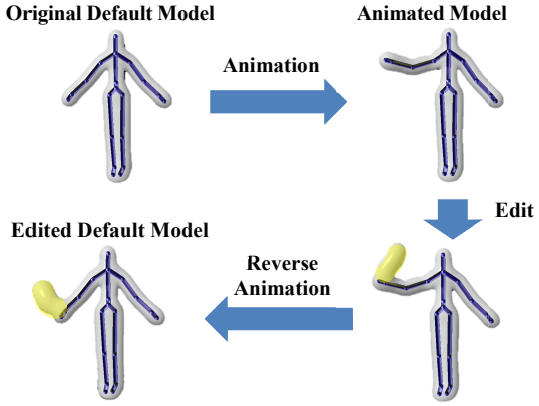

Figure 11. Body-centric mapping. (added shape is highlighted)

In general, this mapping is achieved by a "reverse animation" mechanism. We can imagine the new 3D shape (e.g. horn) being added by an editing gesture as rigidly bound to the nearby avatar section (and in turn the body part that animates the section). Then for each point p_w in the gesture trajectory, we calculate where it should move to (i.e. p_m) if that avatar section is to be animated back to its default pose (Figure 11). To enable this reverse animation, we follow a process similar to animating avatar mesh vertices. However, instead of blending from multiple bones, here we require all points in the gesture trajectory to be driven by one and the same avatar bone, so that the transform remains rigid and unambiguous to the user. To choose which bone to use, we take the "root point" of the added shape (for *Drag* and *Sculpt* this is the starting point of the gesture, and for *Grow* and *Sweep* this is the shoulder joint of the user's arm), and again calculate the accumulated influence of each avatar bone based on the function values of their associated meta-balls at this root

point. Note that to facilitate this calculation, in this step meta-balls are temporarily moved from their default positions to match the current avatar pose. We then select the bone with the largest influence to drive the reverse animation for all points in the gesture trajectory. Again, using meta-ball functions to choose the bone takes the geometry of the avatar into account implicitly. Note although the same bone is used for all points in the gesture trajectory, the actual reverse animation transform for each point may differ depending on the pose of the bone at the time each point is created, hence any motion of the avatar section (and in turn the body part that animates it) during the gesture is also taken into account as mentioned before.

Additional considerations are taken for *Drag* and *Sculpt*. *Drag* requires the created tube to always start on the surface of the avatar, hence needs to first project the starting point p_s of the gesture to the nearby surface. If p_s is outside the avatar (by a small distance, otherwise *Drag* is not allowed), this is simply done by finding the closet point on the surface; if p_s is inside the avatar, we first find the nearby avatar bone using the same method in the previous paragraph, then cast a ray from the bone through p_s to hit the surface and take the intersection point. For *Sculpt*, given its incremental nature, we not only project starting points of both hands onto avatar surface in a similar fashion, but also translate the rest of the gesture trajectory by the same displacement caused by the projection, so that it feels the gesture is always performed on top of the existing surface. For both *Drag* and *Sculpt*, after projection and/or displacement, points in the gesture trajectory are transformed using reverse animation as usual.

Editing Operations

All shape-adding operations (Figure 12b-e) are supported by adding meta-balls to the default model. New avatar bones may be added as needed to associate with the newly added meta-balls, and are connected to the closest joint in the avatar skeleton. Reverse-animated gesture trajectory is used in first-person editing, and original gesture trajectory is used in third-person editing. The avatar mesh is regenerated from the meta-ball model after each operation is completed.

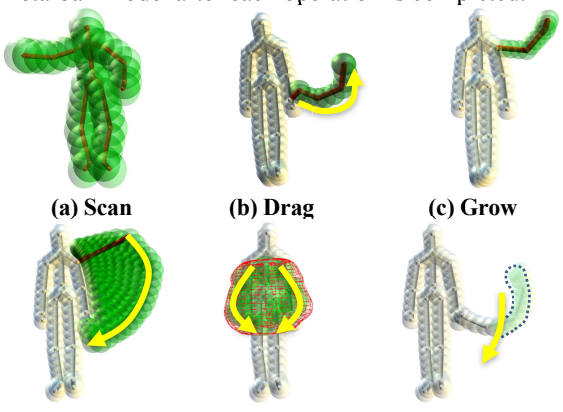

Figure 12. Operation effects on the avatar model. (affected meta-balls and bones are highlighted)

Drag adds a chain of meta-balls along its trajectory, along with a chain of 4 bones sampled from the trajectory.

Grow generates meta-balls along the arm in a fashion similar to *scan*. A replication of the body bones of the arm is added to the avatar skeleton.

