How Important are Detailed Hand Motions for Communication for a Virtual Character Through the Lens of Charades?

ALEX ADKINS, Clemson University, USA

ALINE NORMOYLE, Bryn Mawr College, USA

LORRAINE LIN, Clemson University, USA

YU SUN, Clemson University, USA

YUTING YE, Reality Labs Research @ Meta, USA

MASSIMILIANO DI LUCA, University of Birmingham, UK, Facebook Reality Labs, USA

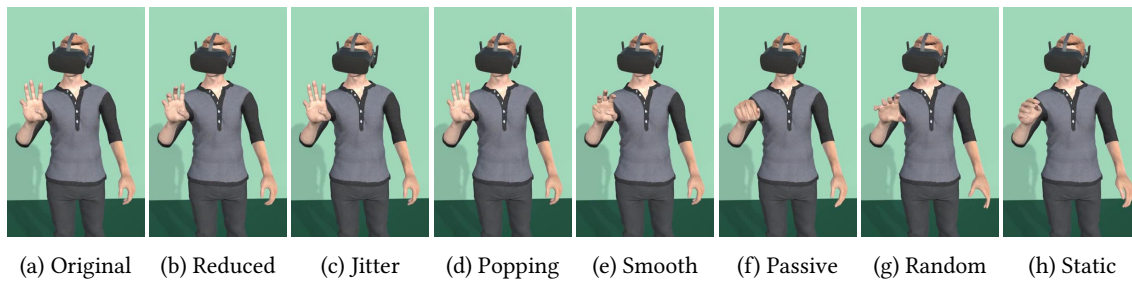SOPHIE JÖRG, Clemson University, USA

| (a) Original | (b) Reduced | (c) Jitter | (d) Popping | (e) Smooth | (f) Passive | (g) Random | (h) Static |

Fig. 1. Frame from the short motion "three" on our realistic character with different conditions modifying the motions of the hands.

Detailed hand motions play an important role in face-to-face communication to emphasize points, describe objects, clarify concepts, or replace words altogether. While shared virtual reality (VR) spaces are becoming more popular, these spaces do not, in most cases, capture and display accurate hand motions. In this paper, we investigate the consequences of such errors in hand and finger motions on comprehension, character perception, social presence, and user comfort. We conduct three perceptual experiments where participants guess words and movie titles based on motion captured movements. We introduce errors and alterations to the hand movements and apply techniques to synthesize or correct hand motions. We collect data from more than 1000 Amazon Mechanical Turk participants in two large experiments, and conduct a third experiment in VR. As results might differ depending on the virtual character used, we investigate all effects on two virtual characters of different levels of realism. We furthermore investigate the effects of clip length in our experiments.

Authors' addresses: Alex Adkins, Clemson University, USA, adkins4@clemson.edu; Aline Normoyle, Bryn Mawr College, USA, anormoyle@brynmawr.edu; Lorraine Lin, Clemson University, USA, lorrain@clemson.edu; Yu Sun, Clemson University, USA, ysun3@g.clemson.edu; Yuting Ye, Reality Labs Research @ Meta, USA, yuting.ye@meta.com; Massimiliano Di Luca, University of Birmingham, UK, and Facebook Reality Labs, USA, m.diluca@bham.ac.uk; Sophie Jörg, Clemson University, USA, sjoerg@clemson.edu.

Amongst other results, we show that the absence of finger motion significantly reduces comprehension and negatively affects people's perception of a virtual character and their social presence. Adding some hand motions, even random ones, does attenuate some of these effects when it comes to the perception of the virtual character or social presence, but it does not necessarily improve comprehension. Slightly inaccurate or erroneous hand motions are sufficient to achieve the same level of comprehension as with accurate hand motions. They might however still affect the viewers' impression of a character. Finally, jittering hand motions should be avoided as they significantly decrease user comfort.

CCS Concepts: • **Computing methodologies** → **Perception**; **Motion capture**; • **Human-centered computing** → *Virtual reality*.

Additional Key Words and Phrases: communication, hand motions, gestures, motion tracking, character animation, virtual humans, avatars

1 INTRODUCTION

We use virtual characters to communicate in many ways: carefully animated as a core element when telling an engaging story in movies and games, as embodied intelligent agents (controlled by algorithms) such as virtual tutors, and as avatars, controlled by people in real-time, e.g., VTubers or users of social VR applications. In effective interpersonal communication, hand movements play a key role [Goldin-Meadow 1999; Kendon 2004; McNeill 2008]. We use gestures to describe, emphasize, and clarify what we want to convey and even to replace words entirely. When it comes to capturing human motions or animating them, the arm and detailed hand motions are often treated separately for technical reasons [Wheatland et al. 2015]. In VR, some head mounted displays (HMDs) include cameras in the headset that can track the hands. However, hand tracking in real-time still fails frequently due to the hand being outside of the camera's field of view, occlusions, or motion blur [Ferstl et al. 2021].

Previous research has shown that small changes in finger motions are perceptible and can change the interpretation of a scenario [Jörg et al. 2010] or the perceived personality of a virtual character [Wang et al. 2016]. However, the impact of errors or lack of accurate hand motions in a communicative setting is not known. What are the consequences of inaccurate or missing hand motions on comprehension or on the impression we get of an avatar? What can we do if hand motions can not be captured at all? It is important to understand how inaccuracies and errors in these motions might affect conversations, from changing the whole message to making a conversational partner appear to be unfriendly or untrustworthy.

In this paper we aim to answer these questions by investigating the role of detailed hand motions when conveying content. We base our experiments on charades, a game which requires players to communicate exclusively with gestures and asks viewers to guess specific answers. Gestures are used differently in charades than in typical conversations. They are made to convey content on purpose and include more gestures that act out objects or concepts. The main advantage, however, of using charades in this research is that they provide us with a straightforward, quantitative way to measure comprehension: the number of correct answers. Further advantages supporting this choice are the option to use charades as videos that can be post-processed offline, and the absence of speech that eliminates a confounding factor.

While our primary goal is to investigate how detailed hand motions affect the viewer's ability to understand the character (comprehension), for a more complete investigation we also examine the effect on the personality of the

character, on the social presence of the viewer, and we ask how comfortable viewers are with the avatar in VR. These effects might be influenced by avatar appearance, which is why we incorporate two characters with different levels of realism into our study.

Our study consists of three experiments. To explore different hand motion alterations, we use Amazon Mechanical Turk to gather large numbers of participants for our first two experiments, in which participants watch videos of our stimuli. We then repeat a small subset of the conditions in a third experiment in VR. In our first experiment, the **Alteration Experiment**, we alter the hand motions of a motion captured character to simulate errors or changes that would typically happen when creating, tracking, or post-processing hand motions (see Figure 1). This experiment asks: How do hand animation and avatar appearance affect the viewer's ability to understand the character and the viewer's impressions of the character? Can we create acceptable hand motions without any data? In our second experiment, the **Intensity Experiment**, we ask: Which intensities of specific alterations, namely Jitter, Popping, and Smooth, are acceptable? What are acceptable thresholds for these errors? We vary their intensity from subtle to extreme to observe their effects. In the third experiment, the **VR Experiment**, we verify some of our results in a virtual environment and evaluate participants' comfort level when watching the character in VR.

We find that the absence of finger motion reduced comprehension and social presence. It also negatively influenced the viewer's perception of a virtual character. However, hand motions with slight inaccuracies and errors, and surprisingly even with larger errors, achieved the same levels of comprehension as accurate hand motions, but they influenced the perception of the character in some cases. A large jitter was seen as most negative, still without affecting comprehension. In the absence of captured hand motions, adding some motions, even random ones, improved the perception of the avatar and the viewer's social presence compared to no motions at all, but it did not improve comprehension. These results have important implications for the animation of engaging virtual characters, for the creation of effective virtual agents, and for the development of VR technologies.

2 RELATED WORK

This work builds up on previous research from several areas, such as motion capture and animation, virtual reality, gestures and communication, and the perception of virtual characters. We provide an overview of previous related research in the following paragraphs.

Capturing and Synthesizing Hand Motions. Capturing detailed hand motions remains an important and difficult problem, which is illustrated by the diversity of suggested approaches. Wheatland et al. [2015] provide an overview of technologies to capture detailed hand motions; Jörg et al. [2020] focus on state of the art virtual reality methods. Approaches include marker-based optical systems, individual and multiple cameras, and sensored gloves [Glauser et al. 2019; Han et al. 2020, 2018; Mueller et al. 2019; Wang et al. 2020]. Further approaches try to synthesize hand motions or to recreate them with incomplete information [Jörg et al. 2012; Schröder et al. 2015; Wheatland et al. 2013; Zhang et al. 2021; Zhao et al. 2013]. While the progress in the past years has been impressive, it is still not possible to capture accurate hand motions with consumer technology [Schneider et al. 2021]. When using HMDs, the limits of the field of view of the camera, visual occlusions, or motion blur create many hand tracking errors [Ferstl et al. 2021]. We investigate the consequences of such errors.

Perception of Hand Motions. Previous work has shown that subtle changes in the timing of hand motions can be noticed and can alter the interpretation of a character's actions [Jörg et al. 2010]. While very subtle changes can be noticed, they are not noticed all of the time: Hoyet et al. [2012] showed that the inaccuracies resulting from using a

reduced marker set with 8 markers compared to 20 markers were only detected at a significant rate for one out of nine different actions. Such changes are by far more subtle than errors in consumer capture devices and the ones we use in our experiments. Wang et al. [2016] found that hand poses and motions can convey personality. For example, spreaded fingers are perceived as conveying extraversion and openness whereas a resting hand pose conveys high emotional stability and agreeableness. Smith and Neff [2017] showed that it is possible to influence the perception of the personality of a virtual character (measured with the Big Five personality model based on the traits extraversion, openness, emotional stability, agreeableness, and conscientiousness) by editing the timing and poses of gestures. While most of the changes apply to the arm motions, they showed that extending the fingers increased extraversion and a slight disfluency — which could be compared to a slight jitter or popping — reduced conscientiousness, agreeableness, and emotional stability.

Gestures, Communication, and Charades. Regardless of whether errors are perceived and what personality is conveyed, hand and arm motions are important for communication [Goldin-Meadow 1999; Kendon 2004; McNeill 1992, 2008]. As McNeill describes, gestures orchestrate speech [McNeill 2015]. Gestures can, for example, be used to emphasize a point or to show the size of an object. Gestures can also replace words altogether, for example, to indicate that something is 'okay' by touching the index and thumb fingers together to form a circle. Altering such gestures can inadvertently misrepresent what a speaker is trying to say. McNeill [1992] defined several categories of gestures which frequently accompany speech: *iconic* where a person acts out an action or object as they describe it; *metaphoric* where a person treats an abstract concept as a physical object, for example, using a giving gesture to indicate generosity; *beat* where the hand moves with the rhythm of speech; and *deictic* which refers to pointing gestures. Different types of gestures occur during different types of speech. Iconic gestures typically accompany utterances when describing concrete objects and events, for example, during narration when telling a story, whereas metaphoric gestures can often be observed when referring to the structure of speech [Goldin-Meadow 2003; McNeill 1992]. There are also other types of categorizations, such as the ones by Kendon or Ekman and Friesen [Ekman and Friesen 1969; Kendon 2004]. Independent of the classification used, the proportions of different types of gestures vary immensely depending on the subject and type of a conversation and many other factors such as cultural background or age [Colletta et al. 2015; Goldin-Meadow 2003].

