Embodiment or Manipulation? Understanding Users' Strategies for Free-Hand Character Control

Benjamin Walther-Franks
TZI, Universität Bremen
CONTACT Software GmbH
bwf@uni-bremen.de

Tanja Döring
TZI, Universität Bremen
tanja.doering@uni-bremen.de

Meltem Yilmaz
TZI, Universität Bremen

Rainer Malaka
TZI, Universität Bremen

ABSTRACT

Controlling a virtual character with your free hands is a useful task for many 3D applications such as games, computer puppetry, or emerging virtual reality applications. So far, only specialist controls have been established in the animation industry. Yet little is known about novices mental models for character control, a key to designing widely usable natural and expressive interfaces. To this end we conducted a gesture elicitation study with twelve participants performing mid-air gestures for thirteen given character motions. The mental models observed fall into two distinct categories: 1) external manipulation of an imagined physical puppet and 2) the gesturing hands embodying the motion of the virtual body part being "controlled". The employed mental model determined hand posture and the mental transformation from gesture to character motion. We present and discuss a gesture set that can inform virtual puppetry interfaces for various application domains.

CCS CONCEPTS

• **Human-centered computing → Gestural input**.

KEYWORDS

Free-hand gestural input; animation; user-defined gestures.

ACM Reference Format:
Benjamin Walther-Franks, Tanja Döring, Meltem Yilmaz, and Rainer Malaka. 2019. Embodiment or Manipulation? Understanding Users' Strategies for Free-Hand Character Control. In *Mensch und Computer 2019 (MuC '19), September 8–11, 2019, Hamburg, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3340764.3344887

MuC '19, September 8–11, 2019, Hamburg, Germany

ACM ISBN 978-1-4503-7198-8/19/09…$15.00
https://doi.org/10.1145/3340764.3344887

1 INTRODUCTION

Character control is a complex task that demands for sophisticated and tailored interaction techniques. To date, these are prevailingly based on traditional desktop input devices or motion capture setups designed for expert users. Yet there are a variety of emerging fields where easy-to-master character control techniques could empower non-expert users to customize avatar movements and expressions, such as social media, computer-aided physiotherapy, games, or layman animation. Motivated by the increasing quality and availability of mid-air gesture technology, we explored gestural input for character control: hands can easily enact fine-grained and complex movements, make use of 3D space, and incorporate timing aspects. While there are many examples of animating fictitious characters in the real world, like puppetry or marionettes, it is unclear how this knowledge can be transferred to free-hand gestural interaction with a virtual character presented on a screen. In this work, we conducted a gesture elicitation study [15] for free-hand gestural interaction for character control revealing the approaches of uninitiated users regarding the gestures' nature, hand form and mental rotations. Based on these results we developed a gesture set for free-hand character control to animate 13 discrete character motions, which animate face, head, torso and arms of a human character.

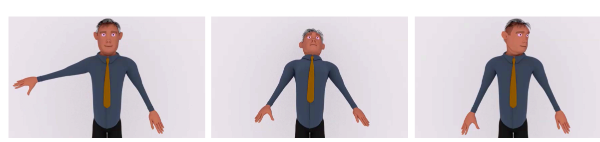

Figure 1: Stills of the videos for the referents "abduct right arm", "lean back" and "turn head".

Our results show that two main control approaches emerge: an embodiment approach where users align their hands with the moving parts of the puppet's body and control this from within, and a manipulation approach where users perform external operations on an imagined puppet that they are facing. We further identified dominating strategies regarding the mapping of fingers, hand and arm to dedicated body parts of the animated character.

