The Role of Avatar Fidelity and Sex on Self-Motion Recognition

Anne Thaler
Max Planck Institute for Biological Cybernetics
Tübingen, Baden-Württemberg
Germany

Anna C. Wellerdiek
Max Planck Institute for Biological Cybernetics
Tübingen, Baden-Württemberg
Germany

Markus Leyrer
Max Planck Institute for Biological Cybernetics
Tübingen, Baden-Württemberg
Germany

Ekaterina Volkova-Volkmar
Max Planck Institute for Biological Cybernetics
Tübingen, Baden-Württemberg
Germany

Nikolaus F. Troje
Queen's University
Kingston, Ontario, Canada

Betty J. Mohler
Technical University of Darmstadt
Darmstadt, Hessen, Germany

ABSTRACT

Avatars are important for games and immersive social media applications. Although avatars are still not complete digital copies of the user, they often aim to represent a user in terms of appearance (color and shape) and motion. Previous studies have shown that humans can recognize their own motions in point-light displays. Here, we investigated whether recognition of self-motion is dependent on the avatar's fidelity and the congruency of the avatar's sex with that of the participants. Participants performed different actions that were captured and subsequently remapped onto three different body representations: a point-light figure, a male, and a female virtual avatar. In the experiment, participants viewed the motions displayed on the three body representations and responded to whether the motion was their own. Our results show that there was no influence of body representation on self-motion recognition performance, participants were equally sensitive to recognize their own motion on the point-light figure and the virtual characters. In line with previous research, recognition performance was dependent on the action. Sensitivity was highest for uncommon actions, such as dancing and playing ping-pong, and was around chance level for running, suggesting that the degree of individuality of performing certain actions affects self-motion recognition performance. Our results show that people were able to recognize their own motions even when individual body shape cues were completely eliminated and when the avatar's sex differed from own. This suggests that people might rely more on kinematic information rather than shape and sex cues for recognizing own motion. This finding has important implications for avatar design in game and immersive social media applications.

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing**; • **Computing methodologies → Motion capture**; • **Applied computing → Psychology**; • **General and reference → Experimentation**; • **Social and professional topics →** *Gender*;

KEYWORDS

Virtual Characters, Shape Cues, Kinematic Information, Self-Motion Recognition, Biological Motion

ACM Reference Format:
Anne Thaler, Anna C. Wellerdiek, Markus Leyrer, Ekaterina Volkova-Volkmar, Nikolaus F. Troje, and Betty J. Mohler. 2018. The Role of Avatar Fidelity and Sex on Self-Motion Recognition. In *SAP '18: ACM Symposium on Applied Perception 2018, August 10–11, 2018, Vancouver, BC, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3225153.3225176

SAP '18, August 10–11, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5894-1/18/08.
https://doi.org/10.1145/3225153.3225176

1 INTRODUCTION

1.1 Avatars in Applications

Avatars are central to immersive social media platforms[1] which allow users to create a digital self that represents them to their friends and family in networked virtual reality applications. Additionally, for many video games a main virtual character is designed as part of the narrative that can often be experienced from the first-person perspective. Video game avatars are typically stylized characters that allow for no modification or adjustments to the user's identity.

In most applications to date, a highly stylized self-avatar is not yet completely able to convey the identity of the user. Many applications allow the configuration of the avatar's sex and visual appearance. However, without the use of a 3D body scanner the exact visual appearance of each user cannot be fully conveyed. Further, in many video games, the identity of a character in the story is forced to be the avatar of all users. The only information that is portrayed through these avatars, by definition, is the actions of the users. In immersive virtual reality we are beginning to see the use of full-body scanners for representing the user [Achenbach et al. 2017; Blom et al. 2014; Fleming et al. 2016; Piryankova et al. 2014]. In addition, the possibility for full body motion of the user conveyed in the avatar is beginning to be possible at a high fidelity [Beck et al. 2013; Lugrin et al. 2016].

Avatars graphically consist of visual appearance cues (color, clothing, and shape) and animation cues (angular movements and

[1] i.e., Microsoft AltspaceVR [Microsoft 2018], Oculus Rooms and Parties [Oculus 2018], Facebook Spaces [Facebook 2018], Linden Labs Second Life [Labs 2018b], and Linden Labs Sansar [Labs 2018a].

deformations of the body shape due to motion). This terminology can also be found in fundamental research on the perception of biological motion where human movement information is broken down into spatial configuration and kinematic cues.

Computer animation and recent advances in modelling realistic human bodies based on statistical data use a very clear and reproducible definition of body shape [Loper et al. 2015]. Body shape and pose are both spatial factors that are modelled statistically based on a population of scanned bodies either generalized across or towards a specific sex. Often, a further distinction is made to body texture that contains color information such as skin, hair, clothing etc. Color information has been shown to be an important aspect of identity when presenting static bodies [Mölbert et al. 2017; Thaler et al. 2018]. Both body shape and texture hereby contain e.g. sex-specific features. In computer animation where fully skinned characters represent the articulation of the body more explicitly and violations of constant segment length are more visible, the veridicality of Cartesian joint locations is often renounced and only joint angles are mapped from one body to another.

1.2 Person Recognition and Biological Motion Perception Research

Motion patterns of living beings are among the most biologically salient events. Research has shown that even when human motion is reduced to the movement of point-lights attached to the joints of an actor's body [Johansson 1973], humans are very efficient in picking up many social cues. For example, observers were found to be able to identify different actions [Dittrich 1993], the actor's sex [Barclay et al. 1978; Hill and Johnston 2001; Mather and Murdoch 1994; Troje 2002], and the actor's identity [Cutting et al. 1977; Hill and Pollick 2000].

