The Perception of Conversational Gestures in Virtual Characters

Grace Lim
Rhodes College
Memphis TN, USA
limhu-23@rhodes.edu

Jacob Justice
Clemson University
Clemson SC, USA
jujusti@clemson.edu

Alex Adkins
Clemson University
Clemson SC, USA
adkins4@clemson.edu

Sophie Jörg
Clemson University
Clemson SC, USA
sjoerg@clemson.edu

ABSTRACT

As social platforms in virtual reality become more common, user hand tracking and the animation of virtual hands in those spaces becomes more and more important. Hand gestures play a key role in supplementing speech and conveying unspoken information. We conducted an experiment to determine the effects of various hand gesture alterations on comprehension and character perception in a conversational setting. Participants on Amazon's Mechanical Turk viewed videos on a screen of four different conditions of hand motion alterations in three different types of conversational settings with or without sound, then indicated how much they perceived certain features in virtual characters. The majority of the responses received were of very low quality, meaning that the set of data fit for analysis was very small. It was found that an interaction between the presence of sound and hand motion conditions had a significant affect on the perceived realism of the characters, but this result is not well supported due to the small size of the data set used.

1 INTRODUCTION

Virtual reality (VR) is slowly becoming a more popular space for social interactions, where users are represented through virtual avatars. Hand gestures are important in face-to-face conversations, used as an aid to emphasize statements, to convey emotions, and to even help visualize ideas or clarify points. Previous works indicate that even minute modifications to hand motions are noticeable and affect how gestures are interpreted [Jörg et al. 2010] and how the virtual avatar's personality is perceived [Wang et al. 2016]. Many VR spaces, however, do not accurately capture or animate hand motions. In this experiment, we modify hand movements and simulate errors and error-correcting methods in hand tracking to answer questions about comprehension, character perception, and user comfort in the presence of hand tracking errors. We use motions from casual conversation, giving directions, and arguing a controversial topic in order to encapsulate a wide variety of gestures used in normal conversation. Participants on Amazon Mechanical Turk were showed these motions with various alterations applied to them, half with the corresponding audio and half without.

Because the quality of the responses from participants was so poor, the results from this experiment cannot be validated. However, an interaction between sound and hand motion conditions was found to have an effect on the perceived realism of virtual characters.

DREU '22 Final Report, July 28-29, 2022, Clemson, SC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 RELATED WORKS

Hand motions have been shown to play an integral part in communication as a way to convey unspoken information that enriches conversation [Goldin-Meadow 1999]. They also have been shown to have a major role in the perception of virtual characters. Previous works indicate that hand motions can display personality as quantified by the Big Five model [Wang et al. 2016]; by modifying hand motions, the perceived personality of the virtual avatar can be manipulated. For example, a resting, relaxed hand position was shown to indicate emotional stability, while fists or hand poses with fingers spread indicates the least stability. The presence of any gestures at all also greatly improves the quality of communication, while the absence of any movements greatly deteriorates communication [Smith and Neff 2018].

There are a variety of ways to capture hand motions, including optical marker-based systems [Han et al. 2018], marker-less real-time tracking through cameras in head mounted displays (HMDs) [Han et al. 2020], and motion-sensing gloves [Glauser et al. 2019]. When motion capture fails, there are also many ways to simulate hand motions, such as a physics-based or neural network-based generation of grasping motions [Zhang et al. 2021; Zhao et al. 2013], or data-driven algorithms [Lee et al. 2019]. However, consumer technology is limited and capturing accurate hand motions is currently not possible. For example, hand tracking by HMDs is restricted by their fields of view, motion blur, and visual obstructions [Ferstl et al. 2021]. We examine the effects of the errors produced by these complications.

Communicative gestures can be split into categories: iconic, metaphoric, beat, and deictic, and are used in conjunction with speech [McNeill 1992]. This experiment builds off of previous research done with charades [Adkins et al. 2022], a gesture-rich form of communication where iconic, deictic, and metaphoric gestures are more common. These include pointing motions, motions that act out what they describe, and motions that treat concepts as physical objects. It was found that the absence of any hand motions negatively impacted character perception, adding even erroneous motions can mitigate some of these effects, and that jittering motions also negatively impact audience comfort. To expand upon this work, this experiment examines gestures used in conversation rather than in charades, where there are more likely to be beat gestures and less likely to be iconic gestures. Additionally, this experiment includes speech that accompanies hand gestures, which greatly affects both the gestures used by the speaker [Goldin-Meadow 1999] and user comprehension and interaction with the virtual character. Since this type of communication is more common, exploring this area and different types of gestures will better inform how hands are tracked and animated in virtual spaces.