The focus of research involving gestures lies on gestures used in combination with speech. McNeill [1992] also uses the word *gesticulation* to specifically denominate speech-based gestures. The gestures during charades are categorized as *pantomime* and *emblems* on the gesture continuum. In contrast to gesticulation, pantomime is made with a purpose, it is improvised, characterized by the absence of speech, and it does not necessarily have gesture phases [McNeill 2015; Żywiczyński et al. 2018]. Emblems, such as the "thumbs up" gesture, have a specific, socially defined meaning and shape. They are also made on purpose to communicate and do not require speech. We provide further details on our charades and how they differ from other types of communication in Sections 3.1 and 7. As the gestures when performing charades are made on purpose to convey meaning and do not involve speech, they are an ideal test case for this study.

Avatar Appearance. Finally, the appearance of the character can affect how viewers interpret its motions. For example, Hodgins et al. [1998] found that viewers could detect differences in motions more easily with a polygonal character than with a stick figure. Chaminade et al. [2007] also showed that artificial and biological motions are perceived differently depending on the character model. McDonnell et al. [2012] found that realistic avatars were more unpleasant than stylistic avatars when they moved unnaturally and Baylor [2011] showed that the appearance of a virtual tutor influences the motivation of the learner. It has also been shown that the presence and appearance of avatars has an

Table 1. Summary of experiments. Both the Alteration and Intensity experiments evaluate participants' comprehension, perception of the character, and social presence based on video clips. VR Compare investigates if those concepts are perceived in the same way in a virtual environment and repeats two conditions of the Alteration experiment in VR. VR Comfort and VR Rank examine participants' comfort level when watching the character in VR and establishes a preference ranking for all conditions.

Experiment	Design	Avatar	Clip Lengths	Motion Alterations	Nb of Conditions	Description	Valid Participants
1. Alteration	Between	Mannequin/Realistic	Long/Short	8	32	Section 4	871
2. Intensity	Between	Mannequin/Realistic	Long/Short	10	40	Section 5	1198
3. VR Compare	Between	Realistic	Short (words)	Original/Static 2		Section 6	31
VR Comfort	Within	Realistic	10s clips	14	14	Section 6	31
VR Rank	Within	Realistic	Long (movies)	8	8	Section 6	31

effect on people's perceptions in virtual reality settings. The use of avatars has been shown to increase social presence in VR [Aseeri and Interrante 2021; Smith and Neff 2018]. Kilteni et al. [2013] demonstrated that people's drumming skills improved when they were given a dark-skinned self-avatar in VR. Further research has shown that the appearance of virtual hands influences the perceived ownership in VR [Argelaguet et al. 2016; Lin and Jörg 2016]. To investigate the influence of the virtual character's appearance, we create two virtual characters with different levels of realism for our experiments: a simple mannequin and a more realistic character.

In summary, hand motions are an essential aspect of body language during communication, contributing to comprehension and influencing the perception of the character. Subtle errors in hand motions have been shown to matter in some cases. In this work, we investigate the effect of such errors in hand motions when we communicate. Our results are important to help us understand how we can successfully convey the information and personality we aim to convey with virtual characters.

3 STUDY OVERVIEW

Our study consists of three experiments: Alteration, Intensity, and VR (see Table 1). In our experiments, among other questions, participants are asked to guess acted movie titles or words displayed on a virtual character, in a similar way to the game "charades" where players pantomime words or phrases.

For both the Alteration and Intensity experiments, our independent variables are the Motion Condition (Motion Alteration or Motion Intensity), Avatar, and Clip Length. Our dependent variables are the participants' comprehension, perception of character, and social presence. Both experiments use a between-group design so that no participant would be asked to guess the same movie or word more than once. Therefore, each participant viewed a series of clips with consistent motion condition, avatar, and clip length. For both experiments, participants were recruited online through Amazon Mechanical Turk.

The VR experiment was conducted in-person at [XXX] and consists of three parts: VR Compare, VR Comfort, and VR Rank. VR Compare repeats the Alteration Experiment with a small subset of conditions. VR Comfort and VR Rank evaluate whether the hand motions influence how comfortable participants feel while watching the character: VR Comfort asks participants for comfort ratings and VR Rank asks them to rank motion conditions based on comfort in direct comparison. The independent variable in all three parts is the Motion Condition. The dependent variables of the first part are the same as in the Alteration and Intensity Experiments; in the second and third part they are the participants' comfort and ranking.

Our hypotheses are as follows:

- **H1, Comprehension**: Missing or inaccurate hand motion data reduces participants' comprehension of a character. Complete absence of motion data reduces it further.
- **H2, Perception of Character**: Missing or inaccurate hand motions and character appearance will affect the perception of a character.
- **H3, Social Presence**: Missing or inaccurate hand motions or a less realistic character will reduce social presence.
- **H4, Comfort**: Missing or inaccurate hand motions will make people feel less comfortable.

3.1 Stimuli Creation

We captured a set of charades using an 18-camera Vicon optical motion capture system specifically set up to accurately capture the detailed hand motions of a standing performer. An actor wore 60 optical markers on his body and 24 markers on each hand. We then asked him to pantomime several movie titles. The actor was told that the virtual character's face would not be animated and that any facial expressions would be lost, so that he focused on his body motions when performing the charades. After verifying which movies could be guessed well based on videos of the capture, we labeled the markers of six movies and computed the skeletons for the body and each hand separately for highest possible accuracy. This process produced three separate joint skeletons – one each for the body, left hand and right hand – which were aligned using aim and point constraints and combined by reparenting each hand to the body's elbow joint. This approach ensured that the captured hand motions were not modified in any way and stayed as accurate as possible. We furthermore configured our virtual characters to match the captured skeleton rather than using retargeting, which can generate slight inaccuracies.

With this procedure we created six long motions (movies) and 15 short motions (words), which are subsets of the long motions. The long motions last between 28 and 80 seconds (mean: 39.1s) and represent the movies Back to the Future, Eat Pray Love, The King's Speech, The Lion King, The Pianist, and The Three Musketeers. Each charade starts with several emblems. Five of our charade motions start with the emblem for movie (right fist describes vertical circles close to the head as if operating an antique video camera while one looks through the left hand that is shaped like a cylinder to represent the lens) and one charade starts with the sign for book (flat hands are opened similar to the pages of a book). Then the number of words is indicated by showing the corresponding number of fingers. Further signs can be used to clarify which word is being pantomimed (first, second, third, fourth) and which syllable of a word is being described. The short motions are between 1 and 17 seconds long (mean: 6.6s) and include individual words from the movie titles such as Eat, Lion, or Three. We include thirteen short motions in our analysis; the motions "four" and "two" were included as attention checks. The long motions give participants more time to notice errors and form an impression of the avatar. The short clips give participants less redundancy to guess the meaning of the motions and allow us to get insights on how the motion alterations might affect individual gestures and on how quickly participants might form an impression of the avatar.

Fig. 2. The two avatars used in our experiments: a stylized wooden mannequin (left) and a realistic character wearing an HMD (right).

We implemented each motion condition as a filter over the original motion in Unity. Finally, for each condition, we exported videos from Unity using the RockVR Video Capture Unity plugin and trimmed them with FFmpeg. The videos had no sound.

All videos are included in the supplemental materials and will be made publicly available.

3.2 Measurements

Our goal is to measure the effect of changes in hand motions on people's comprehension, their perception of the virtual character, their social presence, and their comfort level. The full questionnaire can be seen in the Appendix, Table 4.

Motion comprehension is based on how well participants guessed the movie titles or words. Their answers were rated by two researchers on a scale of 0 - incorrect, 0.5 - partially correct, or 1 - correct. A third researcher solved any discrepancies. Guidelines for rating were established beforehand. If participants guessed a variation of the correct short word, we labeled their response as correct (e.g., "eat" and "eating," "pray" and "praying"). If they wrote down an answer with similar meaning to the correct one for the short motions or if they guessed parts of the long movie title, we labeled their response as partially correct (e.g, any words in "Eat, Pray, Love"). Answers that seemed straight forward based on the animations were also judged as correct (e.g., "monster" for "lion" or "shotgun" for "rifle" for the short clips, movies about famous composers for "The Pianist"). We averaged participants' scores into a final motion comprehension score between 0 and 1. In the Alteration and Intensity experiments, the averages ranged from 0 to 1, with a mean of 0.55. Out of 11882 answers given, 654 had discrepancies between the two initial researchers (5.5%). Inter-rater reliability was very high with an unweighted Cohen's κ of 0.90.

Our perceived comprehension measure is the mean of two 7pt-Likert scale questions that asked participants to judge how well they thought they understood the virtual character. These questions were adapted from Biocca et al.'s Networked Minds Measure of Social Presence Questionnaire [Biocca et al. 2001].

To evaluate the perception of the virtual character, we use McDonnell et al.'s Perception of Virtual Character questionnaire [McDonnell et al. 2012] as well as the Ten Item Personality Inventory (TIPI) [Gosling et al. 2003]. The

TIPI questionnaire includes two measures for each of five personality traits; one question measures the personality trait positively, the other negatively. The questionnaire is based on the Big Five model of personality that has been common in psychology since the 1990s [Costa and McCrae 1992]. It measures extraversion, agreeableness, conscientiousness, emotional stability (reversed neuroticism), and openness to experience. For analysis we follow the procedure set by Gosling et al. and flip the negative measure, then take the average of the two values as the final measure for each personality trait.

Social presence was evaluated based on the questions by Nowak and Biocca [2003]. Our social presence measure is the mean of the five social presence questions.

Finally, in VR Comfort, we asked participants to rate how comfortable they would feel interacting with this character for an extended period of time. In VR Rank, we asked them to rank how comfortable they would feel interacting with each character from most comfortable (1) to least (8).

The exact wording of all questions can be seen in Table 4 of the Appendix.

4 EXPERIMENT 1: ALTERATION EXPERIMENT

The Alteration Experiment examines the role of hand animation accuracy and character appearance on participants' comprehension, perception of the character, and social presence. A between-group design was used, where each participant saw either 15 short clips or 6 long motions on one avatar (out of 2) with one motion alteration (out of 8), leading to a total of 32 different conditions: 8 (Motion Alteration) × 2 (Avatar) × 2 (Clip Length).