2 RELATED WORK

Performance animation, the realtime control of virtual actors with the whole body using motion capture technology, is widely established in the movie and games industries [11] and has been explored in human-computer interaction research [10]. Character control interfaces based on only partial body tracking have also been explored. Such interfaces have the advantages of requiring less tracking space and cheaper and less invasive hardware. Also, they free up body parts for other tasks, such as viewing a simulated space through virtual reality goggles. Examples are finger walking techniques that embody part of the character's motion with the hands [8, 9]. The Jim Henson company has been using electromechanical hand input rigs to puppeteer real and digital character heads [6]. The recent availability of consumer virtual reality equipment has brought forward first ideas for immersive character animation using free-space hand input [4]. Due to these developments and a growing application domain, there is a demand to better understand how to design mid-air hand gesture interaction for character animation and control.

Research on mid-air gestural interaction has been conducted for decades [1, 2, 7]. While free-hand gestural interaction has been explored with gesture elicitation studies in different domains (e.g. for music [5] or TV control [13]), designing a free-hand gesture set for character control based on conducting a gesture elicitation study has, to our knowledge, not been addressed so far.

3 GESTURE ELICITATION STUDY

In order to better understand users' strategies for free-hand character control and to design an appropriate gesture set, we conducted a gesture elicitation study following the methodology applied in earlier studies such as presented by Wobbrock, Vatavu and their colleagues [12–15].

Study Design and Metrics

Analogous to prior work, we presented the effects of commands (i.e. the referents) to participants and asked them to perform corresponding gestures (i.e. which could cause the effects). These referents were presented as short videos. As character control can be highly complex, we had to select a manageable set of commands. We decided to animate face, head, arms and torso of a standing human character waist-up and determined 13 referents that represent basic character motions (see Figure 2; for stills of example referents see Figure 1). 9 of these present changes in single features (e.g. movement of head or one arm), while 4 present two symmetric features (e.g., movement of both arms). While these commands clearly do not cover the full range of possible animations, they present a reasonable set to start with. Furthermore, the commands match the range for muppet-type hand puppets that rarely present or move their legs, so they could as well fit computer puppetry, which could be a use-case for free-hand input in the future.

During the study, we collected demographics data with questionnaires as well as users' judgements on their performed gestures regarding appropriateness, ease and their preference regarding one or both-handed gestures. Furthermore, we videotaped all performed gestures and annotated the video material with an annotation tool regarding the nature of the gestures, mental rotations and hand morphology.

Participants

12 participants (8 female) with an average age of 25.8 years (SD = 7.8) took part in our study. All of them frequently use the computer. While 4 of the users had no prior experience with any form of free-hand gestural interaction, 8 of them had used devices such as Wii, Kinect, or LeapMotion before, mainly in games applications. None of the participants was an expert in character animation, although 4 users had some prior experience. One person was left-handed.

Apparatus and Setup

The study was conducted in a lab, in which participants were seated at a desk with a 21" monitor presenting full-screen referents accompanied by a textual description of the motion and questionnaires in turn. An interaction area in form of a 50 by 50 by 50 cm cube volume in front of the screen was highlighted through rectangular markings of high-visibility tape on the table and monitor. The gestures were captured with two video cameras filming from the top and the side.

Procedure

After an introduction users were presented the 13 referents in the form of short videos in a random order. Participants could demand to view videos again. Afterwards, they were asked to give first a one-handed and then a two-handed gesture while thinking aloud. Directly after performing a gesture, the participants were asked to rate the gesture for their subjective assessment of how appropriate it was and how easy they found it to perform on a 5 point Likert scale. After both gestures for a referent, they also gave their preference for handedness. One session lasted approximately 40 minutes.

4 RESULTS

Overall, we collected 312 gestures: 12 individual user-defined one-handed and 12 two-handed suggestions for each of the 13 referents. In the following, we present the gesture classification we developed on the basis of the video analysis and explain the results about dominating gesture strategies regarding the classification categories. Furthermore, we present agreement scores for the referents, a user-defined gesture set and implications for gesture design for character control.

Classification of Character Control Gestures

The suggested gestures were manually segmented and annotated regarding their *nature*, the hand and finger *morphology*, and the applied *mental rotation*.