Research on identity recognition from motion cues has demonstrated that observers can recognize themselves as well as friends above chance level in point-light walkers [Cutting et al. 1977]. Despite the little visual experience viewing own body motion from a third-person perspective, with the exception of viewing oneself in a video recording or moving in front of reflective surfaces, observers showed a higher sensitivity to recognizing themselves as compared to a friend [Loula et al. 2005; **?**]. Further, whereas recognition performance of friends and family members was found to be better in frontal and half-profile view, own motion recognition seems to be viewpoint independent [Jokisch et al. 2006; Prasad and Shiffrar 2009], suggesting an influence of the motor system that is independent of visual experience. Specifically, the high sensitivity to recognizing own motion might be the result of a perception-action link, implying that the motor system contributes to action perception and thereby enables recognition of self-motion also when viewed from different perspectives [Blake and Shiffrar 2007; Blakemore and Frith 2005; Schütz-Bosbach and Prinz 2007; Wilson and Knoblich 2005]. Whereas most studies on identity recognition from motion have used point-light displays in an attempt to reduce the available information to motion cues, little research has investigated self-motion recognition as a function of the visual appearance of the body [Wellerdiek et al. 2013].

Traditionally, point-light displays of human motion contain various visual cues such as the body configuration (or shape), that is the spatial position of the joints with respect to each other, and kinematic information, the dynamic changes in limb position and orientation over time. The relative contribution of body configuration and kinematic information for self-motion recognition has not been studied yet. This is likely due to the challenge in isolating shape and motion information as they are not independent from each other. The articulated structure of a human body where distances between joints remain approximately constant imposes constraints that generally cannot be met when trying to map the motion of one person onto the body of another. When working with biological motion point-light displays, researchers often relax the articulation constraint: The local motion of individual joints can be left unaltered, but the location of the trajectory of that joint is moved to match the mean location of the target body. In person recognition and biological motion research, spatial body configuration has often been ignored as a visual cue for identity recognition even though it contains rich information about bodily dimensions and might therefore also play a role in identification.

1.3 Recognition Research Using Virtual Characters

When instead of point-light displays, a virtual human character is used, the character also conveys sex in terms of the appearance and body shape details (e.g., secondary sex characteristics) which go beyond relative joint locations. Studies have shown that both body shape and motions of a virtual character influence the perceived sex of the character [McDonnell et al. 2009; Zibrek et al. 2015]. McDonnell et al. [2009] mapped female, male, and neutral motion on different characters (female, male, and ambiguous) and asked participants to categorize the character on a 5-point scale from 'very male' to 'very female'. The results showed that both the appearance and the motion of a character had an influence on the perceived sex. Another study demonstrated that the visual appearance of a virtual character can influence the perceived acceptance and trustworthiness towards this character [Vinayagamoorthy et al. 2006].

Some evidence for the importance of kinematic information for self-motion recognition comes from a study by Cook et al. [2012] where they investigated facial self-recognition when participants' motions were mapped onto a common facial form, thereby eliminating all individual shape cues. They showed that kinematic information mediated the recognition of self-produced motion.

In the present study, we investigated recognition of own whole body motion by eliminating individual visual appearance cues (color and shape) by mapping participants' previously recorded motions onto three different representations of a body (a point-light figure, a male, and a female virtual character). In order to map the motions of different actors onto the same body, the motions were reduced to joint angles between the segments of the body. We hypothesized that if kinematic information is sufficient for self-motion recognition, there should be no influence of the body representation (avatar fidelity) on recognizing own motion. If however identification with the body (fidelity, and sex) contributes to self-motion recognition performance, female observers might be better in recognizing their motions on a female character and male observers on a male character.

2 METHOD

2.1 Participants

Twelve (six female, six male; all right-handed) people from the local university community participated in the experiment. Their age ranged from 19 to 30 years (m = 24.08, sd = 3.99). To ensure that the motion capture equipment fit every participant, only people with body dimensions in the average range were recruited. Participants' mean height was 1.73 meters (sd = 0.08) and they had a body mass index (BMI) between 16.1 and 24.8 (m = 20.97, sd = 2.14). As the motions of the actors were later mapped onto virtual characters, relatively average body dimensions also helped minimizing artifacts caused by the mapping of the motion onto a body with fixed dimensions. None of the participants had professional experience in either dancing or ping pong playing. All participants were naive to the purpose of the experiment and had normal or corrected-to-normal vision. The experiment was approved by the ethics committee of the University of Tübingen, and was performed in accordance with the Declaration of Helsinki. Participants gave written informed consent prior to the experiment and were compensated with € 8 per hour for their participation.

2.2 Motion Capture and Stimuli Generation

2.2.1 Apparatus. The motions of the participants were captured in a large empty hall (15.30 m x 11.70 m x 8.40 m height) using a MVN Motion Capture Suit from Xsens (Netherlands). The lycra suit consists of 17 integrated single sensors, each comprised of a magnetometer, a gyrometer and an accelerometer. The sensor placement in the suit assured correct alignment with anatomical landmarks of the body. In addition, the placement of the sensors and corresponding cables within the suit enabled users to perform free and unrestricted movements and actions. Two master units in the back of the suit sent the data from the sensors wirelessly to a computer, where the MVN Studio software reconstructed the captured data on a participant-scaled human skeleton-like 3D model and recorded the resulting motion output in real-time at 120 Hz.

2.2.2 Actions. All participants performed six different actions. As recognition performance was previously found to be dependent on the action type [Loula et al. 2005], participants both performed common and uncommon actions. The following six actions were captured: (a) greeting by walking towards the experimenter and shaking hands, five times successively, (b) walking four times across the capture hall (distance of 8 meters), (c) running six rounds in the hall, three times clockwise and three times counter-clockwise (diameter of 7 meters), (d) playing ping pong against a ping pong table wall, (e) dancing spontaneously (spon.), and (f) dancing a learned choreography (chor.).