4.1 Motion Alterations

Our baseline motions are the original, unmodified motion captured data (Original) and the complete lack of hand motion (Static). Additionally, we created six motion alterations based on typical errors in the motion capture process or on methods to synthesize or post-process motion data. Based on our results, we realized that our alterations can be grouped into three categories: Full motion data displays the fully accurate motion data, Partial motion data represents data that has been altered from the captured data, and No motion data includes conditions where no information on the hand motions is used and hand motions are either lacking or synthesized from scratch. In the following, we summarize and detail the eight motion conditions.

- Full Motion Data
  - **Original**: unmodified motion captured data
- Partial Motion Data
  - **Reduced**: simplified motion capture
  - **Jitter**: random noise
  - **Popping**: periodic freezes
  - **Smooth**: moving average
- No Motion Data
  - **Passive**: passive hand motion
  - **Random**: unrelated motion capture data
  - **Static**: no movement

Original. Original corresponds to the detailed, unaltered motion captured motion. These motions were recorded with a high-fidelity motion capture system and manually post-processed. This quality can typically not be achieved with real-time, consumer level equipment (yet). It is our most accurate motion.

Reduced. The reduced condition simulates a reduced marker set from Hoyet et al. [2012], assuming only 6 markers, 2 each on the thumb, index, and pinky fingers. We use the markers to get the fingertip positions for the index, pinky, and thumb. The fingertip positions for the middle and ring fingers are computed using linear interpolation. Based on the fingertip positions, we compute rotations for the finger joints using inverse kinematics. This type of motion happens when a hand tracking system is used that only records the fingertip positions [Advanced Realtime Tracking 2022].

Jitter. Jitter induces random rotation movement (jitter) along the primary rotational axes of the wrist, fingers, and thumb. This condition simulates the effects of noise from sensors, which can cause jumpiness and small fluctuations in the animation. For each frame, we compute a small random rotation pertubation by sampling an angle θ from a normal distribution: $\theta \sim \mathcal{N}(0, \sigma)$. J. Segen and S. Kumar [1998] examined the ranges of jitter in hand tracking and proposed that typical jitter in orientation are less than 2 degrees, which also corresponds to our experience. We stay consistent with this result when setting the variance σ to 0.667 to create jitter. With this setting, the angle θ stays within -2 and 2 in 99.7% of cases. We also stay within the range used by Toothman and Neff [2019] who add jitter to whole body motions to evaluate the impact of avatar tracking errors in virtual reality. They apply a rotational jitter between 0 and 0.5 degrees, then between 0 and 1 degree, and finally between 0 and 6 degrees. Jitter is also encountered in current consumer equipment and can increase in low light conditions [Oculus VR 2021].

Popping. The popping condition periodically freezes the joints of the wrist, fingers, and thumb and then pops them back to their current rotations. It simulates the effects of abrupt transitions in the motion such as those caused by temporary occlusions or loss of tracking. This type of error is common with head-mounted inside-out hand-tracking technologies when the hands leave the tracking space [Ferstl et al. 2021]. We induce popping with a freeze duration of 0.8 seconds at intervals between 7 and 9 seconds to prevent the popping from looking too regular. Pops are more visible if the hands are moving a lot. We ensured each clip had at least one pop.

Smooth. Most systems perform smoothing to counteract jitter from sensors. We implement this condition by applying an exponentially weighted average on the original animation curves of the wrist, fingers, and thumb, sampled at 30 frames per second. This smoothing technique blends the incoming frame f^{orig} with the previous computed frame f^{t-1} such that $f^t = f^{orig}\alpha + f^{t-1}(1 - \alpha)$. Choosing a lower α weights the previous values over the new value, which produces a smoother curve at the expense of loss of detail. We set α to 0.2 to simulate a slight, not too obvious smoothing that would also be used in practice in such applications.

Passive. The Passive condition uses the method developed by Neff and Seidel [2006] to implement digits that move solely under the effect of gravity. The result is a hand that seems uncontrolled and lax. The authors provide the results of simulation in a table, driven by wrist orientation, which we implement directly. The motivation for including this condition is that in cases where no information on the finger and thumb motion is available, it might look more realistic to add some motion than to have none at all.

Random Based on the same motivation as the Passive condition (some hand motion might be preferred to none), Random adds captured hand motions that might not fit the body motions: for each charade, we applied the hand motion from the next charade (order is alphabetical by title), starting at the middle of the charade to avoid similar beginnings. This technique creates somewhat random hand motions within the same style. The short clips were extracted from the resulting long motions.

Static The hand does not move. We set the wrist, fingers, and the thumb to a relaxed pose to make the effect more subtle. This condition occurs when an avatar's hands are shown but there is no detailed hand tracking, for example when using simple controllers.

4.2 Method

4.2.1 Participants. For the Alteration and Intensity experiments combined (they were run together), we recruited 1940 online participants using Amazon Mechanical Turk. A technical failure resulted in the loss of the motion comprehension data for 840 participants but preserved all other data. Participants found our experiment listed as a HIT (Human Intelligence Task) on the Mechanical Turk portal. We restricted participants to those who had an approval rate of over 95%, were located in the United States, and had not previously taken any other questionnaires distributed as part of the same project. Participants were compensated with $1 for a task that took about 10 minutes.

4.2.2 Cleaning Data. Research suggests that recruiting participants from Mechanical Turk does not lead to a significant degradation in data quality [Bartneck et al. 2015; Sprouse 2011] as long as some quality assurance is performed on the responses. Text responses were checked manually. Across both the Alteration and Intensity experiments, we excluded 39 participants due to nonsense written answers or errors playing back the video. An additional 36 participants were omitted due to non-consent or missing data. Ultimately, 1865 (96.13%) online participants remained for further processing.

To ensure the quality of the survey responses, we computed Pearson correlation coefficients for each grouping of questions as they were shown to participants: two for the TIPI measure, one for McDonnell et al.'s questions on the perception of the character, and one for the questions on social presence. Participants whose answers greatly differed from the mean in a grouping compared with others in their same condition (same motion condition, same avatar, and same clip length, 19.3 participants on average) were flagged. Those with three flags or more were omitted from the analysis (resulting in 265 being omitted), leaving 1600 (82.47% of 1940 recruited) participants total for analysis. Similar quality assurance techniques have been used in various crowd-sourced studies [Smith and Neff 2017; Sprouse 2011] with the assumption that not too many responses from an individual should deviate greatly compared to the other responses in that condition. We followed Smith and Neff's [2017] example and, after extensive testing, used the same small threshold of 0.15 in order to preserve as many responses as possible. Spot checks revealed that this method correctly excluded participants whose ratings did not seem thought through, e.g., when the same rating was given to every question, or who did not answer our attention check motions correctly.

Out of the 1600 participants (927 of which had motion comprehension data), a total of 871 participants (505 with full data) were analyzed as part of the Alteration experiment, and 1198 participants (685 with full data) were analyzed as part of the Intensity experiment. There was an overlap of 469 participants (263 with full data) because the conditions Original, Jitter(Low), Popping(Low), and Smooth(Low) were considered in both experiments. The Alteration experiment had an average of 27.2 participants for each combination of conditions, 108.9 participants per motion alteration, and 425.5 per clip length and avatar.

4.2.3 Procedure. Participants were directed to a Qualtrics survey. Participants started by signing a consent form and providing demographic information. Participants watching the long clips were introduced to the rules of charades and told that they would be asked to guess a movie title. They were then asked to select the movies they were familiar with from a selection of 45 movie covers that included the titles for the charades used in the experiment. We had planned to use this information to eliminate participants that were not familiar with the movies they had to guess, but found

that many participants were able to guess the movie titles even if they were not familiar with the actual movies. We therefore did not use this information. Participants who watched the short clips were told that they would be asked to guess a noun or verb.

For both clip lengths, participants watched the sequence of animation clips in randomized order and typed in their responses. Participants could only watch each clip once. After the video section, participants answered questions about their perceived comprehension, perception of the character, and social presence. The last question was open-ended and asked for comments and feedback.

4.3 Results and Discussion

If not otherwise mentioned, results were analyzed with an 8x2x2 repeated measures ANOVA with between-subjects factors Motion Alteration (8), Avatar, and Clip Length. As typical tests for normality do not provide reliable answers for large datasets, we inspected the distribution of the answers in the histograms. As the number of analyses run was large, p-values were adjusted for Type I error using False Discovery Rate (FDR) control over all values from the 15 measures [Benjamini and Hochberg 1995]. If significance was found post-hoc testing used Tukey HSD comparisons. Only significant results are reported. Statistics for the Alteration experiment are provided in Table 1 of the Appendix. We follow the order in Table 1 when presenting and discussing our results, starting with main effects of Motion Alteration, Avatar, and Clip Length followed by any interaction effects for each examined concept.

Comprehension. Our analysis revealed a main effect of Motion Alteration for Motion Comprehension and Perceived Comprehension; the No Motion Data conditions performed significantly worse than the Partial and Full Motion Data conditions, with the exception of a non-significant difference between Passive and Reduced for Perceived Comprehension, see Figure 3.
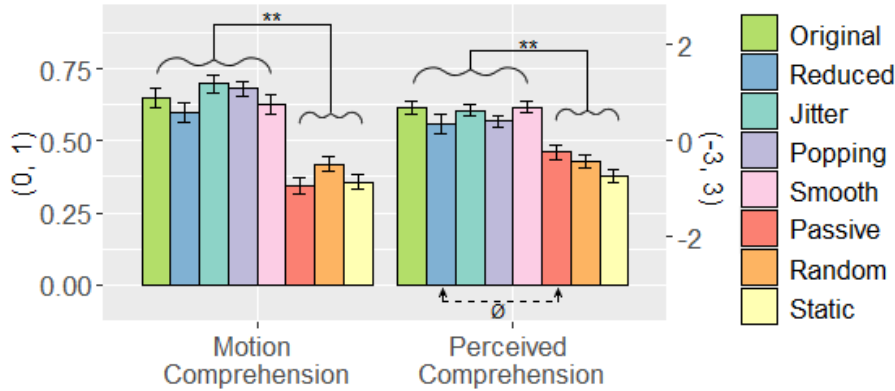

Fig. 3. Alteration Experiment. Main effects of Motion Alteration for Motion Comprehension and Perceived Comprehension: For both measures the No Motion Data conditions (Passive, Random, and Static) performed significantly worse than the Partial Data (Reduced, Jitter, Popping, Smoothing) and Full Data (Original) conditions with the exception that the difference between the Reduced and Passive motion alterations is not significant (∅) for Perceived Comprehension. Motion Comprehension values range from 0 to 1 and Perceived Comprehension is visualized in the graph as between -3 and 3 (no numbers were shown on the actual Likert scales). Error bars represent the standard error of the mean in all graphs. *** indicates a p-value of less than 0.001, ** a p-value <0.01, and * a p-value <0.05. If multiple conditions are grouped, the lowest p-value is used. These symbols are consistent throughout the paper.