Sweep tessellates the surface swept by the arm with a grid of meta-balls. This surface is defined by joining the line segments connecting the user's hand and shoulder at each frame during the gesture. The surface is treated as rigid, thus only one bone is added to animate it as a whole.

The "bigger/smaller" speech commands affect the radius of the meta-balls created by *Drag*, *Grow*, and *Sweep*.

Sculpt requires generating a rotund volume from two 3D hand trajectories only. To infer this volume, we take each pair of points along the two trajectories, and use them as the endpoints to define a 120° arc, whose radius points towards the user's body. Joining all these arcs results in a 3D curved surface that defines the outer side of the target volume (shown as a wireframe in Figure 12e). The inner side is defined by the existing surface of the avatar. Then a number of meta-balls are added to approximate this volume, using a greedy algorithm to fill the space with as few and as large meta-balls as possible.

Cut deletes all meta-balls the gesture trajectory intersects with. If all meta-balls associated with an avatar bone are deleted, the bone will also be deleted. If the model breaks into several disconnected components after the cut, the component containing the earliest created meta-ball or a bone closest to the root joint is kept as the main body, and the other components are deleted.

Unlike other operations, *Paint* affects the avatar's texture map but not the geometry or structure. Instead of using reverse animation, here we adopt a screen-based 2D painting metaphor since the color is always applied to the avatar surface. Taking the user's hand position, we find a point on the avatar surface that occupies the same 2D location on the screen, by casting a ray from the virtual rendering camera through the hand to hit the avatar surface. The texture color at this position is modified accordingly.

IMPLEMENTATION

The BodyAvatar software is implemented in C++ and runs on Windows OS in real-time. It uses Kinect for Windows SDK for body tracking, DirectX 11 for graphics rendering and GPU acceleration, and Microsoft Speech Platform for speech command recognition.

USER TRIAL

To understand the effectiveness of BodyAvatar, we conducted an initial user trial. Six volunteers (3 female), aged 22-23, participated in the trial (noted as P1-P6). Three had experience with 3D modeling software (P1 for 4 years, P2, P5 each for 1 year), two had drawing experience (P1 for 14 years, P5 for 3 years), and 5 had played Kinect games. Although the current participants were from a relatively uniform age group due to availability, they represented a major prospective user population for BodyAvatar, i.e. young adults. In the future we plan to conduct further trials with children and other age groups.

Procedure

Each participant participated in the trial individually. The BodyAvatar interface is displayed on a 32-inch LCD screen. The participant stood and walked between 2.8-4.5 meters in front of the screen determined by the sensing range of Kinect. Considering variable accents of the participants, to prevent them from being distracted by recognition errors of speech commands, a researcher entered correct commands through hotkeys according to their speech when necessary.

The researcher first gave a demonstration of BodyAvatar, then walked the participant through each operation by letting them try themselves. Some examples the researchers created using BodyAvatar were also shown as inspirations. The participant then spent 5 minutes to freely explore the system until comfortable.

The participant was asked to create 3 avatars: (1) to mimic any one of three example avatars shown to them: a butterfly, a lobster, and a horned monster; (2) to create an avatar based on any one of four abstract textual descriptions: dragon, fish, big-headed baby, and alien; (3) to freely create any avatar they specified themselves. Note that although we provided candidate tasks in (1) and (2), these were merely used to inspire the participant to gradually open up their creativity, and we were not concerned about how faithfully the avatar created matches the chosen subject.

For a conceptual comparison, after the BodyAvatar trial, the participant also tried ShapeShop [16], a fully functional sketch-based 3D modeling tool. The participant watched a demo video and was walked through the main functions of ShapeShop. They were then asked to explore the tool and do a free creation of any 3D model they like.

The participant was asked to think aloud throughout the entire trial. We observed their behaviors such as workflow, frequent actions, engagement, etc. After the session, we interviewed the participant about their experience. Each participant session lasted 75 to 100 minutes in total.

Findings

All the participants completed 3 avatars using BodyAvatar without difficulty. They spent on average 11 minutes 33 seconds creating each avatar, of which 6 minutes 48 seconds was for modeling the shape and 4 minutes 45 seconds for painting colors (we should note that an earlier, less robust version of the coloring algorithm was used in the user trial, which contributed to the overrepresented time spent on painting). Participants were satisfied with their creations for 17 of the total 18 avatars. Figure 13 shows examples.

Figure 13. Avatars created during user trial.