To maintain individual kinematic information, free walking and free running was chosen in contrast to other studies (e.g., Loula et al. [2005]) that used a treadmill for walking and running. By walking and running on a treadmill with fixed speed, the variation of the natural gait would have been reduced. While greeting, walking, and running are rather common actions, playing ping pong, spontaneous dancing, and performing a choreographed dance are not. For this reason, all participants practiced the movements prior to the recording. Playing ping pong was practiced for five minutes, followed by a recording time of two to four minutes depending on the playing performance. Both for practice and movement recording, a real ping pong paddle and ball were used. For the choreographed dance, a simple choreography that could easily be learned was chosen ('Macarena' by Los del Rio). Participants were trained until they were able to dance the choreography, that was then recorded by looping the refrain of the song seven times. For spontaneous dancing, participants were asked to dance for approximately two minutes to a club song ('Lady hear me tonight' by Mojo). To make them feel comfortable, all lights were switched off except for the computer monitor that was needed for operating the motion recording system. For all actions, participants were instructed to move as naturally as possible. The starting location of all actions was the same across all participants.

2.2.3 Stimulus Generation. Based on the motion capture data, for each participant and recorded action, five motion clips with a length of five seconds each were created. All motion clips were chosen to be representative of the specific action. For running, three clips contained running clockwise and two clips running counter-clockwise. For dancing (chor.), due to the short length of the motion clips, three clips showed the participants only moving their arms according to the choreography and the other two clips showed the whole jump for turning by 90 degrees. All 300 motion clips (5 actions x 5 motion clips x 12 participants) were prepared in Moven Studio from Xsens and exported for subsequent use in Unity (Unity Technologies).

To remove all shape cues of the individual participants while maintaining kinematic cues (joint angle motions), the motions were re-targeted onto three different visual representations of a body: a point-light figure, a male character, and a female character (Figure 1). The characters were average size male and female characters from the Rocketbox Studios GmbH Complete Characters Library. The characters' faces were covered with a black texture to put the focus on the body and avoid distraction by the non-animated face. The recorded motions were directly mapped onto the human skeleton model thereby keeping angular but not positional information. The point-light figure was created by attaching 16 spheres to the joints of the invisible female skeleton. All 300 motion clips were shown on each of the three visual representations, resulting in 900 video clips in total.

In the video clips, the size of the visual stimuli differed depending on the displayed action. The motion in every clip started in the center of the display and was shown from a frontal view. Dancing and ping pong playing were stationary actions, resulting in video clips where the stimuli maintained the same size. The walking motions were presented from a frontal view showing the person approaching. For running, the video clips showed the actor running from the display center to the right or left side. To ensure visibility of the actions during the whole clip, running motions were presented smaller on the screen as compared to the other actions.

2.3 Experimental Setup

During the experiment, participants sat in front of a computer screen at approximately 60 cm viewing distance. The visual stimuli

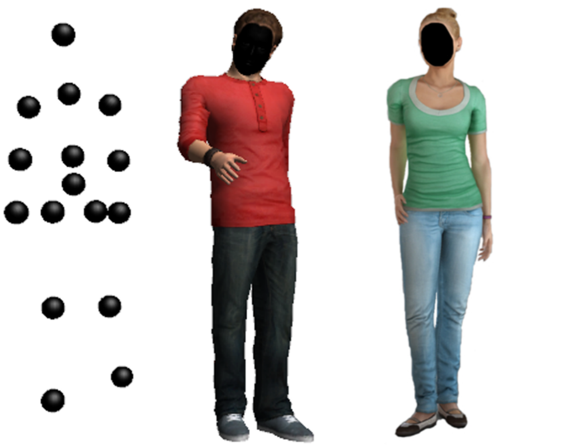

Figure 1: Visual body representations that were used in the experiment: a point-light figure, a male and a female virtual character.

were presented on a Dell UltraSharp 2007FPb 20 inch monitor running at 60 Hz with a resolution of 1280x1024 pixels. The experiment was programmed in Unity 4.0 (Unity Technologies) and was run on a Dell Precision M6400 laptop (Intel Core2 Duo 2.8 GHz, 4GB RAM, nVidia Quadro FX 3700M graphics card with 1024MB VRAM).

2.4 Design and Procedure

The experimental procedure was conducted across two sessions. In the first session, participants' motions were captured. In the second session, participants completed the self-motion recognition experiment. In the experiment, participants saw their own motions as well as the motions of all other eleven participants. The sessions were separated by six to eleven weeks (m = 9.01) similar to the inter-session time interval used by Loula et al. [2005]. This time interval was chosen to ensure that participants did not remember their specific movements during the motion capture session. The recording time of several minutes for the uncommon actions (dancing and playing ping pong) additionally reduced the likelihood that participants remembered their action performance.

Before the self-motion recognition experiment, participants were instructed that in each trial they would see either their own movement or the movement of other people on one of three different representations of a body (a point-light figure, a male character, or a female character) and their task would be to judge whether the displayed motion was their own or not.

To familiarize participants with the procedure, the experiment started with 12 practice trials, in which participants saw one greeting animation from each actor. The actions were displayed either on a point-light figure, a male character, or a female character (Figure 1). Each representation was shown four times in the practice phase, randomly chosen for each actor. Each trial started with a fixation cross that was displayed for 500 ms, followed by the five second video clip. Subsequently, two buttons with the labels 'Me' and 'Not Me' appeared on the screen and participants responded by clicking on one of the buttons using the computer mouse and confirmed their response by pressing the enter key on the keyboard.