These results **support part of H1, that the complete absence of motion data reduces comprehension**. This effect could not be diminished by adding synthesized motions as in the Random and Passive conditions. However, **the first part of H1 was not supported, as errors or reduced information**, at least up to the levels we tested, **did not affect comprehension** in our experiment. The hand motion data in our Partial Data conditions was sufficient to understand the meaning of our clips as correctly as when the accurate hand motion was depicted.

We also found main effects of Avatar for Motion Comprehension and Perceived Comprehension. As shown in Figure 4, participants were on average able to guess more words or movies with the Realistic avatar than with the Mannequin. Despite efforts to keep the avatars as similar as possible, including their degrees of freedom and primary colors, participants were not able to understand the Mannequin as well as the Realistic avatar. This result could be due to the fact that the shading of the hands lead to slightly less contrast for the Mannequin, or due to increased familiarity with the Realistic avatar. This result does show how important the design of the avatar is when accurate comprehension is key.
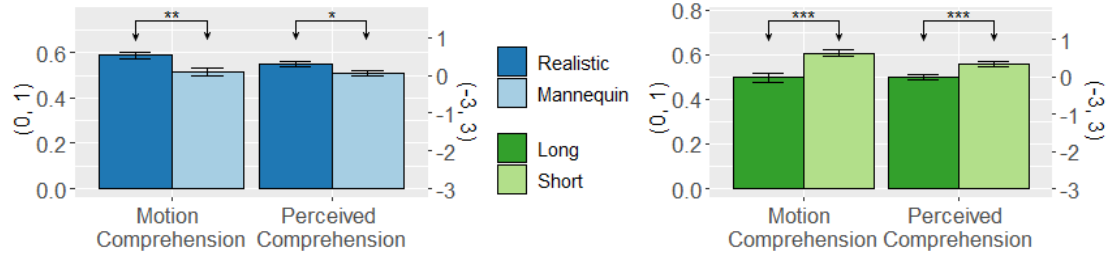

Fig. 4. Alteration Experiment. Main effects of Avatar for Motion Comprehension and Perceived Comprehension (left). Participants correctly understood the Realistic avatar significantly more often than the Mannequin. Main effects of Clip Length for both Motion Comprehension and Perceived Comprehension (right). The Long movies were and seemed significantly more difficult to comprehend than the Short words.

Furthermore, we found main effects of Clip Length for both Motion Comprehension and Perceived Comprehension, as the Long movies received lower ratings than the Short clips for both comprehension measures. The better results for the Short clips could be due to the fact that they were taken from the most comprehensible segments of the Long movies or maybe guessing words is an easier task then guessing movies. For the Long movies, participants might have guessed parts of the answer correctly but did not manage to infer the correct movie, which may have contributed to the lower comprehension scores.

Finally, there was a significant interaction effect between Motion Alteration and Clip Length for Perceived Comprehension, see Figure 5. The effect occurs because for the Full and Partial Motion Data conditions, the Long and Short clips are perceived to be similarly comprehensible, whereas for the No Motion Data conditions the Long movies were perceived to be less comprehensible than the Short clips. Interestingly, this interaction effect is not present when it comes to actual Motion Comprehension. This result may imply that the user had enough time when viewing the Long clips to realize that not everything could be understood, leading to a lower perceived comprehension. Or this difference could be attributed to the differences in tasks and a different perception of task difficulty. One conclusion could be that to achieve a high level of perceived comprehension, accurately tracked hand motions are more important in longer interactions.
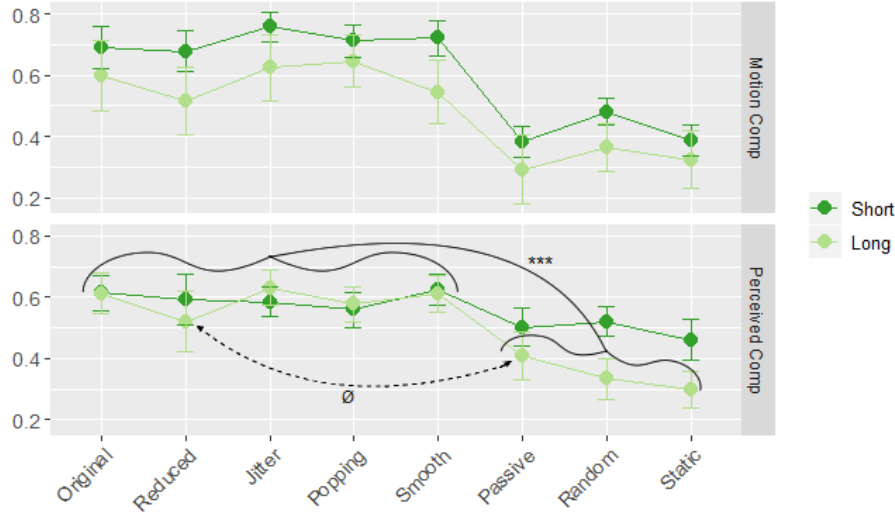
Fig. 5. Alteration Experiment. Interaction effect of Motion Alteration and Clip Length for Perceived Comprehension (bottom): The No Data alterations (Passive, Random, Static) as Long motions were rated significantly lower than many other conditions. The graphs for Motion Comprehension are shown at the top for comparison, but there is no interaction effect.

Perception of Character. Main effects of Motion Alteration were present for nine of the twelve Perception of Character measures, see Figure 6. Agreeableness, Extraversion, and Emotional Stability were the exceptions. For each measure, some of the No Motion Data conditions were rated as significantly worse than some of the Full or Partial Data conditions. In most cases the Static condition received the least positive results. The only additional significant differences affect the Naturalness measure: the Jitter condition was rated as significantly less natural than the Original condition and Random was perceived as significantly more natural than Static. The detailed significant differences for each measure are listed in Table 1 in the Appendix.

To quantify these results, we counted how often each condition received a significantly higher value (+1) or a significantly lower value (-1) than any other condition for all measures related to the perception of the character. We found the following results: Original 12, Reduced 7, Jitter 3, Popping 5, Smooth 13, Passive -5, Random -9, Static -26, confirming our observations. According to these sums, the conditions can be divided into four groups: Original and Smooth were rated most favorably followed by Reduced, Jitter, and Popping. The next group consists of Passive and Random, seen as less positive than the Reduced, Jitter, and Popping conditions. Finally, in the Static condition the character was perceived least favorably by far.

These results **strongly support the first part of H2, that changes to the hand motions will affect the participants' perception of the character**. The significant effects are nearly all based on the No Motion Data conditions being rated less favorably. These results imply that hand motions are important when it comes to a positive impression of a virtual character. Surprisingly, the Partial Motion Data conditions did not significantly change participants' perception of the character when compared to the Full Motion Data condition, meaning that **errors in hand tracking did not significantly affect how people perceive a virtual character** at least up to the levels of error we tested in this first experiment. One exception is Jitter, but even Jitter only reduced the perceived Naturalness of the character, not other measures such as Familiarity.
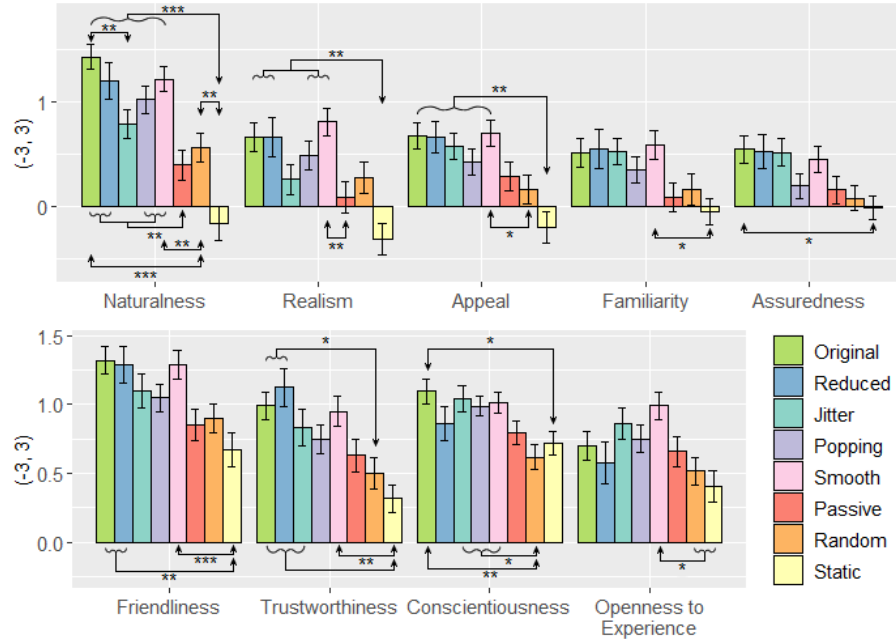
Fig. 6. Alteration Experiment. Main effects of Motion Alteration for questions on the perception of the character: Some of the No Motion Data conditions were rated significantly more negatively than the Full or Partial Data conditions in many cases, with the Static condition rated worst most often. Questions were asked on a 7-pt Likert scale, and are represented here on a -3 to 3 scale.

A closer look at our results reveals further insights:

- When some type of hand motion is added (Passive and Random conditions), our virtual characters less often receive lower ratings than without any hand motion (Static). While these conditions still perform significantly worse than selected Full Motion Data or Partial Motion Data conditions for some measures, adding some motion and having correct wrist motions seems to be advantageous.
- While there were no significant differences between the Partial and the Full Motion Data conditions (except for the naturalness of Jitter), the Original and Smooth conditions were more often significantly different from the No Motion Data conditions than the other Partial Data conditions, so Original and Smooth were rated most positively overall.

Significant main effects of Avatar were found for Realism, Appeal, Familiarity, Assuredness, and Agreeableness. In all cases, the Mannequin avatar was ranked significantly lower than the Realistic avatar. These results **strongly support the second part of H2, that changes to character appearance will affect the participants' perception of the character**. It furthermore shows that the design of an avatar is a crucial element of any application where interaction with a virtual character is important.

