Do We Measure What We Perceive? Comparison of Perceptual and Computed Differences between Hand Animations

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

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

Tommy Dong
Clemson University
Clemson SC, USA
dong6@clemson.edu

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

ABSTRACT

An increased interest in public motion capture data has allowed for the use of data-driven animation algorithms through neural networks. While motion capture data is increasingly accessible, data sets have become too large to sort through manually. Similarity metrics quantify how different two motions are and can be used to search databases much faster when compared to manual searches as well as to train neural networks. However, the most popular similarity metrics are not informed by human perception, resulting in the potential for data that is not perceptually similar being labeled as such by these metrics. We conducted an experiment with hand motions to identify how large the differences between human perception and common similarity metrics are. In this study, participants watched two animations of hand motions, one altered and the other unaltered, and scored their similarity on a 7-point Likert scale. In our comparisons, we found that none of the tested similarity metrics correlated with human judged scores of similarity.

CCS CONCEPTS

• **Computing methodologies → Perception**; **Motion processing**.

ACM Reference Format:
Jacob Justice, Alex Adkins, Tommy Dong, and Sophie Jörg. 2022. Do We Measure What We Perceive? Comparison of Perceptual and Computed Differences between Hand Animations. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '22 Posters), August 07-11, 2022*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532719.3543233

1 INTRODUCTION

More motion capture data is readily available than ever before. This increase in data has encouraged the use of data-driven animation algorithms based on neural networks, as well as opened the doors for more people to use motion capture data in their projects. Similarity metrics are used to find motions in large databases, to train neural networks and to evaluate new algorithms [Sedmidubsky et al. 2021]. The accuracy of a similarity metric is therefore crucial for the success of many different algorithms. However, popular similarity metrics are purely mathematical and do not take human perception into account. It is likely that data marked as similar by

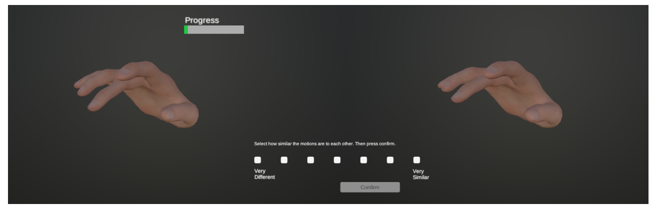

Figure 1: The web client participants interacted with to judge similarity between motions.

these metrics are not perceptually similar. Ideally, we could compute the similarity of any type of motion data as a human would. To investigate this, we designed an experiment to determine if and when the perception of hand motions correlates well with popular existing metrics. In this study we focus on hand motions because the space of possible changes is smaller for hands than for body motions and it has been shown that hand motions are important in our perception of motions [Jörg et al. 2010; Wang et al. 2016].

2 METHOD

Our experiment is designed to measure the discrepancy between common motion similarity metrics and human perception of similarity. In our experiment, participants viewed altered and unaltered animations, each about 10 seconds long, played side-by-side simultaneously in the web client shown in Figure 1. The alterations consist of error modifiers applied to the motion data, inspired by common technical issues that occur during motion capture.

2.1 Modifiers

We applied five modifiers to each animation: Original, Smooth, Offset, Jitter and Static. We used these modifiers to examine correlations between popular similarity metrics and human perception.

- **Original**: No changes are made to the animation. This modifier was used as an attention check.
- **Smooth**: The finger motions are averaged over frames. The length of the motion is not changed.
- **Offset**: The finger joints rotations are offset by a constant angle throughout the whole animation.
- **Jitter**: In each frame, a random rotation value is applied to each finger joint orientations.
- **Static**: The fingers do not move; the wrist moves normally.

The Smooth, Offset and Jitter modifiers have three levels of intensity, leading to a total of 11 conditions. The intensities were adjusted so that each modifier had the same Euclidean position distance as the others for each intensity level.

SIGGRAPH '22 Posters, August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9361-4/22/08.
https://doi.org/10.1145/3532719.3543233

2.2 Procedure

Participants were recruited using Amazon's Mechanical Turk. Participants were first shown examples of very similar and very dissimilar motions to establish the full range of animation differences. They would then, in random order, view all 11 conditions. The played animations were randomly selected from a pool of 11 motions. The altered animation appeared randomly on either the left or right side of the screen. Then, they rated similarity on a 7-point Likert scale from "Very Similar" (1) to "Very Dissimilar" (7). Animations could be replayed. The participants completed this process three times for a total of 33 comparisons.

3 RESULTS AND DISCUSSION

We evaluated the perceptual comparison results from 62 participants against Euclidean distance, calculated from various motion features and an angular velocity metric. We use Euclidean distance because it is a common choice when comparing motion data [Sedmidubsky et al. 2021].