In the test trials, the remaining five actions (walking, running, playing ping pong, dancing (chor.), and dancing (spon.)) were shown to the participants. The 900 video clips were presented in random order. Participants were not aware that they would only see their own movement in 1 out of 12 trials. After every 60 trials (duration approximately 10 minutes), participants took a short break. In total, the experiment took between two and three hours. After the experiment, participants filled out a post-questionnaire in which they rated how difficult the judgments were for each body representation on a 10 point Likert scale (1 = 'very easy', 10 = 'very hard').

3 RESULTS

On average, participants recognized their own motion in 63.6% (sd = 15.2%, range: 40% to 90.67%) of the 75 trials (25 per body representation) where their own motion was displayed (Figure 2). On the trials where participant saw the motion of another person, they erroneously responded that the presented motion was their own in 31.87% (sd = 12.24%, range: 16.12% to 56.85%) of the 825 trials (11 actors x 5 actions x 5 recordings x 3 body representations).

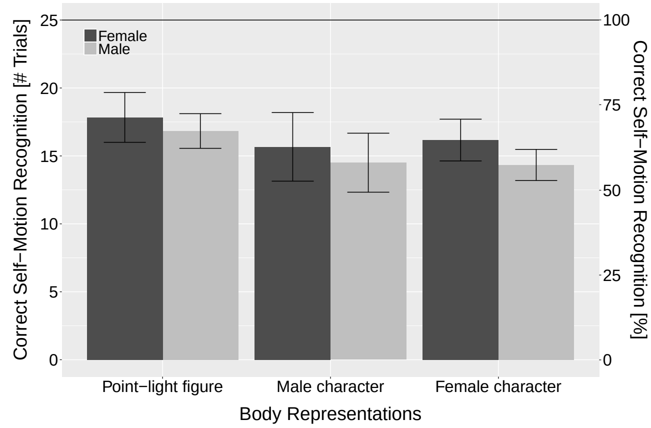

Figure 2: Performance accuracy for trials where own motion was displayed as a function of body representation and participant sex. Error bars represent one standard error of the mean. The solid line indicates perfect self-motion recognition performance.

For getting a measure of the self-motion recognition performance that is independent of the proportion of trials where participants' own motions were shown, a signal detection theory approach was used. Based on the proportion of hits (answer 'Me' when own motion was shown) and false alarms (answer 'Me' when another actor's motion was shown), d-prime (d') and the response bias (criterion, c) were calculated. A higher d' (d' > 0) indicates that the signal can be more readily detected, whereas a d' value of zero indicates chance performance. As some participants achieved a perfect hit rate or had no false alarms for certain actions and body representations, these data points were normalized using the approach described in Macmillan and Kaplan [1985] for calculating the corresponding d' and c values. For the measure of response bias, negative criterion values indicate a loose criterion (many hits, many false alarms, i.e., a bias towards giving 'Me' responses), whereas positive values indicate a strict criterion (few hits and few false alarms, i.e., a bias towards giving 'Not Me' responses).

3.1 Discrimination Sensitivity

The d-prime values are shown in Figure 3. To analyze the sensitivity of detecting own motion for the different body representations and actions, a repeated-measures analysis of variance (ANOVA) was conducted on the sensitivity measure d' with body representation (point-light figure, male character, female character) and action (running, walking, dancing (spon.), dancing (chor.), playing ping pong) as within-subject factors, and participant sex as a between-subjects factor. Because action type failed the test of sphericity, the results reported are Greenhouse-Geisser corrected.

The ANOVA revealed no influence of body representation on the sensitivity towards detecting own motion, $F(2, 20) = 0.12$, $p = .88$. This indicates that the appearance of the body had no influence on recognizing own motion, even if displayed on a virtual character of the opposite sex. Further, there was no influence of participant sex, $F(1, 10) = 1.5$, $p = .25$, suggesting that male and female participants were overall equally sensitive to recognizing their own motion.

There was a main effect of action type, $F(2.1, 20.95) = 8.84$, $p < .001$, $\eta^2 = 0.22$, suggesting that the sensitivity towards recognizing own motion was dependent on the displayed action. Sensitivity was highest for recognizing dancing (spon.) (m = 1.95, se = 0.17), followed by dancing (chor.) (m = 1.13, se = 0.16), and playing ping pong (m = 0.82, se = 0.2). Sensitivity was lowest for running (m = 0.66, se = 0.22) and walking (m = 0.55, se = 0.14). Post-hoc pairwise comparisons using Bonferroni-correction revealed a significant difference in d' between dancing (spon.) and running ($p < .001$), and dancing (spon.) and walking ($p = .004$). One-sample t-tests showed that the d' values were significantly different from 0 for dancing (chor.) ($t(11) = 4.4$, $p = .001$), dancing (spon.) ($t(11) = 7.28$, $p < .001$), playing ping pong ($t(11) = 2.57$, $p = .03$), and walking ($t(11) = 2.71$, $p = .02$). This indicates that the sensitivity towards recognizing own motion was above chance level for those actions. For running, participants had difficulties recognizing their own motion and performed around chance level as indicated by d' for this action not being different from 0, $t(11) = 1.88$, $p = .09$.

The sensitivity to the presented action varied based on participant sex, $F(2.1, 20.95) = 5.28$, $p = .001$, $\eta^2 = 0.15$. The largest difference was observed for ping pong playing, where females showed a high sensitivity to recognizing own motion as compared to males who performed around chance level. No other significant two-way or three-way interactions were found.