Nature. Four categories emerged for a gesture's nature: *embodiment*, *manipulation*, *deictic*, and *abstract* gestures, representing different mental modals of control. *Embodiment* gestures are enacted in the tradition of physical hand puppets, where hand and fingers directly represent and animate body parts of the puppet. In the second category, *manipulation*, the mental model applied goes back to physical plush toys or marionettes, where the animator touches the puppet and performs movements. Free-hand gestures that fall into this category are initiated by touching a virtual target, e.g. imagining holding and moving the virtual character's arm. The third category depicts *deictic* gestures that apply pointing gestures to the referents, e.g. using the thumb or fingers to indicate a movement of a body part. Last, *abstract* gestures follow an arbitrary mapping.

Morphology. Inspired by the notation of form features proposed by Bressem [3], we applied 5 form clusters for hand configuration *fist*, *flat hand*, *single finger*, *combination of fingers (grasping)*, and *combination of fingers (not grasping)*.

Mental rotation. This aspect describes the mental rotation necessary to align user input motion to virtual character motion. We clustered it into the 4 categories *opposed*, *mirrored*, *aligned*, and *other*. Gestures in the *opposed* category take the scene "as is", i.e. the character facing the user in the way of the setup. Gestures with a *mirrored* transform assume a character facing the user, but with a mirrored alignment, i.e. the user's right relates to the character left. In *aligned* gestures, the user mentally aligns herself with the character, like a puppeteer standing behind a puppet. The user's left side is aligned to the characters left. Other mental transformations (e.g., performing gestures towards a horizontal tabletop surface) were grouped in the *other* category.

Gesture Strategies Regarding the Classification

Two researchers independently coded the videos regarding the classification criteria, also taking verbal expressions of participants into account. Results show that in the *nature* dimension gestures fall with an overwhelming majority into the two categories *embodiment* (57,7%) and *manipulation* (41,3%). Only three gestures in total were in the *abstract* and *deictic* categories (all performed by a single participant). Figure 2 shows the distribution of nature categories per referent. While overall a tendency towards embodiment for most gestures can be observed, we found clear preferences for manipulation gestures for the referents "roll eyes" and "smile". This is very likely due to the fine nature of these features – eyes and corners of the mouth – that suggest to be mentally touched and moved with fingers. An example for a referent with a strong tendency towards an embodiment gesture is "speak", which was often performed with the flat index to little fingers describing a hinge joint and opening and closing to move the jaw.

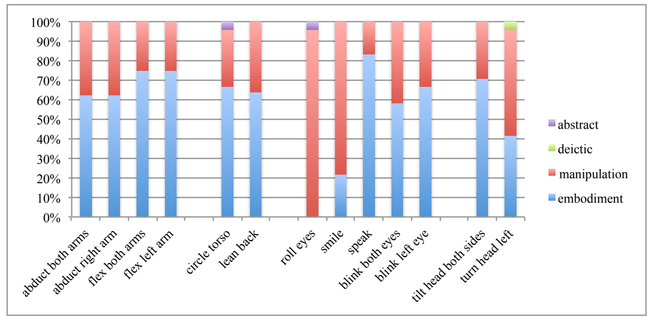

Figure 2: Classification by nature category for all gestures.

Looking at individual preferences regarding gestural nature choice per participant, we found that there seem to be "embodiment" and "manipulation" types of users. 6 participants performed mainly embodied gestures (10 or more of the 13 referents for each input) and 3 mainly chose manipulation gestures (either 12 or all of the 13 referents for each input). Only 3 participants applied a mixed approach. Analyzing the gestures' morphology in relation to their nature, we found that embodiment gestures used only the *flat hand* (71,1%) and *combinations of fingers (not grasping)* (28,9%), while manipulation gestures were spread over all categories but had a focus on *combinations of fingers (grasping)* (44,2%).