We use Equation 1 to calculate the Euclidean distance d between a pose $\mathbf{q}$ from the altered motion and a pose $\mathbf{p}$ from the unaltered motion, where n is the total number of joints in the hand (20) and x, y, and z represent the coordinates of each joint i. We compute the distance for each frame and average over the total number of frames to find our final distance.

$$d(\mathbf{p}, \mathbf{q}) = \frac{1}{n} \sum_{i=1}^{n} \sqrt{(q_{xi} - p_{xi})^2 + (q_{yi} - p_{yi})^2 + (q_{zi} - p_{zi})^2} \quad (1)$$

Equation 2 compares the angular rotations of the current frame B to the previous frame A for each Euler rotation coordinates of each joint and gives the pose's total angular velocity.

$$d(\mathbf{p}, \mathbf{q}) = \frac{1}{n} \sum_{i=1}^{n} |(\vec{q}_{Bi} - \vec{q}_{Ai}) - (\vec{p}_{Bi} - \vec{p}_{Ai})| \quad (2)$$

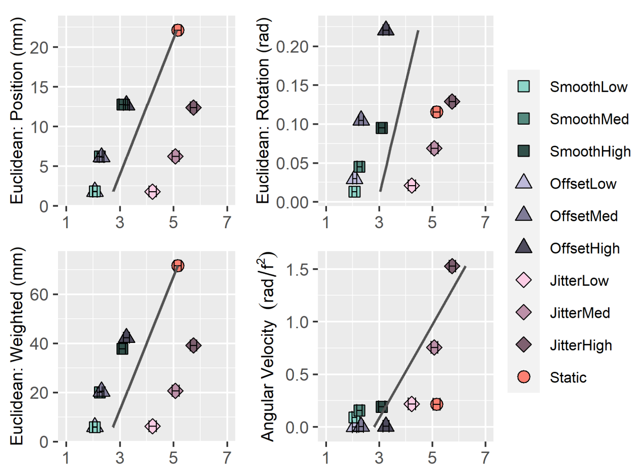

Figure 2: A comparison of similarity metrics against the mean Likert scale results for each condition. Error bars show the standard error of the mean.

3.1 Metric Analysis

We use linear regression to see how well the metrics correlate with human perception as seen in Figure 2. For each metric we obtain an R^2 value, where larger values indicate more agreement with human responses, and thus a more successful metric.

Euclidean distance: Position. $R^2 = 0.0644$, an approximate 6.4% correlation between the perceived differences and the computed metric.

Euclidean distance: Rotation. This metric compares the Euler angle coordinates. $R^2 = 0.0298$, an approximate 3.0% correlation.

Euclidean distance: Weighted. In this variant of the position metric we apply weights to the joints, giving more significance to some joints over others. The weights were 4, 3, 2 and 1 for each of the digits, starting at the tips and ending at the base, respectively. The root hand joint's weight was also one.

Angular Velocity. Finally, we tested an angular velocity metric that uses Equation 2. $R^2 = 0.1515$, an approximate 15.2% correlation.

3.2 Discussion and Limitations

Our study shows that typical similarity metrics do not accurately represent how humans perceive the differences between finger motions. While this conclusion is not surprising, the correlations are disappointingly low. Testing a variety of simple metrics provided no sufficient linear correlation between human perception and the mean metric scores for each condition This study supports recent related work with a similar result [Durupinar 2021]. Our experiment also gave us insight into which kinds of error modifiers create a perceptual discrepancy, i.e. the Offset, Smooth and Jitter conditions seem to correlate linearly, though Jitter runs along a different slope.

Although our study is limited to finger motions, it opens the stage for future work to evaluate more complex similarity metrics, apply additional motion modifiers, and compare full body motions. An accurate perceptual distance metric would be incredibly useful in evaluating the quality of new animation algorithms, allow for perceptually accurate motion database queries and enable a standardized baseline for comparisons. This study is the first step in establishing a pipeline to collect data on perceptual similarities. With enough data and insight on the perception of motion similarity it would be possible to devise such a distance metric.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1652210. We thank the attendees of the 2019 McGill University Bellairs workshop for ideas and discussions.

REFERENCES

Funda Durupinar. 2021. Perception of Human Motion Similarity Based on Laban Movement Analysis. In *ACM Symposium on Applied Perception 2021* (Virtual Event, France) *(SAP '21)*. ACM, New York, NY, USA, Article 8, 7 pages. https://doi.org/10.1145/3474451.3476241

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)*. ACM, New York, NY, USA, 129–133. https://doi.org/10.1145/1836248.1836273

Jan Sedmidubsky, Petr Elias, Petra Budikova, and Pavel Zezula. 2021. Content-Based Management of Human Motion Data: Survey and Challenges. *IEEE Access* 9 (2021), 64241–64255. https://doi.org/10.1109/ACCESS.2021.3075766

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.