3.2 Response Bias

The results of the response bias (criterion) are shown in Figure 4. To analyze whether the response bias was influenced by the stimulus type, the criterion (c) for each participant, representation and action was calculated. A repeated-measures ANOVA was conducted on the criterion value with body representation (point-light figure, male character, female character) and action (running, walking, dancing (spon.), dancing (chor.), playing ping pong) as within-subject factors, and participant sex as a between-subjects factor. Because action type and body representation failed the test of sphericity, the results reported are Greenhouse-Geisser corrected.

There was a main effect of body representation, $F(1.68, 16.83) = 6.16$, $p = .008$, $\eta^2 = 0.06$. Participants were more willing to answer 'Me' to motions mapped onto the point-light figure (m = -0.14, se = 0.08) as compared to the virtual characters (female: m = 0.13, se = 0.08; male: m = 0.17, se = 0.09). Bonferroni-corrected pairwise t-tests showed that this difference in response bias between the point-light figure and the virtual characters was statistically significant (female character: $p = .014$; male character: $p = .023$).

There was a main effect of action type, $F(1.87, 18.68) = 3.75$, $p = .01$, $\eta^2 = 0.13$. Participants showed a bias towards responding 'Me', as indicated by low criterion values, for dancing (chor.) (m = -0.07, se = 0.08) and dancing (spon.) (m = -0.24, se = 0.09), and a bias towards responding 'Not Me' for playing ping pong (m = 0.04, se = 0.09), running (m = 0.18, se = 0.12) and walking (m = 0.36, se = 0.13). This effect however was dependent on participant sex, as indicated by the significant interaction between participant sex and action, $F(1.87, 18.68) = 3.75$, $p = .01$, $\eta^2 = 0.13$.

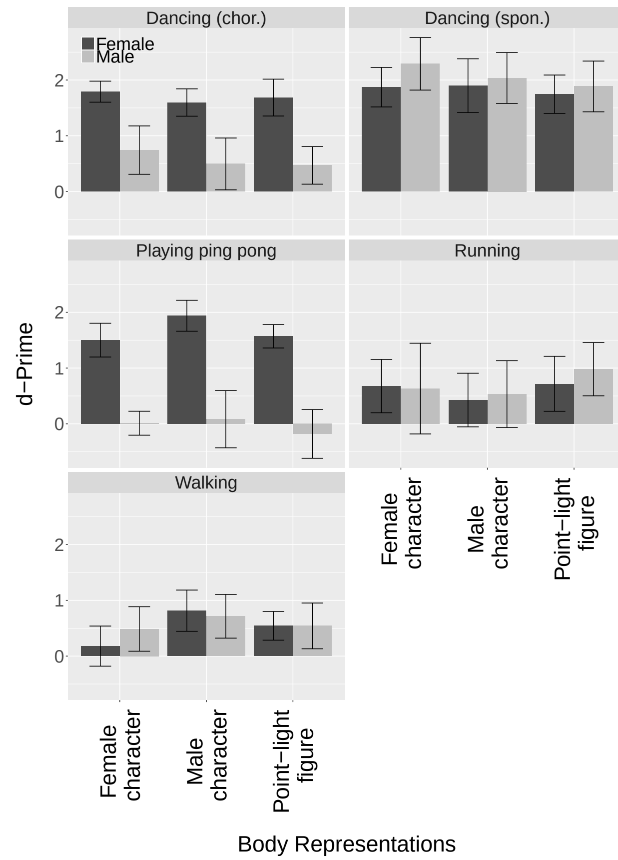

Figure 3: d-Prime as a function of displayed action, body representation and participant sex. Higher d-prime values indicate that own motion was recognized more readily, a d-prime value of 0 indicates chance performance. Error bars represent one standard error of the mean.

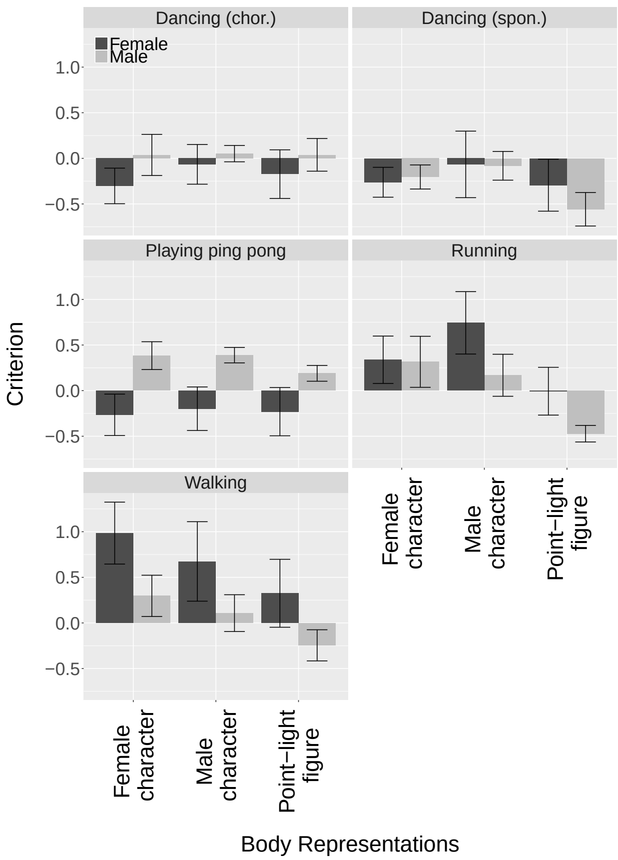

Figure 4: Response bias (criterion) as a function of displayed action, body representation, and participant sex. Higher values indicate a stricter criterion (few hits and few false alarms, i.e., a bias towards giving 'Not Me' responses). Error bars represent one standard error of the mean.

There was also a significant interaction between representation and action, $F(4.43, 44.31) = 2.92$, $p = .007$, $\eta^2 = 0.05$ indicating that the response bias for the different actions was dependent on the body representation, as can be seen in Figure 4.