Intuitiveness, Immersion, Playfulness

BodyAvatar successfully achieved its goal to be intuitive, immersive, and playful:

Intuitive: participants all found the concepts and operations natural, simple, and straightforward to understand and perform. None had difficulties understanding the 3D structure of the avatar and the effects of the gestures. Five of the 6 participants said the system was very easy to learn. The only exception P1 said she needed some time to memorize the mapping between gestures and operations, especially when the same hand pose is reused for different operations under different contexts (e.g. *Grow* and *Fill Color*). This challenge was partially due to the limited set of hand poses we can detect using the gesture glove, and may be overcome with future technologies. Indeed, we did not observe the participants' learning speed of BodyAvatar to be dependent on their existing experience with 3D modeling, drawing, or Kinect gaming. In comparison, the 3 participants with 3D modeling experience (P1, P2, P5) succeeded in creating a simple rigid model using ShapeShop (e.g. a toy car), while the other 3 found it difficult to judge the 3D structure of the object and perform operations to their desired effects, and gave up as a result despite the fact that they all appreciated the higher precision offered by ShapeShop's interface.

Immersive: the first-person metaphor resulted in a high degree of immersion for the participants, as they felt themselves to be the avatar. As P1 put: "*I even forgot where I was. I imagined I was right in there (the screen), messing around in that closed space about myself.*"

Playful: the entire creation process was filled with laughter of all participants throughout – when they finished their avatar, when they achieved a good result from their gesture, and even when they got an unexpected result. The feeling that they were looking into a mirror and fiddling with themselves seemed to have made everything laughable. During the interview, every participant expressed that using BodyAvatar was a very enjoyable experience.

In fact, immersion and playfulness together resulted in a high level of engagement from the participants, side-evidenced by their perception of time. During the interview we asked the participant how long they thought they spent with BodyAvatar, and all quoted a much lower number than the actual time passed - "*I didn't feel that long had passed, so was amazed when I saw it was half past eight.*" (P3)

First-Person Metaphor

We were particularly interested in participants' experience of the first-person metaphor in BodyAvatar. Although we did not have access to other gesture-based 3D modeling tools for comparison, the third-person operations in BodyAvatar captured some of their aspects, thus can be considered as a delegation for informal conceptual comparison.

All the participants agreed that first-person operations were easy and helpful for the creation process ("*you can feel your avatar is really alive in the first person mode, and it's much easier to control it than in the third person mode*", P3). The foremost advantage of first-person operations was the intuitive physical frame of reference. When making edits in third-person style, all the participants encountered challenges in positioning their hand relative to the avatar especially along the depth dimension, and had to try the gesture several times before confirming – a typical issue in gesture-based 3D modeling tools. Such difficulties did not happen in the first-person status: participants simply looked at the screen and used their proprioception to reach the appropriate position. As a result, participants were more active and confident in first-person status, and showed no hesitation when performing gestures.

Although the avatar could be moved and rotated in both first-person (by walking and turning one's own body) and third-person (manipulating using gestures) style, participants used these two ways for very different purposes. First-person

rotation was used very frequently, almost subconsciously, by all participants to examine the 3D structure of the avatar, as if turning around in front of a mirror to check new clothes – an intuitive action directly from everyday life. In contrast, third-person rotation was almost only used to prepare for an editing operation, and participants always had to pause to think about the desired rotation before performing it – similar observations were made with ShapeShop, where participants seldom rotated the model for examining its 3D structure.

On the other hand, aside from inherent limitations of the first-person metaphor such as the difficulty to edit the backside of the avatar, technical limitations of Kinect body tracking capability also made some first-person operations vulnerable to tracking noise/error when certain body parts occluded or contacted each other (e.g., bending down to touch one's foot). Third-person manipulation and editing operations helped participants to circumvent these issues.

Two-Person Creation

In addition to the main user trial that focused on the single user experience, we conducted an informal trial session afterwards with a pair of users (who also participated in the main trial) on the two-person creation experience, which again yielded positive feedback. In particular, the two-person editing style (Figure 9b) made some editing steps considerably easier. Even when the avatar section being edited (e.g. rear legs of a centaur) was somewhat offset from the first user's body, the second user could still use the first user's body as a starting point to quickly extrapolate and locate the section in space, without trying repeatedly like in a pure third-person metaphor.

DISCUSSION & CONCLUSION

Current BodyAvatar interactions are most suited for creating organic-shaped avatars, not necessarily for models with polyhedral features. This is consistent with our objective to create living creatures representing the user. If a user wishes to create mechanical-looking objects instead, a different interaction style or system may be preferable, e.g. one that is based on assembling primitives. On the other hand, there is no algorithmic limitation in the range of shapes that can be modeled using our meta-ball representation together with potential technical extensions such as blob-trees as demonstrated by ShapeShop [16].