Main effects of Clip Length were present for Naturalness, Realism, Appeal, Familiarity, and Openness to Experience. The Long movies were rated worse than the Short words in all cases. This result is in line with our results for Motion Comprehension and Perceived Comprehension, where longer movies also performed worse. These results indicate that negative effects are more noticeable when the motions are seen for longer times. Viewers might have more time to notice

errors and imperfections. Finally, we found interaction effects between Avatar and Clip Length for several measures related to the perception of the character measures: Naturalness, Realism, Appeal, Familiarity, and Trustworthiness (Figure 7). In most cases, the Long movie clips with the Mannequin were rated significantly worse than all other conditions (see Appendix Table 1 for details).
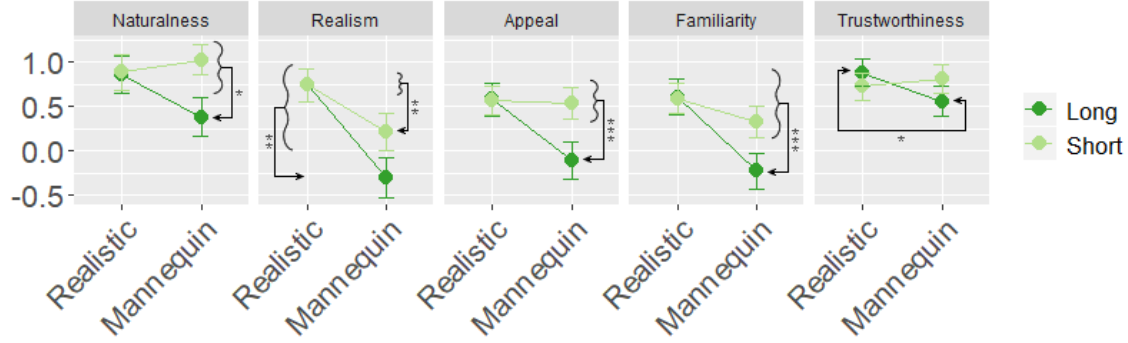

Fig. 7. Alteration Experiment. Interaction effects of Avatar and Clip Length for the questions related to the Perception of the Character were mostly due to the low ratings when the Mannequin was seen in the Long movies.

Social Presence. We found a main effect of Motion Alteration for Social Presence. Participants who watched the Static condition found that Social Presence was significantly lower than participants who watched any of the Full and Partial data conditions with the exception of Reduced, see Figure 8, left. Furthermore, there was a main effect of Avatar, with the Realistic avatar leading to significantly higher ratings than the Mannequin (Figure 8, right).
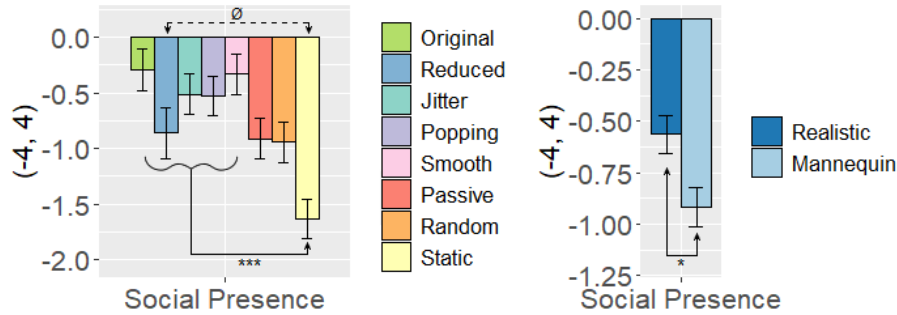

Fig. 8. Alteration Experiment. Main effects of Motion Alteration and Avatar for Social presence. Social presence responses were asked on a 9-pt Likert scale, and are represented here on a -4 to 4 scale.

These results **support H3, that less natural hand motions or a less realistic character will reduce social presence**, as the Static condition and the Mannequin both had that effect. The other No Motion Data conditions, Passive and Random, did reduce the perceived social presence on average, but considerably less and without reaching significance, implying that some motion, even if partly incorrect, is still better than none. It is unclear why the difference between Reduced and Static was not significant, maybe the decrease in detail did impact social presence for the Reduced condition. Based on these results, we also recommend to use a more realistic avatar when high social presence is desired.
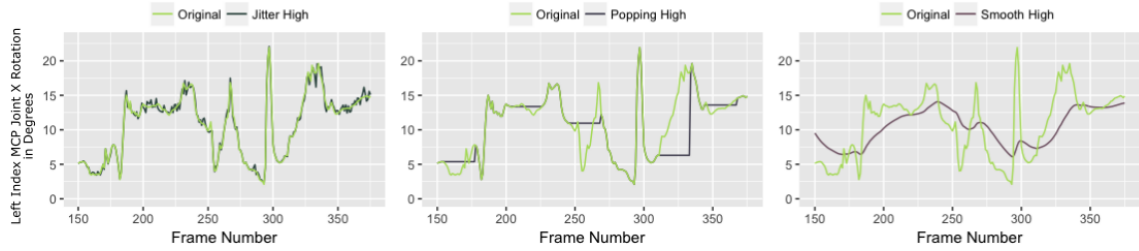
Fig. 9. Intensity Experiment. Example of JitterHigh, PoppingHigh, and SmoothHigh conditions against the Original condition curve for the first joint of the left hand index finger in the motion *The Pianist*.

5 EXPERIMENT 2: INTENSITY EXPERIMENT

In our first experiment, we found that our Partial Motion conditions did not lead to many differences compared the our Original condition. They did not reduce Motion Comprehension or Perceived Comprehension at all. However, if the intensities of these errors are increased, at some point we expect them to influence comprehension as there is no meaningful data left. For example, smoothing a motion to an extreme point would result in a static, averaged hand pose, which corresponds to our Static condition. Increasing the intensity of Popping to an extreme level would result in one or a few random poses being held for a long time, and exaggerating the Jitter condition would result in an erratic, random-type motion. The levels of errors we added were very reasonable. Therefore, in our second experiment, we test higher levels of errors with the goal of finding thresholds up to which the errors would be acceptable.

5.1 Motion Conditions

The Intensity experiment uses the same design as the Alteration experiment, but changes the levels of intensity of specific motion alterations. We include the Original condition into our analysis as a baseline. This experiment tests three different intensities (low, medium, and high) for each of the motion alterations Jitter, Popping, and Smooth, see Figure 9. The low intensities are identical to the motion alterations from the Alteration experiment (e.g., PoppingLow in the Intensity experiment is Popping in the Alteration experiment). Medium intensity doubles the parameters and high intensity quadruples them. The Jitter alteration samples a normal distribution to obtain an offset to apply to the original rotation. This distribution has variance of 0.667 degrees for low, 1.32 for medium, and 2.67 for high. To increase the intensity for Popping, we decrease the time between the pops from 7-9 seconds (low), to 3-5 seconds (medium), and 1-3 seconds (high). Note that for our Short word data set, we could not test all intensities of popping as some of the clips were too short. We ensured each clip had at least one pop. The Smooth alteration is implemented using an exponential moving average. To increase the intensity, we decrease the parameter α we use for blending. We use values of 0.2 for low smoothing, 0.1 for medium smoothing, and 0.05 for high smoothing.

5.2 Method

The Alteration and Intensity experiments were run in parallel on Mechanical Turk. The procedure in both experiments was identical as was the process to clean the data, described in Section 4.2.2. After post-processing, the Intensity experiment had 1198 participants, 685 had Motion Comprehension data. On average there were 30.0 people in each combination of conditions, 119.8 per motion intensity, and 599 per clip length and avatar.

5.3 Results

We perform the analysis of our second experiment in a similar way to our first experiment: For each of the three alterations that we varied in intensity (Jitter, Popping, and Smooth), we ran a three-way 4x2x2 ANOVA with the between-subjects factors Motion Intensity (4; Original and Low, Med, High Intensities), Clip Length (2), and Avatar (2). P-values were adjusted for Type I error using False Discovery Rate control; FDR was run over the p-values of the 15 measures for each intensity. If significant main or interaction effects were found, a post-hoc Tukey HSD revealed the detailed significant differences between conditions. Detailed results are listed in Table 2 of the Appendix.

Comprehension. There were no significant differences of Motion Intensity for Motion Comprehension or Perceived Comprehension for Jitter, Popping, or Smooth, see Figure 10.
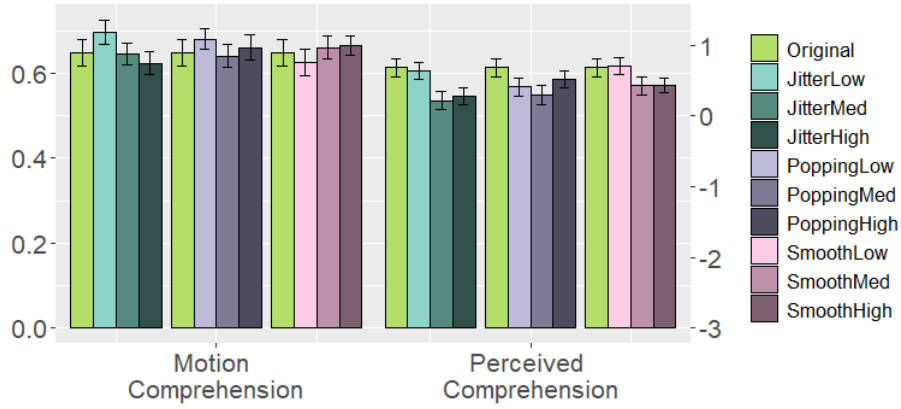

Fig. 10. Intensity Experiment. Contrary to what we expected, there were no significant differences of Motion Intensity for Motion Comprehension or Perceived Comprehension for Jitter, Popping, or Smooth even higher intensities.

This result comes as a surprise. When the changes to the original data become larger, one can see less and less of the original information. We expected comprehension measures to decrease as a result. We did not find this effect in our collected data. We conclude that relatively large errors can be applied before comprehension is affected in a significant way (at least in the way we measure it), which is good news for developers in that area. We can again **not support the first part of H1, that partially missing or inexact hand motion data reduces participants' comprehension** of a character.

We found main effects of Clip Length for Motion Comprehension for all three types of errors. The Short words were guessed correctly more often than the Long movies, which is in line with our results in the Alteration experiment. A main effect of Clip Length for Perceived Comprehension could not be confirmed in the Intensity experiment.

For Popping there was a main effect of Avatar for Perceived Comprehension; the Realistic avatar was perceived to be easier to understand than the Mannequin. This result supports our findings from the first experiment. There were no significant effects of Avatar for Motion Comprehension or for Perceived Comprehension for Jitter or Smooth, however, the Mannequin lead to lower scores on average in all five cases as well, suggesting a consistent trend.

Perception of Character. We found main effects of Motion Intensity for several measures related to the perception of the character for Jitter, see Figure 11, namely Naturalness, Realism, Appeal, Assuredness, Conscientiousness, and Emotional

Stability. In each case, JitterHigh was ranked as significantly worse than Original. For Naturalness, Assuredness, and Emotional Stability, the differences between JitterMedium and Original reached significance. JitterLow was only rated significantly worse than Original for Naturalness. Further details can be found in the Appendix in Table 2. The impact of jitter on personality is in line with results from Wang et al. [2016] and Smith and Neff [2017], who found that a resting hand pose conveys high emotional stability (jitter would be the most opposite to a resting hand pose) and disfluency in the arm motions reduces conscientiousness and emotional stability.
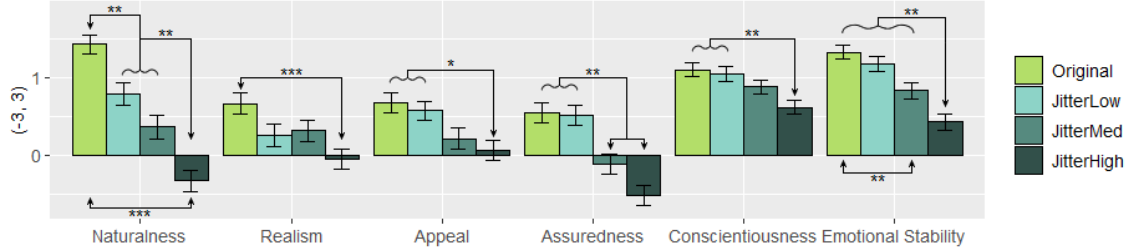

Fig. 11. Intensity Experiment, Jitter. Main effects of Motion Intensity: JitterHigh is rated significantly lower than Original in six measurements, JitterMedium in three, and JitterLow in one.