3.3 False Alarms

Since previous research has found that observers can determine the actor's sex in point-light displays [Barclay et al. 1978; Hill and Johnston 2001; Mather and Murdoch 1994], we further examined whether participants' responses were influenced by whether the displayed motion was performed by a male or female actor (Figure 5). A repeated-measures ANOVA was conducted on the false alarm rates with actor sex (male, female) and body representation (point-light figure, male character, female character) as within-subject factors, and participant sex as a between-subjects factor. As body representation failed the test of sphericity, the results reported are Greenhouse-Geisser corrected.

There was a main effect of representation, $F(1.77, 17.69) = 13.39$, $p < .001$, $\eta^2 = 0.11$. Participants had a higher false alarm rate for the point-light figure (m = 38.18%, se = 4.4%) as compared to the male (m = 27.71%, se = 3.5%) and female character (m = 29.56%, se = 3.6%). Further, there was a main effect of actor sex, $F(1, 10) = 6.23$, $p = .03$, $\eta^2 = 0.06$. Participants were more willing to answer 'Me' when viewing the motion of a female actor (m = 34.96, se = 4.24) as compared to a male actor (m = 28.68, se = 3.46). There was no interaction between actor sex and participant sex or body representation indicating that viewing the motion of an actor of one's own or the other sex did not have any influence on the probability to respond that the motion was one's own.

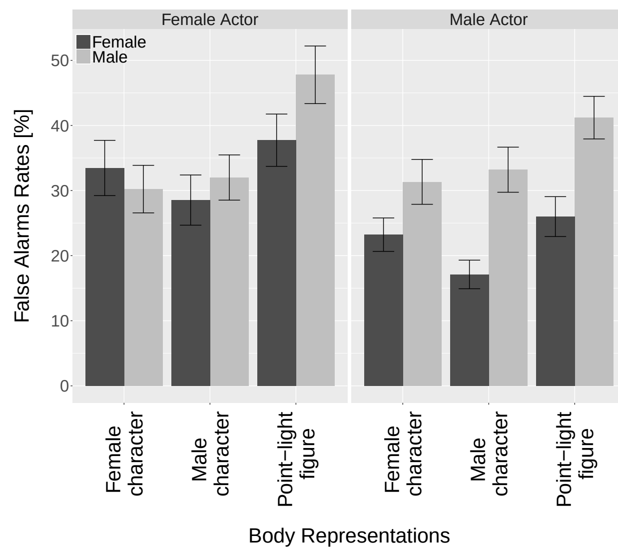

Figure 5: False alarm rates as a function of actor sex, body representation, and participant sex. Error bars represent one standard error of the mean.

3.4 Difficulty Ratings

The results of the post-questionnaire difficulty ratings are shown in Table 1. To investigate whether the difficulty ratings were dependent on the body representation, the difficulty ratings were statistically compared. Since the data was not normally distributed, the results of the difficulty ratings were analyzed using non-parametric Wilcoxon signed-rank tests. Participants rated the difficulty to judge whether their own motion was displayed on a point-light figure or a male or female character as equally difficult (male character & point-light figure: $W = 14.5$, $p = .06$; male character & female character: $W = 22$, $p = 1$; female character & point-light figure: $W = 18.5$, $p = .11$).

Further, when analyzing whether difficulty ratings were dependent on whether the sex of the virtual character was congruent or incongruent with own sex, the results show that participants rated the difficulty to judge own motions on a virtual character of own sex as easier, $W = 6$, $p = .03$ (congruent: Mdn = 5; incongruent: Mdn = 6).

4 GENERAL DISCUSSION

Avatars and avatar design are increasingly becoming relevant for immersive video games and social media. Current state-of-the art

Table 1: Participants' difficulty judgments for the visual representations on a 10-point Likert scale (1 = 'very easy', 10 = 'very hard'); MD = median.

	Point-light figure		Male character		Female character	
	range	MD	range	MD	range	MD
All participants	3-10	8	4-7	5	3-9	5
Female participants	3-8	8	4-7	6	3-5	4.5
Male participants	3-10	7.5	4-7	5	4-9	6

for animated self-avatars conveys some actions, but little to no visual fidelity of the user. The current research is therefore important for these applications as it suggests that motion information might sometimes be sufficient for self-recognition regardless of the visual fidelity and sex of the avatar.

Previous studies on self-motion recognition have used point-light displays in an attempt to reduce the available identity information to motion cues [Loula et al. 2005; ?]. However, point-light displays often contain residual body form cues, i.e., the spatial position of the joints with respect to each other, thereby indicating the actor's body shape. The aim of the current study was to investigate the importance of avatar fidelity and sex on self-motion recognition performance. We eliminated individual body configuration and shape cues by mapping the participants' own motions onto three different body representations (a point-light figure, a male and a female virtual character).

Our results show that participants were equally sensitive towards recognizing their own motions presented on a point-light figure and a virtual character of their own and opposite sex. Given that participants' motions were all mapped onto the same body representations, this finding suggests that recognizing one's own whole body motion might be mainly based on kinematic information rather than the underlying individual body shape and bodily appearance. This result is in line with previous research on self-recognition of an avatar's facial motion [Cook et al. 2012]. Similar to our approach, Cook et al. [2012] mapped individual facial motions onto a common facial form of an avatar, thereby eliminating facial shape cues. Their results suggest that recognition of self-produced facial motion was mediated mainly by kinematic information.