In summary, hand motions are important to communication, adding another dimension to speech which improves comprehension and affects perception of virtual characters. In previous research, the effects of errors from motion tracking in a game of charades was investigated. This experiment extends that work by exploring different types of gestures and contexts to further improve the way virtual characters are implemented in VR.

3 METHOD

Motions were taken from three different databases from [Jörg et al. 2012], which contain gestures used in casual conversation, debate, and giving directions. For the conversation database, an actor was asked to converse about various topics, such as a past vacation or project. For the debate database, a different actor was asked to talk about various controversial topics such as creationism being taught in schools and the viability of nuclear power. For the directions database, an actress was asked to give directions to and from certain locations. Their gestures were captured using optical Vicon motion capture systems with 13 to 18 cameras, and with 56 markers on each actor's body and 22 on each hand. The skeletons for the body and each hand were computed separately and then assembled using aim and point constraints to ensure that the hand motions were kept as accurate as possible. For our experiment, motion conditions were applied to hands only.

We use three different models, shown in Figure 1, two male and one female to match the gestures and audio from the corresponding actors. These models were created using Ready Player Me's avatar platform. We animated each model's eyes with the Realistic Eye Movements Unity package [Knabe 2022], with which the models look at the camera about 10% of the time to simulate natural eye movements during conversation. Audio quality was improved and synced with animations using Audacity, and the models' mouths are animated with the Oculus Lipsync package for Unity [Oculus 2019], which simulates mouth movements by analyzing audio input. Videos for each motion condition were generated by the RockVR Video Capture Unity plugin and trimmed with FFmpeg, then cut into 10-30 second length clips. Clips were chosen based on audio quality and variety in hand gestures and content.

For this experiment, we implemented a total of four different motion conditions: the original captured motion, a complete lack of hand motions, altered motions which create random noise, and a condition to simulate when hands go out of bounds of a HMD's field of view. Each motion condition is outlined below.

- **Original**. This is the unmodified, accurately captured motion.
- **Static**. The hand does not move and is set in a relaxed position to make motions look more natural.
- **Jitter**. Small, random rotations are applied to the primary axes of the hands, fingers, and thumbs, which causes the hands to jerk and twitch unnaturally, simulating random noise being picked up by sensors. At each frame, the rotation values are sampled from a normal distribution. In the work done by Adkins et. al [2022], which this experiment extends, jitter was found to have a significant effect on the perceived naturalness of virtual characters and greatly decreased user comfort. However in modern motion tracking, smoothing is applied to hand motions, meaning jitter rarely occurs in a typical VR setting. As such, only a low intensity jitter is used, where rotations are sampled from a normal distribution with a standard deviation of 0.667.
- **Out of Bounds**. When hands leave an area that approximates an Oculus Quest's field of view, they are frozen in place. Once they reenter the area, they move as normal again, causing the hands to snap into their new position. This looks like a sudden jolt in hand movements, and mimics hand tracking being lost by a headset once hands exit the range that the HMD can cover [Ferstl et al. 2021].

Figure 1: Each of the three virtual characters participants saw in each clip.

3.1 Procedure

Participants were recruited using Amazon Mechanical Turk and were shown 12 different clips in a Qualtrics survey online. After participants signed a consent form, all clips were shown in random order. No clip was shown to each participant more than once, and each of the four conditions was shown three times, once for each database. Each clip was played automatically, and participants were only allowed to watch them once. Half of the participants were shown all clips with their original audio, and half were shown all clips without any audio. They were both asked about the subject matter of each clip, their perceived comprehension and perception of the virtual character, and their comfort with it. At the end, the participants were asked for comments and feedback.

Participants were also shown one to two attention checks. Those who saw the clips with audio saw one visual attention check and two attention checks to determine if they listened to the audio. Participants who saw the clips without audio only saw the visual attention check.

The quality of responses from participants in this version of the experiment were incredibly poor. Out of the 20 participants from the first trial, only three passed the attention checks in a satisfactory manner. We adjusted this version of the experiment two times, once to raise the reward for completion from $1.50 to $2.00, and once to make the wording of the visual attention checks clearer. With

Survey version	Number of questions per clip	Conditions shown	Number of clips shown	Reward	Estimated duration (minutes)	Average duration (minutes)	Attention check success rate	Number of unique locations[*]
1	15	All	12	$1.50	20-25	32.63	3/20	6
2	15	All	12	$2.00	20-25	33.55	4/18	11
3[**]	15	All	12	$2.00	20-25	29.96	0/20	7
4	7	All	12	$2.00	20-25	33.16	6/20	5
5	15	Original	5	$1.50	5-10	24.92	5/20	7

Figure 2: Details of each iteration of the survey.