There were no main effects of Motion Intensity for Popping or Smooth, which again is surprising considering the large errors that are being introduced.

As in the Alteration experiment, a main effect of Avatar was present for multiple measures related to the perception of the character with the Mannequin avatar always yielding lower scores than the Realistic avatar. This effect was present for all three types of errors for Realism, Familiarity, and Assuredness; for Jitter and Smooth it was additionally found for Appeal. These results are expected and **further support H2, that changes to character appearance will affect the participants' perception of the character.**

Finally, we found an interaction effect of Clip Length and Avatar for the Realism measure when analyzing the intensities of Smooth, mainly based on the fact that, when watching the Long movies, the Mannequin was perceived as significantly less realistic than in all other combinations of Avatar and Clip Length (see Figure 12, right), which is in line with our results from the Alteration Experiment.

Social Presence. There was a main effect of Motion Intensity for Social Presence when analyzing the Jitter intensities. The post-hoc test revealed that Social Presence was reduced in the condition with the highest level of jitter compared to the Original condition. For Jitter, we furthermore found an interaction effect of Motion Intensity and Clip Length, visualized in Figure 12, left, mainly because participants rated social presence as significantly higher when watching the Short clips in the Original condition than for several of the other combinations with higher error intensities. For the Smooth intensities, a main effect of Avatar was based on a lower Social Presence rating for the Mannequin compared to the Realistic avatar, which is again in line with our results from the Alteration experiment. Together, these results **support H3, that less natural hand motions or a less realistic character will reduce social presence,** at least in some cases.
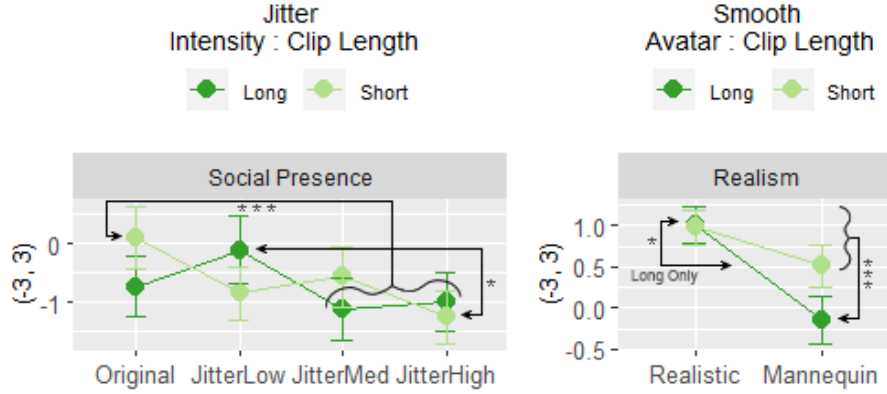
Fig. 12. Intensity Experiment. Interaction effects: For Jitter there was an interaction effect of Motion Intensity and Clip Length (left) and for Smooth we found an interaction effect of Clip Length and Avatar for the Realism measurement.

6 EXPERIMENT 3: VIRTUAL REALITY EXPERIMENT

6.1 Method

The main goal of our third experiment is to investigate if selected findings from our first experiment also apply for an avatar observed in a virtual environment or if the virtual character is perceived differently in VR. As users might have preferences that are not reflected in our measurements (comprehension, perception of the character, and social presence), we furthermore compare all of our conditions from the previous experiments in a within-subjects design and examine the viewers' comfort level with every condition and their preferences between conditions. The experiment has three parts, all three use the Realistic avatar.

The first part, **VR Compare**, recreates the Alteration experiment in virtual reality with the Original and Static conditions only. Participants wear an Oculus Rift head-mounted display (HMD) and are integrated into the same Unity scene used to generate the videos. We follow the procedure of the Alteration experiment, with the change that during the word guessing phase participants say their answers out loud so that they do not have to take off their HMD. The experimenter writes the answers down and starts the next clip. Participants are randomly assigned to see either the Original or the Static motion condition; all participants see all 15 of the Short word clips in their assigned condition. After viewing all of the motions, participants briefly remove their HMD to answer the same post-experiment questionnaire as in the Alteration and Intensity experiments, then put the HMD back on.

The second part, **VR Comfort**, asks participants to judge the viewing comfort and perceived naturalness of all motion conditions from the Alteration and Intensity experiments (14 conditions in total). In this experiment, we use random 10 second clips from each of the six charade motions. Between each clip, the experiment pauses to ask participants two questions: "How comfortable would you feel interacting with this character for an extended period of time?" (comfort) and "Please rate the naturalness of the character's motions" (naturalness). This experiment is self paced and participants choose their answers on a 7-point Likert scale using a gamepad controller. Each motion condition is shown twice (in random order), once with the Likert scales initialized with the lowest value selected and once with the scale initialized with the highest value. So each participant rates a total of 28 clips.

Fig. 13. Scene from VR Rank. Participants assigned rankings from 1 (most comfortable) to 8 (least comfortable) to each avatar. Each avatar was animated with a different condition. Participants were placed in the center of the avatars.

The third part, **VR Rank**, asks participants to rank each of the motion conditions from the Alteration Experiment from most comfortable to interact with (1) to least (8). In this experiment, participants are surrounded by eight clones in a slightly more than half circle as shown in Figure 13. Each clone is animated with a different motion condition. Placement is randomized. This configuration allows participants to make side-by-side comparisons. We decided not to show all 14 conditions based on pilot tests as the task becomes more complex and confusing with that many animated clones. This experiment is also self-paced, with no time limit. Participants assign a unique rank to each character using a gamepad controller.

We recruited 31 in-person participants through emails, flyers, and word of mouth. Upon arrival, participants fill out a consent form along with a demographics questionnaire. Next, participants put on the HMD. They start each experiment part in a virtual welcome room where they can become comfortable with the VR environment. They see a welcome screen with introductory text, which allows the experimenter to adjust the focus if necessary. During each experiment part participants can see the character in virtual reality as if they were standing in front of it. They have no virtual body of their own. Participants complete VR Compare, VR Comfort, and VR Rank. To wrap up, participants are asked open ended exit questions and compensated with a $5 gift card. For most participants, the experiment took 20-25 minutes to complete. The experiment was approved by our ethics committee.

6.2 Results

The significant results for VR Compare, VR Comfort, and VR Rank are also reported in Table 3 in the Appendix.

6.2.1 VR Compare. We used one-way ANOVAs to analyze VR Compare data with False Discovery Rate to correct for Type I errors. A main effect of Motion Alteration was found for Motion Comprehension. Participants who watched the Original condition were able to guess the words correctly significantly more often than participants viewing the Static condition (Figure 14, left). This result corresponds to our results in our first experiment and allows us to **confirm Hypothesis 1, that the absence of hand motion data reduces comprehension, for virtual environments** as well.

We also found a significant effect for Conscientiousness, with the character in the Static condition rated as significantly less conscientious than in the Original condition, which is an effect we also found in our first experiment.

There were no significant differences for our other measurements, which could be due to the smaller number of participants in this experiment. Results that are similar to the ones in previous experiments, such as the differences in Perceived Comprehension visualized in Figure 14 that did not reach significance, can be seen as supporting that explanation. However, it is also possible that the differences seen are less apparent in a virtual environment. The fact that the viewer can look around more freely in VR could contribute to these results.
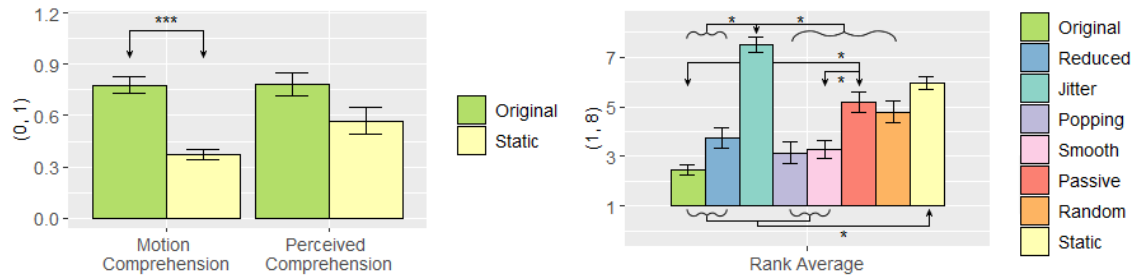

Fig. 14. VR Compare (left): Main effect of Motion Condition for Motion Comprehension. VR Rank (right): Participants were asked to rank how comfortable they would feel interacting with a character with each motion alteration from most comfortable (1) to least comfortable (8). The Original condition was rated best, the Jitter condition worst.
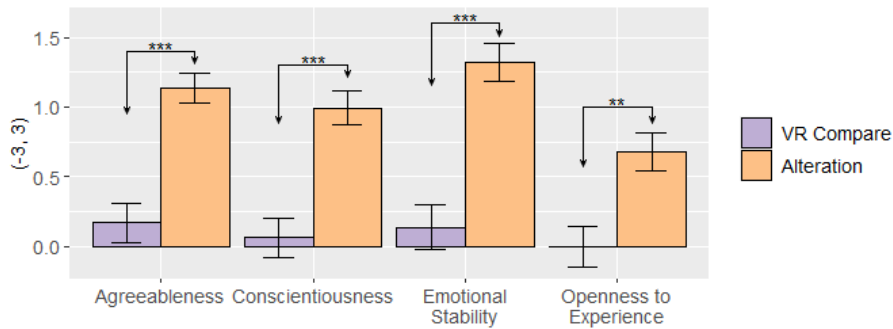

Fig. 15. Significant differences between the Alteration Experiment and VR Compare. The avatar was rated significantly more favorable when seen in videos than when seen in VR for four of our measures.
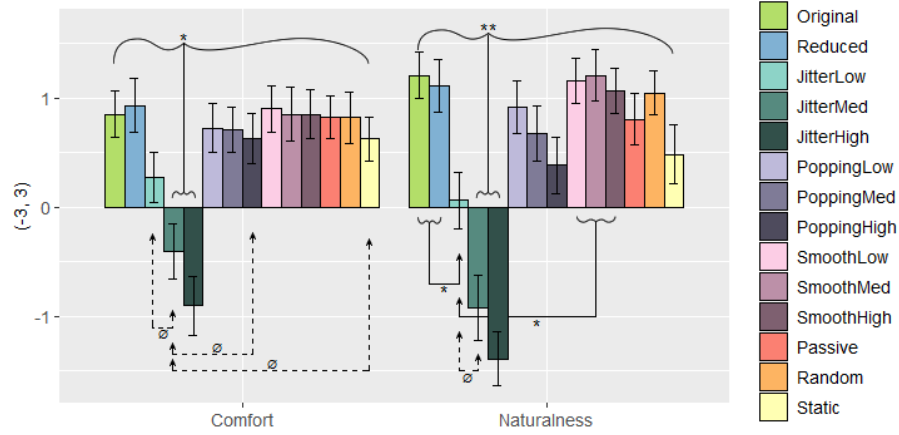
Fig. 16. VR Comfort. Main effect of Motion Condition for Comfort and Naturalness in VR. The JitterMed and JitterHigh alterations were rated significantly worse than all other conditions with few exceptions. Significant effects are visualized in the graphs; full details are available in Table 3 of the Appendix. Both questions were asked on a -3 to 3 Likert scale.