Our analysis of the response bias towards accepting motions mapped onto the three different body representations as one's own, revealed that participants showed a bias towards stating the displayed motion was their own when shown on the point-light figure. Previous studies have found that when the same motions were displayed on a point-light figure and human-like characters, participants tended to rate motions on a point-light figure as more biological and when displayed on a human-like character as more artificial [Chaminade et al. 2007]. Possibly, participants in our study also felt that the motions of the point-light figure looked more biological and therefore stated more often it was their own motion. Interestingly however, participants indicated in the post-questionnaire that they found it equally difficult to judge whether they saw their own motion for all three body representations, but congruency of the avatar's sex with one's own was rated to facilitate judgments.

In line with Loula et al. [2005], our results show that self-motion recognition performance was dependent on the action type. Specifically, we found a higher sensitivity to recognize infrequent actions such as spontaneous dancing, dancing a choreography and playing ping pong, as compared to frequent actions such as walking and running. The movements of infrequent actions were potentially more distinct [McDonnell et al. 2009] and therefore easier to recognize. For some of the infrequent actions, such as dancing (chor.) and playing ping pong, female participants showed a higher sensitivity towards recognizing their own motion, suggesting that for females there might have been greater inter-subject variability in performing those actions. Interestingly, even though individual kinematic information in the running actions was maintained by using free running rather than running on a treadmill, recognition performance was still at chance level. This could be due to the fact that the running stimuli were presented smaller on the screen as compared to the other actions and self-specific motion features might have therefore been more difficult to detect.

A limitation of the present study concerns that the point-light figure stimuli for *both* male and female participants were created by attaching the point lights onto the skeleton of the *female* character. Previous studies have found that observers are able to identify the actor's sex in point-light displays [Dittrich 1993]. The body configuration of the female character in our study, as defined by the joint locations, is visibly different from that of the male character. However, despite the female body configuration of the point-light figure, our results show that body representation did not have any influence on self-motion recognition thereby indicating that participants did not use information about the body configuration for making their judgments.

Future studies should consider including skill-level ratings especially for uncommon actions, such as dancing and playing ping pong, as both high and low skill levels could influence the distinctiveness of the movements. In addition, it would be interesting to evaluate the distinctiveness and attractiveness of the motions of different actors more directly and test whether this correlates with self-motion recognition performance. For example, Hoyet et al. [2013] investigated motion distinctiveness by presenting participants with video clip of three actors performing the same action side by side, followed by a single actor performing the same action. Participants' task was to indicate whether the single actor was present or absent in the previous group of three. The results showed that dancing movements were rated as more distinct as compared to walking and jogging. Interestingly, distinctiveness and attractiveness were found to be correlated such that average (less distinct) movements were rated as more attractive for common actions (running, walking), however not for uncommon actions (dancing) [Hoyet et al. 2013]. Possibly, the better self-motion recognition performance for uncommon actions in our study is related to the distinctiveness of performing these actions.

In the current study, we used five different common and uncommon actions. It would be interesting to further explore how other action types, such as conversing body motion influence self-motion recognition as audio and visual cues have previously been found to affect the ability to determine the sex of virtual characters [McDonnell and O'Sullivan 2010]. Especially for actions with known gender differences, it would be preferable to present participants only with motions of same-sex actors when investigating self-motion recognition. Previous studies on point-light displays have shown that actor sex can be identified [Barclay et al. 1978; Hill and Johnston 2001; Mather and Murdoch 1994; Troje 2002], and this might allow disregarding some movement stimuli because of obvious sex-differences in performing these actions. In our study, the analysis of the false alarm rates revealed that participants' erroneous 'Me' responses were not influenced by whether a male or female actor performed the displayed action. Thus, detecting sex-specific movement characteristics might partially depend on the stimulus type.

Further, future research should not focus just on self-motion recognition of avatars, but also on recognizing a colleague or friend's avatar in terms of motion or visual fidelity. This may be important both for people interested in being easily recognized by others, and those that are wanting to protect their privacy and not be recognized in anonymous situations. Also, in this study we only considered two biological sex categories: female and male, and did not investigate a broader spectrum of gender identity. This might especially be interesting in digital self-representation since altering visual appearance to match the full gender spectrum could easily be achievable, such that people could create a more diverse avatar to represent themselves. It would be interesting to expand the research of Dunn and Guadagno [2012] of self-avatar selection and the relationship to gender identity and personality. Specifically, future research could focus on the relevance of motion and shape cues in the process of avatar selection. Another possible future research direction is to investigate which motions are recognizable or passable as own when displayed on a self-avatar (i.e. a body scan), or avatars that varied in their visual likeness to the participant's identity. Specifically, to systematically investigate the influence of shape and motion cues on self-recognition.

Our study has important strengths. Mapping the motion of different actors onto common body representations allowed us to completely eliminate individual form cues and thereby investigate the importance of kinematic information on self-motion recognition performance. Further, by using body representations that are commonly used in modern media (virtual characters), as well as a point-light display, we were able to investigate the influence of the visual fidelity and sex of the virtual avatar on self-motion recognition performance.

ACKNOWLEDGMENTS

We wish to thank Joachim Tesch and Heinrich H. Bülthoff for technical and advisory support.

REFERENCES

Jascha Achenbach, Thomas Waltemate, Marc Erich Latoschik, and Mario Botsch. 2017. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 12.

Catharine D Barclay, James E Cutting, and Lynn T Kozlowski. 1978. Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics* 23, 2 (1978), 145–152.

Stephan Beck, Andre Kunert, Alexander Kulik, and Bernd Froehlich. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 616–625.

Randolph Blake and Maggie Shiffrar. 2007. Perception of human motion. *Annual Review of Psychology* 58, 1 (2007), 47–73.

Sarah-Jayne Blakemore and Chris Frith. 2005. The role of motor contagion in the prediction of action. *Neuropsychologia* 43, 2 (2005), 260–267.