Analyzing the mapping of users' fingers, hands, and arms to figures' body parts, we found that prevailing strategies were to use fingers for face animation, hands for head animation, arms for arms animation and arms and hands together for torso animation. Regarding mental rotation, *opposed* was the dominating choice for manipulation gestures (which can be expected as this is the most natural) with 98,4%, whereas among embodiment gestures 55% used opposed, 26% aligned, 16% mirrored and 3% other rotations.

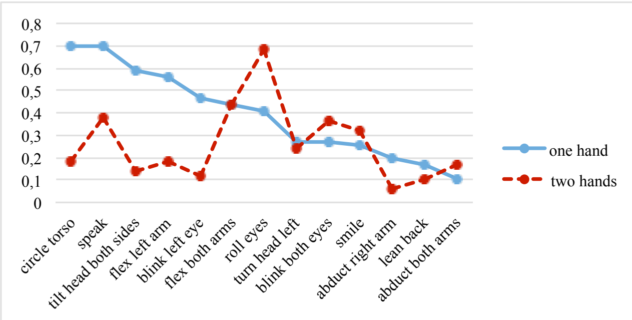

Figure 3: Agreement scores for the 13 referents.

Agreement and User Preferences

We grouped the gestures into clusters of identical gestures for each referent and calculated the agreement scores that reflect the consensus among participants [12]. The overall agreement for all one- and two-handed gestures was A_{1H}=0.38 and A_{2H}=0.26, respectively. Individual rates are given in Figure 3. For 6 of the referents one-handed gestures were preferred. For 5 (the two-sided referents and "lean backward") two-handed gestures were preferred. For the two head referents the votes were equal.

User-defined and Refined Gesture Set

In order to derive a user-defined gesture set we first chose the most frequently performed gesture for each referent. Secondly, this gesture set was refined regarding conflicts and mismatches. For example, the "abduct/adduct right arm" and "flex/extend left arm" referents were assigned the same gesture and "blink left eye" and "speak" were performed in the same way. There was one symmetry mismatch ("abduct right arm" and "abduct/adduct both arms" had different types of gestures) and one nature mismatch in that gestures for the same body parts should have the same nature.

Conflicts and mismatches were manually resolved as follows. The gesture for "turn head left" was replaced with the next-most popular one-handed embodied gesture to align it with the one-handed embodied gesture for "tilt head both sides". The gesture for "abduct/adduct right arm" was replaced with the second-most popular one-handed embodied gesture to avoid conflict with the gesture for "flex right arm". The gesture for "abduct/adduct both arms" was replaced with the two-handed version of the gesture for "abduct/adduct right arm". The gesture for "blink left eye" was replaced with second-most popular one-handed embodied gesture to avoid conflict with the gesture for "speak". Accordingly, the gesture for "blink both eyes" was replaced with the two-handed version of the gesture for "blink left eye".

The resulting final gesture set (see Figure 4) is free of conflicts and mismatches. In accordance to our classification results, it consistently maps user to puppet space – the puppet's upper body is controlled by whole arm gestures, its head is controlled by whole-hand gestures and the face by finger gestures. Furthermore, it uses the two-handed version of a one-hand control to control the symmetrical features arms and eyes. Last, the set applies mainly the same nature categories for each body region, with the exception of the face, where the embodied gestures for "speaking" and "blinking eyes" were quite strong, while for "roll eyes" and "smile" manipulation gestures were selected.

5 DISCUSSION AND CONCLUSION

In this paper we addressed non-expert users' strategies for free-hand character control and conducted a gesture elicitation study. Among our results are classification criteria for free-hand character control gestures, an analysis of the users' strategies based on these criteria, and a user-defined gesture set for character control. Our study revealed that users applied two dominating strategies with different underlying mental models: embodiment and manipulation. While embodiment gestures occurred more often in general, certain referents that control small elements were overall preferred to be performed by manipulation gestures. Furthermore, we found that participants preferred to use their fingers to control the character's face, their hands to control the character's head and their arms to control the character's upper body. Some implications for design derive from our work: first, manipulation and embodiment both seem to be suitable for this domain. Second, there seem to be users with a preference for embodiment and others with a preference for manipulation, which they consistently apply throughout almost all gestures; further research needs to be conducted whether this is due to specific experiences and could be applied to dedicated user groups.