For further insights, we directly compared participants' reactions in VR Compare to those in the Alteration experiment with a two-way ANOVA with between-subjects factors Experiment (2) and Motion Alteration (2), including only participant results from the Alteration experiment under the same conditions as in VR Compare (Realistic avatar, Short word clips, Static and Original alterations). Here again, we used False Discovery Rate corrections. There were no significant effects for Motion Comprehension or Perceived Comprehension. We found several main effects of Experiment: for Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience (see Figure 15). In all cases the VR Comfort participants ranked these measures as lower on average than the participants from the Alteration experiment. So participants had a less favorable perception of the character when viewed in VR. Character design might be even more important in VR than it already is in videos.

There were no interaction effects, showing that the changes in hand motions might have a similar effect when in VR compared to when watching videos, or at least that differences were not strong enough to reach significance. The lack of interaction effects (which we expected) is not a proof that errors in hand motion have the same effect in VR than when watching videos. It is always possible that effects are present but that our experiment did not reach the power or design to reveal them. However, it is likely that any differences would be small.

A surprising result was the lack of a significant difference in social presence, which we would have expected between a video and a scene in virtual reality. We suspect that this result is due to the lack of reference. An experiment with a within-subjects design could verify this assumption.

6.2.2 VR Comfort. Using a one-way ANOVA and FDR to correct for Type I errors, we found main effects of Motion Condition for Comfort and Naturalness. In both cases, the JitterHigh and JitterMed conditions were rated significantly worse than nearly all other conditions. For Naturalness, even a small amount of jitter (JitterLow) had a significant negative effect compared to some other conditions (see Figure 16 and Table 3 for details). These results also **support Hypothesis 2**.

Comparing the results from Naturalness to the two previous experiments shows many similar tendencies as one would expect. Interestingly, the average ratings for the No Motion data conditions are much higher than they were in

the Alteration experiment. This observation could be due to the fact that participants watched a specific condition for a much shorter timespan in the VR Comfort experiment (10 seconds in the VR Comfort experiment vs. 99 seconds (Short clips) or 235 seconds (Long movies) in the Alteration and Intensity experiments). They furthermore were not asked to understand the character, but only to watch and rate it. Finally, they had other conditions to compare the motions to, which can also influence the results.

The main effect of Motion Condition for Comfort **supports Hypothesis 4, that less natural hand motions will make people feel less comfortable**. However, while this effect is strong when adding medium or high levels of jitter, the other Partial or No motion data conditions did not result in a significantly lower perceived comfort. Considering the differences between the Short clips and Long movies in the previous experiment, we think that for some of these conditions (such as Static), this result could be due to the short viewing times of each condition. In retrospect, we should have shown each condition for a longer time period to give conclusive results with regard to comfort.

6.2.3   VR Rank. Averaging the given rankings across all participants results in the following ordering: Original (most comfortable), Popping, Smooth, Reduced, Random, Passive, Static, Jitter (least comfortable).

A Friedman's ANOVA was used to analyze the ranked data between Motion Conditions (Figure 14, right) with Friedman Multiple Comparisons to determine individual differences. We found a significant main effect of Motion Alteration. Participants ranked the Jitter, Passive, and Static conditions as significantly less comfortable than several other conditions. Detailed significant differences can be found in Table 3 in the Appendix and are visualized in Figure 14. Not all of the differences are significant. For example, the differences between Original, Popping, and Smooth are statistically meaningless.

These results give further **support to Hypothesis 4, that less natural hand motions will make people feel less comfortable**. They show that Jitter should definitely be avoided, that some of the No Motion conditions reduce comfort, and that the differences between the Original motion and most of the Partial Motion conditions (Reduced, Popping, and Smooth) are not significant when it comes to comfort levels.

7   CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this paper, we investigate the effects of errors in or the lack of hand motions on comprehension, perception of a virtual character, perceived social presence, and comfort when watching the character. We summarize our key findings in Table 2 and as follows:

(1) Lack of hand motion data significantly reduces comprehension and social presence, as well as negatively affects the perception of the character, for example, appeal, friendliness, and conscientiousness are reduced.

(2) Partial or erroneous hand data at the levels we tested is sufficient in many cases to avoid negative effects. Comprehension with partial hand data is not reduced compared to accurate hand motions even with the large errors we tested; the character is perceived similarly and social presence is similar to having accurate hand motions.

(3) Adding unrelated motion to the digits and correct wrist motion does not improve comprehension but can reduce the negative effects of a fully static hand when it comes to social presence or the perception of the character. For example, social presence is not significantly reduced in the Passive and Random conditions compared to the full and partial data conditions whereas it is significantly reduced in the Static condition.

(4) Jittery motions should be avoided. While the presence of jitter did not affect comprehension, our Jitter condition was preferred least and rated lowest for comfort and naturalness.

Hand motions	Compre-hension	Perception of character	Social presence	User comfort
Fully accurate	++	++	++	++
Partial, no jitter	++	+	++	++
Partial, with jitter	++	+	+	-
Unrelated to content	-	-	+	+
No motion	-	- -	-	+

Table 2. High-level summary of our results. This summary simplifies some of the details of our results. It is intended to give a quick idea of the consequences when choosing the accuracy of hand motions. The categories ++, +, -, and − are relative and do not reflect the importance of each category in a specific application. The Original condition represents the "Fully accurate" hand motions; "Partial, no jitter" includes the Reduced, Popping, and Smooth conditions; "Partial, with jitter" includes the Jitter conditions; Passive and Random are the conditions with hand motions "Unrelated to content"; and "No motion" is the Static condition.

(5) Our more realistic avatar performed better: comprehension was higher in some cases and many of the personality ratings were more positive. The negative effects of the mannequin were more pronounced when the viewers watched longer motions.

(6) Comprehension of our realistic character for the tested conditions was similar (and not significantly different) in VR than when watching videos.

(7) Watching our character in VR created a less favorable perception of the character as opposed to watching the same character on a screen.

Our experiments confirm the importance of detailed hand motions for communication, social presence, and for accurately conveying personality. Furthermore, we found several surprising results. We expected to see negative effects when showing static hands, but also when seeing hand motions with large errors. However, errors in hand motions did not reduce comprehension, even when the errors were very obvious and larger than what one would encounter in practice as was the case in our Intensity experiment. We assume that the redundancy in motions in large enough, so that viewers are able to extract or infer the information necessary for comprehension or forming impressions of a character even if the data is incomplete or noisy. The thresholds for reducing comprehension were higher than the values we tested. Jitter was perceived negatively at lower intensities in some cases, but still did not reduce comprehension even with a high intensity.

While adding random or passive motions to the hands in the absence of data did not help with comprehension, it did at least improve social presence and how the character was perceived to some degree. Both chosen methods (conditions Passive and Random) were rather simple. It would be interesting to see if other, more complex methods of creating hand motions when no data is available might help with motion comprehension or increase, for example, the perceived naturalness.

Finally, we found that character design is important (not surprisingly), and that it might be even more important when the character is seen for longer times and in virtual reality. This might be due to viewers noticing more details when given more time and when sharing a virtual environment with a character. While in our case the realistic character was perceived more positively than the mannequin when seen for longer times, this result might be different based on the exact design of the character, e.g., if we had used a highly appealing cartoony character our results might be different.

While we were able to answer many questions with our experiments, it also has limitations. In these experiments, the virtual character plays charades. We chose this type of motions as we specifically were looking for tasks with expressive motions where body motions and hand motions might be important and for a quantitative way to measure comprehension. We were furthermore trying to avoid any confounding effects that audio might have. However, gestures are used differently when playing charades where one might use more iconic and metaphoric gestures than during typical conversations where beat gestures are more common and detailed hand motions might therefore be less important. To quantify those differences, the gestures of all of the charades motions used in this study were labeled by two graduate students and classified as iconic, metaphoric, beat, or deictic based on the descriptions by McNeill [1992; 2015]. In the 3 minutes and one second of charades, we detected 32 iconic, 37 metaphoric, 8 deictic, and not a single beat gesture. As a comparison, the same process was applied to two motion databases from Jörg et al. [2012], one called *Conversations database* that includes 8 minutes and 5 seconds of narrations and one called *Debates database* from the same actor as the charades and with 9 minutes and 34 seconds of debates. In the narrations, 42 gestures were coded as iconic, 27 as metaphoric, 9 as deictic, and 66 as beat gestures. In the debates, 28 gestures we categorized as iconic, 77 as metaphoric, 29 as deictic, and 74 as beat gestures. As expected, the charades show a higher frequency of iconic and metaphoric gestures and a lower frequency of beat gestures than the conversations or the debates. We also compare this distribution to findings from the literature: McNeill provides statistics of six cartoon narratives by English-language speaking university students. There are 261 iconic, 43 metaphoric, 28 deictic, and 268 beat gestures in an estimated 49 minutes of narration. Such a distribution would be expected in a narrative scenario and is closest to our conversations database.

While results might differ depending on the exact type of communication, accurate hand motions are likely to be less relevant in a conversational scenario with audio and more beat gestures. We conclude that, if errors in hand motions did not reduce motion comprehension when playing charades, they are unlikely to affect comprehension during a typical conversation with audio. Still, gestures are an integral part of conversations and the knowledge that the complete lack of hand motion reduces comprehension in some cases might be enough to attempt to add at least some hand motions all of the time.

A second design choice and limitation was to not animate the face of the realistic avatar to avoid confounding factors with the mannequin. Of course, the presence of detailed facial animations is likely to influence our results. We assume that differences between conditions would be less pronounced as facial animation might convey additional information and distract the viewer from the hands. Changes are likely to depend on the detailedness of the facial animation. In current VR social rooms, facial animation, if present, typically only includes motion of the jaw matching the audio. As that animation is very limited, we assume that our results apply well to current VR scenarios. However, future work will have to show the influence of accurate hand motions when detailed facial motions are present.