Kristopher J Blom, Anna I Bellido Rivas, Xenxo Alvarez, Ozan Cetinaslan, Bruno Oliveira, Verónica Orvalho, and Mel Slater. 2014. Achieving participant acceptance of their avatars. *Presence* 23, 3 (2014), 287–299.

Thierry Chaminade, Jessica Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social cognitive and affective neuroscience* 2, 3 (2007), 206–216.

Richard Cook, Alan Johnston, and Cecilia Heyes. 2012. Self-recognition of avatar motion: how do I know it's me? *Proceedings of the Royal Society of London B: Biological Sciences* 279, 1729 (2012), 669–674.

James E Cutting, Lynn T Kozlowski, et al. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society* 9, 5 (1977), 353–356.

Winand H Dittrich. 1993. Action categories and the perception of biological motion. *Perception* 22, 1 (1993), 15–22.

Robert A Dunn and Rosanna E Guadagno. 2012. My avatar and me–Gender and personality predictors of avatar-self discrepancy. *Computers in Human Behavior* 28, 1 (2012), 97–106.

Facebook. 2018. Facebook Spaces VR is better with friends. https://www.facebook.com/spaces. (2018). Accessed: 2018-03-30.

Reuben Fleming, Betty J Mohler, Javier Romero, Michael J Black, and Martin Breidt. 2016. Appealing avatars from 3D body scans: Perceptual effects of stylization. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 175–196.

Harold Hill and Alan Johnston. 2001. Categorizing sex and identity from the biological motion of faces. *Current Biology* 11, 11 (2001), 880–885.

Harold Hill and Frank E Pollick. 2000. Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science* 11, 3 (2000), 223–228.

Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehee Lee, Jessica Hodgins, and Carol O'sullivan. 2013. Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 204.

Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14, 2 (1973), 201–211.

Daniel Jokisch, Irene Daum, and Nikolaus F Troje. 2006. Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception* 35, 7 (2006), 911–920.

Linden Labs. 2018a. Sansar. https://www.sansar.com/. (2018). Accessed: 2018-03-30.

Linden Labs. 2018b. Second Life. http://secondlife.com/. (2018). Accessed: 2018-03-30.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. 2005. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance* 31, 1 (2005), 210.

Jean-Luc Lugrin, David Zilch, Daniel Roth, Gary Bente, and Marc Erich Latoschik. 2016. Facebo: Real-time face and body tracking for faithful avatar synthesis. In *Virtual Reality (VR), 2016 IEEE*. IEEE, 225–226.

Neil A Macmillan and Howard L Kaplan. 1985. Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin* 98, 1 (1985), 185.

George Mather and Linda Murdoch. 1994. Gender discrimination in biological motion displays based on dynamic cues. In *Proc. R. Soc. Lond. B*, Vol. 258. The Royal Society, 273–279.

Rachel McDonnell, Sophie Jörg, Jessica K Hodgins, Fiona Newell, and Carol O'sullivan. 2009. Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Transactions on Applied Perception (TAP)* 5, 4 (2009), 20.

Rachel McDonnell and Carol O'Sullivan. 2010. Movements and voices affect perceived sex of virtual conversers. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*. ACM, 125–128.

Microsoft. 2018. AltspaceVR Be there together. https://altvr.com/. (2018). Accessed: 2018-03-30.

Simone C Mölbert, Anne Thaler, Betty J Mohler, Stephan Streuber, Javier Romero, Michael J Black, Stephan Zipfel, Hans-Otto Karnath, and Katrin E Giel. 2017. Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted. *Psychological medicine* (2017), 1–12.

Oculus. 2018. Oculus Rooms and Parties Join friends in VR with Oculus Rooms and Parties. https://www.oculus.com/blog/join-friends-in-vr-with-oculus-rooms-and-parties/. (2018). Accessed: 2018-03-30.

Ivelina V Piryankova, Jeanine K Stefanucci, Javier Romero, Stephan De La Rosa, Michael J Black, and Betty J Mohler. 2014. Can I recognize my body's weight? The influence of shape and texture on the perception of self. *ACM Transactions on Applied Perception (TAP)* 11, 3 (2014), 13.

Sapna Prasad and Maggie Shiffrar. 2009. Viewpoint and the recognition of people from their movements. *Journal of Experimental Psychology: Human Perception and Performance* 35, 1 (2009), 39–49.

Simone Schütz-Bosbach and Wolfgang Prinz. 2007. Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences* 11, 8 (2007), 349–355.

Anne Thaler, Michael N. Geuss, Simone C. Mölbert, Katrin E. Giel, Stephan Streuber, Javier Romero, Michael J. Black, and Betty J. Mohler. 2018. Body size estimation of self and others in females varying in BMI. *PLOS ONE* 13, 2 (02 2018), 1–24. https://doi.org/10.1371/journal.pone.0192152

Nikolaus F Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision* 2, 5 (2002), 2–2.

Vinoba Vinayagamoorthy, Marco Gillies, Anthony Steed, Emmanuel Tanguy, Xueni Pan, Celine Loscos, Mel Slater, et al. 2006. Building expression into virtual characters. (2006).

Anna C Wellerdiek, Markus Leyrer, Ekaterina Volkova, Dong-Seon Chang, and Betty Mohler. 2013. Recognizing your own motions on virtual avatars: is it me or not?. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 138–138.

Margaret Wilson and Günther Knoblich. 2005. The case for motor involvement in perceiving conspecifics. *Psychological Bulletin* 131, 3 (2005), 460–473.

Katja Zibrek, Ludovic Hoyet, Kerstin Ruhland, and Rachel Mcdonnell. 2015. Exploring the effect of motion type and emotions on the perception of gender in virtual humans. *ACM Transactions on Applied Perception (TAP)* 12, 3 (2015), 11.