[*] Over all iterations of the survey. Out of a total of 91 participants, 36 had unique physical locations.
[**] Visual attention checks were made clearer in this version

these changes, only four out of a total of 40 participants passed these attention checks to a satisfactory degree.

In order to address any sort of fatigue participants may have experienced while taking the original version of the survey, a new version was created, which only asked participants to answer one third of the original questions, the questions about how the virtual characters were perceived. In this version, six out of the 20 participants successfully passed the attention checks. Another version was created that only asked participants to view one hand motion condition, which included three different clips and one attention check, and only five out of 20 participants who took this version passed the attention check. Details for every iteration of the survey are described in Figure 2.

Data Cleaning. Due to time restraints, these five iterations of the survey were considered for analysis and no other trials were conducted. The version that only displayed one condition to participants was disregarded because the data was not complete enough. Additionally, to be consistent with the version of the survey that had a reduced amount of questions, only half of the set of questions was considered. These questions were about how they perceived the characters: their naturalness, realism, appeal, familiarity, assuredness, and trustworthiness. These questions were used in prior works investigating the perception of virtual characters [McDonnell et al. 2012]. They ask participants to rank how much each feature was present in the virtual characters on a 7-point Likert scale.

The data was further cleaned based on three different criteria. One, all participants who passed the attention check and provided seemingly meaningful responses, as opposed to entering the same response for every question, were included. Two, all participants who did not pass the attention check, but provided seemingly meaningful responses and provided feedback for the researchers that indicated that they paid attention to the survey were included. Three, any participant with a unique physical location at the time of taking the survey as recorded by Qualtrics were included. Over one third of the participants over all iterations of the survey had a recorded location and demographic information that was exactly the same as at least one other participant, indicating that the survey was somehow being taken by the same people multiple times. These participants were not considered in the analysis.

After the data was cleaned, only 15 sets of responses remained to be used in the analysis.

4 RESULTS AND DISCUSSION

We analyzed the data we had using factorial repeated measures ANOVA, and found that there was one significant interaction between the two independent variables, the presence of sound and hand motion conditions, on the perceived realism of the virtual characters, with $F(3, 39)=3.88$ and $p<0.05$. A Bonferroni post-hoc test showed no significant differences between the two. The distribution of results of perceived realism are shown in Figure 3, and the interaction graph is as shown in Figure 4. When viewing the out of bounds (OOB) and Static conditions with sound, the perceived realism appears to decrease significantly.

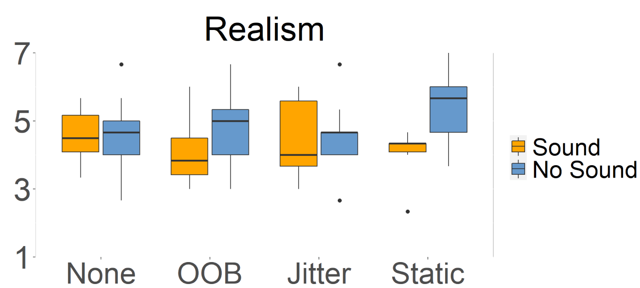

Figure 3: Distribution of results for the perceived realism of virtual characters.

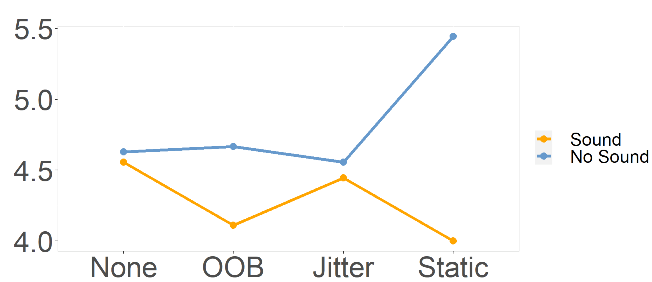

Figure 4: Interaction graph for realism with sound and hand motion condition.

4.1 Discussion and Limitations

Conditions that made the motion of the virtual characters seem unnatural, such as jitter and the out of bounds condition, were expected to have a much greater effect on how characters were perceived. However, because the set of data used in the analysis was so small, the results obtained may not be meaningful, and may have looked different if there were more high-quality responses. As limitations of this experiment largely came from the quality of responses from Amazon's Mechanical Turk, recommended future work is to run the experiment through a different service with quality control over its participants or in person, in order to ensure that a high quantity of high quality data can be gathered.

As seen in Figure 5, the distribution of results before the data was cleaned had a consistently larger spread and a less varied mean than the results after the data was cleaned, as seen in 6.