Future work would be able to further investigate these effects and answer further questions. Would people in a live scenario adapt and move differently or speak more clearly if errors occur in the motions? Do the results vary depending on the expressiveness of gestures, the personalities of the performers, the information conveyed, and the emotional content of the conversation? For a complete picture, many variables need to be taken into account in future work. The influence of the design of the avatar from stylized floating upper bodies with floating hands to realistic virtual characters should be investigated further. It would also be interesting to find out if hand motions can be learned that actually contribute to comprehension. Our experimental setup could serve as a test bed for such approaches.

Based on our findings, we have several recommendations for developers and animators to consider when creating virtual characters or interactions with avatars in VR. We recommend to capture at least partial hand motion whenever

possible even if they contain some errors. Smaller errors and even most of the larger ones we tested did not affect comprehension or social presence or how the character was perceived. The main exception was jitter, which should be avoided or smoothed. However, even highly jittery motion contributes to comprehension. If no hand motions can be acquired, creating some substitute motions is still better than leaving the fingers immobile when it comes to social presence and how the character is perceived.

ACKNOWLEDGEMENTS

REFERENCES

Advanced Realtime Tracking. 2022. Fingertracking. https://ar-tracking.com/en/product-program/fingertracking

Ferran Argelaguet, Ludovic Hoyet, Michael Trico, and Anatole Lecuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*. 3–10. https://doi.org/10.1109/VR.2016.7504682

Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783

Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. 2015. Comparing the Similarity of Responses Received from Studies in Amazon's Mechanical Turk to Studies Conducted Online and with Direct Recruitment. *PLOS ONE* 10, 4 (04 2015), 1–23. https://doi.org/10.1371/journal.pone.0121595

Amy L. Baylor. 2011. The design of motivational agents and avatars. *Educational Technology Research and Development* 59, 2 (April 2011), 291–300. https://doi.org/10.1007/s11423-011-9196-3

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*. 1–9.

Thierry Chaminade, Jessica Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience* 2, 3 (05 2007), 206–216. https://doi.org/10.1093/scan/nsm017

Jean-Marc Colletta, Michèle Guidetti, Olga Capirci, Carla Cristilli, Ozlem Ece Demir, Ramona N Kunene-Nicolas, and Susan Levine. 2015. Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *Journal of Child Language* 42, 1 (2015), 122–145. https://doi.org/10.1017/S0305000913000585

Paul T. Costa and Robert R. McCrae. 1992. Four ways five factors are basic. *Personality and Individual Differences* 13, 6 (1992), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I

Paul Ekman and Wallace V. Friesen. 1969. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica* 1, 1 (1969), 49–98. https://doi.org/10.1515/semi.1969.1.1.49

Ylva Ferstl, Rachel McDonnell, and Michael Neff. 2021. Evaluating Study Design and Strategies for Mitigating the Impact of Hand Tracking Loss. In *ACM Symposium on Applied Perception 2021* (Virtual Event, France) *(SAP '21)*. Article 3, 12 pages. https://doi.org/10.1145/3474451.3476235

Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. 2019. Interactive Hand Pose Estimation Using a Stretch-sensing Soft Glove. *ACM Trans. Graph.* 38, 4, Article 41 (July 2019), 15 pages. https://doi.org/10.1145/3306346.3322957

Susan Goldin-Meadow. 1999. The role of gesture in communication and thinking. *Trends in cognitive sciences* 3, 11 (1999), 419–429.

Susan Goldin-Meadow. 2003. *Hearing Gesture: How Our Hands Help Us Think.* Belknap Press.

Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D. Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, Asaf Nitzan, Gang Dong, Yuting Ye, Lingling Tao, Chengde Wan, and Robert Wang. 2020. MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality. *ACM Trans. Graph.* 39, 4, Article 87 (jul 2020), 13 pages. https://doi.org/10.1145/3386569.3392452

Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D. Twigg, and Kenrick Kin. 2018. Online Optical Marker-based Hand Tracking with Deep Labels. *ACM Trans. Graph.* 37, 4, Article 166 (July 2018), 10 pages. https://doi.org/10.1145/3197517.3201399

Jessica K. Hodgins, James F. O'Brien, and Jack Tumblin. 1998. Perception of Human Motion With Different Geometric Models. *IEEE Transactions on Visualization and Computer Graphics* 4, 4 (Oct. 1998), 307–316. https://doi.org/10.1109/2945.765325

Ludovic Hoyet, Kenneth Ryall, Rachel McDonnell, and Carol O'Sullivan. 2012. Sleight of Hand: Perception of Finger Motion from Reduced Marker Sets. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (Costa Mesa, California) *(I3D '12)*. 79–86. https://doi.org/10.1145/2159616.2159630

Sophie Jörg, Jessica Hodgins, and Carol O'Sullivan. 2010. The perception of finger motions. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*. ACM, 129–133.

Sophie Jörg, Jessica Hodgins, and Alla Safonova. 2012. Data-driven Finger Motion Synthesis for Gesturing Characters. *ACM Trans. Graph.* 31, 6, Article 189 (Nov. 2012), 7 pages. https://doi.org/10.1145/2366145.2366208

Sophie Jörg, Yuting Ye, Franziska Mueller, Michael Neff, and Victor Zordan. 2020. Virtual Hands in VR: Motion Capture, Synthesis, and Perception. In *SIGGRAPH Asia 2020 Courses* (Virtual Event) *(SA '20)*. Article 2, 32 pages. https://doi.org/10.1145/3415263.3419155

Adam Kendon. 2004. *Gesture: Visible action as utterance.* Cambridge University Press.

Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. 2013. Drumming in immersive virtual reality: the body shapes the way we play. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 597–605. https://doi.org/10.1109/TVCG.2013.29

Lorraine Lin and Sophie Jörg. 2016. Need a Hand? How Appearance Affects the Virtual Hand Illusion. In *Proceedings of the ACM Symposium on Applied Perception* (Anaheim, California) *(SAP '16)*. 69–-76. https://doi.org/10.1145/2931002.2931006

Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render Me Real?: Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Trans. Graph.* 31, 4, Article 91 (July 2012), 11 pages. https://doi.org/10.1145/2185520.2185587

David McNeill. 1992. *Hand and Mind.* University of Chicago press.

David McNeill. 2008. *Gesture and thought.* University of Chicago press.

David McNeill. 2015. *Why We Gesture: The Surprising Role of Hand Movements in Communication.* Cambridge University Press. https://doi.org/10.1017/CBO9781316480526

Franziska Mueller, Micah Davis, Florian Bernard, Oleksandr Sotnychenko, Mickeal Verschoor, Miguel A. Otaduy, Dan Casas, and Christian Theobalt. 2019. Real-time Pose and Shape Reconstruction of Two Interacting Hands With a Single Depth Camera. *ACM Transactions on Graphics (TOG)* 38, 4 (2019).

Michael Neff and Hans-Peter Seidel. 2006. Modeling Relaxed Hand Shape for Character Animation. In *Proceedings of the 4th International Conference on Articulated Motion and Deformable Objects* (Mallorca, Spain) *(AMDO'06)*. Springer-Verlag, Berlin, Heidelberg, 262–270. https://doi.org/10.1007/11789239_27

Kristine Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. 12 (10 2003), 481–494.

Oculus VR. 2021. *Introducing High Frequency Hand Tracking.* https://developer.oculus.com/blog/high-frequency-hand-tracking/

Daniel Schneider, Verena Biener, Alexander Otte, Travis Gesslein, Philipp Gagel, Cuauhtli Campos, Klen Čopič Pucihar, Matjazz Kljun, Eyal Ofek, Michel Pahud, Per Ola Kristensson, and Jens Grubert. 2021. Accuracy Evaluation of Touch Tasks in Commodity Virtual and Augmented Reality Head-Mounted Displays. In *Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Article 7, 11 pages. https://doi.org/10.1145/3485279.3485283

Matthias Schröder, Jonathan Maycock, and Mario Botsch. 2015. Reduced Marker Layouts for Optical Motion Capture of Hands. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (Paris, France) *(MIG '15)*. 7–16. https://doi.org/10.1145/2822013.2822026

Jakub Segen and Senthil Kumar. 1998. Human-computer interaction using gesture recognition and 3D hand tracking. *Proceedings 1998 International Conference on Image Processing* 3 (1998), 188–192. https://doi.org/10.1109/ICIP.1998.727164

Harrison Jesse Smith and Michael Neff. 2017. Understanding the Impact of Animated Gesture Performance on Personality Perceptions. *ACM Trans. Graph.* 36, 4, Article 49 (July 2017), 12 pages. https://doi.org/10.1145/3072959.3073697

Harrison Jesse Smith and Michael Neff. 2018. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 289.

Jon Sprouse. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods* 43, 1 (2011), 155–167. https://doi.org/10.3758/s13428-010-0039-7

Nicholas Toothman and Michael Neff. 2019. The Impact of Avatar Tracking Errors on User Experience in VR. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), 756–766. https://doi.org/10.1109/VR.2019.8798108

Jiayi Wang, Franziska Mueller, Florian Bernard, Suzanne Sorli, Oleksandr Sotnychenko, Neng Qian, Miguel A. Otaduy, Dan Casas, and Christian Theobalt. 2020. RGB2Hands: Real-Time Tracking of 3D Hand Interactions from Monocular RGB Video. *ACM Transactions on Graphics (TOG)* 39, 6 (12 2020).

Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Trans. Appl. Percept.* 13, 2, Article 9 (March 2016), 23 pages. https://doi.org/10.1145/2874357

Nkenge Wheatland, Sophie Jörg, and Victor Zordan. 2013. Automatic Hand-Over Animation Using Principle Component Analysis. In *Proceedings of Motion on Games* (Dublin 2, Ireland) *(MIG '13)*. Article 175, 6 pages. https://doi.org/10.1145/2522628.2522656

Nkenge Wheatland, Yingying Wang, Huaguang Song, Michael Neff, Victor Zordan, and Sophie Jörg. 2015. State of the art in hand and finger modeling and animation. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 735–760.

He Zhang, Yuting Ye, Takaaki Shiratori, and Taku Komura. 2021. ManipNet: Neural Manipulation Synthesis with a Hand-Object Spatial Representation. *ACM Trans. Graph.* 40, 4, Article 121 (jul 2021), 14 pages. https://doi.org/10.1145/3450626.3459830

Wenping Zhao, Jianjie Zhang, Jianyuan Min, and Jinxiang Chai. 2013. Robust realtime physics-based motion control for human grasping. *ACM Transactions on Graphics (TOG)* 32, 6, Article 207 (2013), 12 pages.

Przemysław Żywiczyński, Sławomir Wacewicz, and Marta Sibierska. 2018. Defining Pantomime for Language Evolution Research. *Topoi* 37, 2 (June 2018), 307–318. https://doi.org/10.1007/s11245-016-9425-9