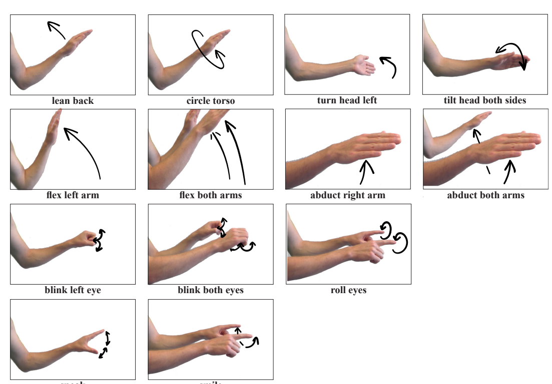

Figure 4: The final gesture set for free-space hand/arm-control of common character motions.

REFERENCES

[1] Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: Remote Control of Objects Using Free-hand Gestures. *Commun. ACM* 36, 7 (July 1993), 28–35. https://doi.org/10.1145/159544.159562

[2] Richard A. Bolt. 1980. Put-that-there: Voice and Gesture at the Graphics Interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. https://doi.org/10.1145/965105.807503

[3] Jana Bressem. 2013. *A linguistic perspective on the notation of form features in gestures*. De Gruyter: Mouton, Berlin, Boston, 1079–1098. https://doi.org/10.1515/9783110261318.1079

[4] Liat Clark. 2015. Animating in virtual reality with Oculus Rift. (Friday 8 May 2015). http://www.wired.co.uk/article/animating-in-vr-with-oculus

[5] Niels Henze, Andreas Löcken, Susanne Boll, Tobias Hesselmann, and Martin Pielot. 2010. Free-hand Gestures for Music Playback: Deriving Gestures with a User-centred Process. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (MUM '10)*. ACM, New York, NY, USA, Article 16, 10 pages. https://doi.org/10.1145/1899475.1899491

[6] John Jurgensen. 2008. From Muppets to Digital Puppets. http://www.youtube.com/watch?v=GN8WbHomQJg (last accessed July 29, 2019). http://www.youtube.com/watch?v=GN8WbHomQJg

[7] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. 1985. VIDEOPLACE - an Artificial Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '85)*. ACM, New York, NY, USA, 35–40. https://doi.org/10.1145/317456.317463

[8] Wai C. Lam, Feng Zou, and Taku Komura. 2004. Motion editing with data glove. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE '04)*. ACM, New York, NY, USA, 337–342. https://doi.org/10.1145/1067343.1067393

[9] Noah Lockwood and Karan Singh. 2012. Finger Walking: Motion Editing with Contact-Based Hand Performance. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '12)*. Eurographics Association, Aire-la-Ville, Switzerland, 43–52. https://doi.org/citation.cfm?id=2421739

[10] Ali Mazalek, Sanjay Chandrasekharan, Michael Nitsche, Tim Welsh, Paul Clifton, Andrew Quitmeyer, Firaz Peer, Friedrich Kirschner, and Dilip Athreya. 2011. I'M in the Game: Embodied Puppet Interface Improves Avatar Control. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '11)*. ACM, New York, NY, USA, 129–136. https://doi.org/10.1145/1935701.1935727

[11] Alberto Menache. 2011. *Understanding Motion Capture for Computer Animation* (second ed.). Morgan Kaufmann.

[12] Radu D. Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[13] Radu D. Vatavu and Ionut A. Zaiti. 2014. Leap Gestures for TV: Insights from an Elicitation Study. In *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video (TVX '14)*. ACM, New York, NY, USA, 131–138. https://doi.org/10.1145/2602299.2602316

[14] Jacob O. Wobbrock, Htet H. Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[15] Jacob O. Wobbrock, Meredith R. Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866