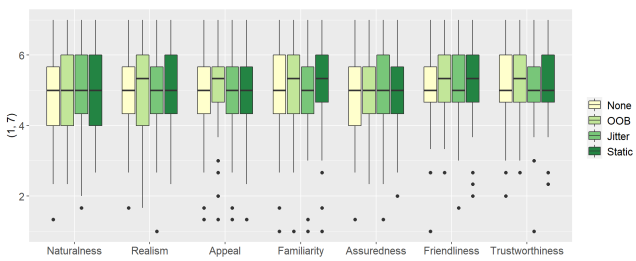

Figure 5: Distribution of results for seven of the questions in the survey for all responses in the first four iterations of the survey.

Figure 6: Distribution of results for seven of the questions in the survey for responses that remained after data cleaning in the first four iterations of the survey.

Additionally, we recommend implementing other hand motion conditions in future works. Possible other conditions include database-driven hand motion simulation, which is a potential way to correct for missing hand motions and has been validated for a certain number of gestures [Jörg et al. 2012], and a condition where hand motions are smoothed, which is a more common method of error-correction in hand-tracking.

ACKNOWLEDGMENTS

This research was supported by AccessComputing through the Computing Research Association and the National Science Foundation under Grant No. CNS-0540615, CNS-0837508, CNS-1042260, CNS-1539179, CNS-2137312 and Grant No. IIS-1652210.

REFERENCES

Alex Adkins, A Normoyle, L Lin, Y Sun, Y Ye, M Di Luca, and S Jörg. 2022. How important are detailed hand motions for communication in a virtual character? *ACM Transactions on Graphics (TOG)* (2022).

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 (SAP '21)*. Article 3, 12 pages. https://doi.org/10.1145/3474451.3476235

Oliver Glauser, Shihao Wu, Daniele Panozzo, Otman Hilliges, and Olga Sorkine-Hornung. 2019. Interactive Hand Pose Estimation Using a Stretch-sensing Soft Glove. *ACM Transactions on Graphics (TOG)* 38, 41 (2019), 1–15. https://doi.org/10.1145/3306346.3322957

S Goldin-Meadow. 1999. The Role of Gesture in Communication and Thinking. *Trends Cogn Sci.* 11, 41 (1999), 419–429. https://doi.org/10.1016/s1364-6613(99)01397-2

Sangchen 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 Transactions on Graphics (TOG)* 39, 87 (2020), 87:1–87:13. https://doi.org/10.1145/3386569.3392452

Sangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D Twigg, and Kenrick Kin. 2018. Online Optical Marker-based Hand Tracking with Deep Labels. *ACM Transactions on Graphics (TOG)* 37, 166 (2018), 1–10. https://doi.org/10.1145/3197517.3201399

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* (Los Angeles, California) *(APGV '10)*. 129–133. https://doi.org/10.1145/1836248.1836273

Sophie Jörg, Jessica K. Hodgins, and Alla Safonova. 2012. Data-driven Finger Motion Synthesis for Gesturing Characters. *ACM Transactions on Graphics* 31, 6 (November 2012), 189:1–189:7. https://doi.org/10.1145/2366145.2366208

Tore Knabe. 2022. Realistic Eye Movements. https://assetstore.unity.com/packages/tools/animation/realistic-eye-movements-29168. Accessed July 2022..

Gilwoo Lee, Zhiwei Deng, Shugao Ma, Takaaki Shiratori, Siddhartha Srinivasa, and Yaser Sheikh. 2019. Talking With Hands 16.2M: A Large-Scale Dataset of Synchronized Body-Finger Motion and Audio for Conversational Motion Analysis and Synthesis. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 763–772. https://doi.org/10.1109/ICCV.2019.00085

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 Transactions on Graphics (TOG)* 31, 91 (2012), 1–11. https://doi.org/10.1145/2185520.2185587

David McNeill. 1992. *Hand and Mind: What Gestures Reveal about Thought.* University of Chicago Press., Chicago, Illinois, USA.

Oculus. 2017-2019. Oculus Lipsync. https://developer.oculus.com/downloads/package/oculus-lipsync-unity/. Accessed July 2022..

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 (CHI '18)*. Article 289, 12 pages. https://doi.org/10.1145/3173574.3173863

Yingying Wang, Jean E Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the impact of hand motion on virtual character personality. *ACM Transactions on Applied Perception (TAP)* 13, 2 (2016), 1–23. https://doi.org/10.1145/2874357

He Zhang, Yuting Ye, Takaaki Shiratori, and Taku Komura. 2021. ManipNet: Neural manipulation synthesis with a hand-object spatial representation. *ACM Transactions on Graphics (TOG)* 40, 121 (2021), 1–14. 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, 207 (2013), 1–12. https://doi.org/10.1145/2508363.2508412