The level of detail that BodyAvatar can create could be limited by the physical granularity of the gestures (both in terms of the user's motor control capability and in terms of Kinect's sensing capability) under the first-person metaphor, where the user and the avatar are typically close to a one-to-one scale. However, under the third-person metaphor the user may freely scale up the avatar to do detailed edits on a portion of it, although this may again highlight the gesture alignment challenge as seen in other gestural 3D modeling systems. Considering all factors, we believe our choice to focus on the first-person metaphor is rational given our goal to support novice users' creation, where intuitiveness and simplicity takes priority over precision and sophistication.

There exist many possible extensions to BodyAvatar. For example, in the future we are interested in incorporating other elements into the avatar, such as real-world objects (e.g. a hat) scanned by Kinect, or predesigned components (e.g. facial features or polyhedral primitives) retrieved using gestures (similarly to [6]). These will further expand the range of avatars BodyAvatar can create, making it more generalizable to other application domains, such as designing 3D-printed articulated toys, interactive storytelling, or prototyping characters and animations for movie production.

In conclusion, BodyAvatar provides an intuitive, immersive, and playful experience for novice users to create freeform 3D avatars using their own body. Its first-person interaction metaphor is unique from existing gesture-based 3D modeling tools, and well accepted by our users.

ACKNOWLEDGEMENTS

We sincerely thank Jaron Lanier, Shahram Izadi, and Jiawen Chen for invaluable research discussion and help, and anonymous participants of the user trial.

REFERENCES

1. Bae, S-H, Balakrishnan, R., Singh, K. (2008). ILoveSketch: as-natural-as-possible sketching system for creating 3D curve models. *UIST*. p. 151-160.
2. Baran, I., Popovic, J. (2007). Automatic rigging and animation of 3D characters. *SIGGRAPH*.
3. Bloomenthal, J. (1988). Polygonization of implicit surfaces. *Computer Aided Geometric Design*, 5(4), p.341-355.
4. Chen, J., Izadi, S., Fitzgibbon, A. (2012). KinÊtre: animating the world with the human body. *UIST*, p. 435-444.
5. Follmer, S., Ishii, H. (2012). KidCAD: digitally remixing toys through tangible tools. *CHI*. p. 2401-2410.
6. Holz, C., Wilson, A.D. (2011). Data miming: inferring spatial object descriptions from human gesture. *CHI*, p. 811–820.
7. Igarashi, T., Matsuoka, S., Tanaka, H. (1999). Teddy: a sketching interface for 3D freeform design. *SIGGRAPH*, p.409-416.
8. Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *Transactions on Office Information Systems*, 2(1), p. 26–41.
9. Keskin, C., Kirac, F., Kara, Y.E., Akarun, L. (2012). Hand pose estimation and hand shape classification using multi-layered randomized decision forests. *ECCV*. p. 852-863.
10. Lewis, J. P., Cordner, M., Fong, N. (2000). Pose space deformations: a unified approach to shape interpolation and skeleton-driven deformation. *SIGGRAPH*. p. 165–172.
11. McDonnell, K., Qin, H., & Wlodarczyk, R. (2001). Virtual clay: A real-time sculpting system with haptic toolkits. *I3D*. p. 179-190.
12. Nishimura, H., Hirai, M., Kawai, T., Kawata, T., Shirakawa, I., Omura, K. (1985). Object modeling by distribution function and a method of image generation. *Electronics Communications Conference*, p. 718–725.
13. Nishino, H., Utsumiya, K., & Korida, K. (1998). 3D object modeling using spatial and pictographic gestures. *VRST*. p. 51-58.
14. Rossignac, J., et al. (2003). Finger sculpting with digital clay: 3D shape input and output through a computer-controlled real surface. *Shape Modeling International*. p. 229-234.
15. Schkolne, S., Pruett, M., & Schroeder, P. (2001). Surface drawing: creating organic 3D shapes with the hand and tangible tools. *CHI*. p. 261-268.
16. Schmidt, R., Wyvill, B., Sousa, M. C., Jorge, J.A. (2005). ShapeShop: sketch-based solid modeling with blobTrees. *Eurographics Workshop on Sketch-Based Interfaces and Modeling*. p. 53-62.
17. Sheng, J., Balakrishnan, R., Singh, K. (2006). An interface for 3D sculpting via physical proxy. *GRAPHITE*. p. 213-220.
18. Shin, H.J., Lee, J., Shin, S.Y., Gleicher, M. (2001). Computer puppetry: an importance-based approach. *Trans. on Graphics*. 20(2), p. 67-94.
19. Song, P., Goh, W.B., Hutama, W., Fu, C., Liu, X. (2012). A handle bar metaphor for virtual object manipulation with mid-air interaction. *CHI*. p. 1297-1306.
20. Zeleznik, R.C., Herndon, K., & Hughes, J. (1996). SKETCH: an interface for sketching 3D scenes. *SIGGRAPH*. p. 163-170.