The Correlation Between Distance Metrics and Perception for Hand Animations

Katherine Gemperline
Clemson University
kgemper@clemson.edu

Alex Adkins
Clemson University
adkins4@clemson.edu

Sophie Jörg
Clemson University
sjoerg@clemson.edu

ABSTRACT

When animating virtual characters, motions are typically used multiple times for games and movies, and developers modify these motions, for example, concatenating and blending multiple of them together. Distance metrics calculate the similarities or differences between pairs of animations. One commonly used distance metric is based on the Root Mean Square Error, using the rotations of the x, y, and z axes of each joint. However, while comparing two animations, viewers might perceive a large difference between the two animations while there is a low distance metric value and vice versa. In this paper, we are researching the correlation between distance metric values and viewer perception of animations to explore the strength of the correlation and the accuracy of distance metrics. We investigate how these motions are perceived to help develop more perceptually valid distance metrics in the future.

CCS CONCEPTS

• **Character Animation → Distance Metrics**; *Perceptual Similarity*;

KEYWORDS

distance metric, Root Mean Square Error, Character Animation, Perceptual Similarity.

1 INTRODUCTION

Distance Metrics are useful in many ways in the animation and game development industries. For instance, a distance metric can de- termine the best point to transition from one animation to another or can be used when searching for similar motions in databases. As Pejsa and Pandzic explain [3], motion retrieval from databases of motion is difficult as they contain vast quantities and are often unlabeled and unrecognizable. Jörg et al. [4] found that details in hand motions can be perceived and change our interpretation of a scene.

One of the most common distance metrics, based on the Root Mean Square Error, calculates the differences between two animations by comparing the rotations of the x, y, and z axes of the joints in the animation in each frame as follows

$$\sqrt{\frac{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}{n}}$$

where n is the number of degrees of freedom of the skeleton, which in our case is the number of joints times 3. Each joint has an x, y, and z rotation axis. A value is calculated for each frame using the above formula. The resulting distances are averaged together for a single value representing the difference between the two animations.

This specific distance metric averages all the joints, and therefore, if one joint is given a large offset, and the distance metric for the new and old animations are calculated, it would be similar to if all joints were given a slight offset, and the distance metric was taken of that animation with the original.

As shown in other research on distance metrics, such as Tang et al. [2], the main issue with the Root Mean Square Error, along with other distance metrics, is their lack of correlation with perceived differences. Jörg et al. [4] showed that finger motion is an important, even if unconscious, factor in understanding motion, but which changes in motion affect the perception of differences the most?

2 EXPERIMENT

We created an experiment to compare two motions from two angles and administered the experiment via Amazon Mechanical Turk to 120 participants. The experiment questionnaire was comprised of 4 sets of 36 questions; each participant only answered questions from one set. Each set was comprised of one base motion with four types of changes and four degrees of each change. Each question showed two camera angles of two videos: a base animation and the same animation after a change has been applied, see Figure 5. After the video, participants were asked to compare the two animations on a seven-point Likert Scale ranging from Extremely Similar to Extremely Dissimilar.

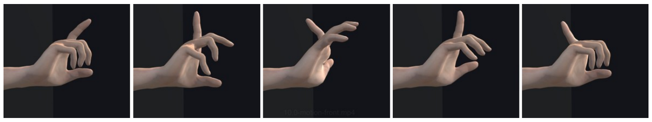

Figure 1: Base Motion Attention: (Left to Right) Original, Jitter, Added Motion, Offset Hand, Offset Index

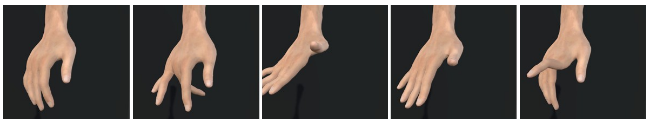

Figure 2: Base Motion Idle: (Left to Right) Original, Jitter, Added Motion, Offset Hand, Offset Index

Figure 3: Base Motion Shrug: (Left to Right) Original, Jitter, Added Motion, Offset Hand, Offset Index

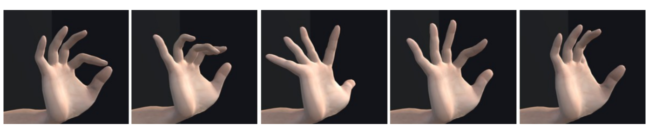

Figure 4: Base Motion Small: (Left to Right) Original, Jitter, Added Motion, Offset Hand, Offset Index

2.1 Stimuli

The four base motion animations are attention, idle, shrug, and small. They are segmented from the motions in the Large Gesture Database from Jörg et al. [1]. The animation segments are 68 frames long, played at 30 frames per second.

Attention, shrug, and small were all originally animated on the right hand of the model, while idle was the left hand's animation of the attention gesture. We mirrored the animation to the right hand so that all gestures were on the same hand for consistency.

As shown in Figures 1 - 3, there were four types of errors in addition to the original motion: offsetIndex, offsetHand, addedMotion, and jitter. Offset index added an offset to the y-axis of the metacarpophalangeal joint of the index joint. Offset hand added an offset to the y-axis of the metacarpophalangeal and proximal interphalangeal joints in each finger and the z-axis of the carpometacarpal and the metacar- pophalangeal joints in the thumb. Jitter added random movement to the fingers and thumb. Motion added an increasing offset to the y-axis of the fingers and z-axis of the thumb. The offset value is equal to the previous frame's offset value plus the increase value. The increase value was hard-coded and changed for each degree of change in regard to the distance metric values. The initial value for each motion was 1 to be consistent. The increase values are listed in Table 1.

Error Type	Distance Metric Value	Variable Value
		Standard Deviation
added Motion	2.5 5 7.5 10	.333 .729 .1.124 1.519
		offsetValue
jitter	2.5 5 7.5 10	4.5 9 13 18
		offsetValue
offsetIndex	2.5 5 7.5 10	20 40 60 80
		increase Value
offsetHand	2.5 5 7.5 10	4.48 8.95 13.43 17.9

Table 1: Error Type with the degree at which the original motion was changed and the value used to achieve that distance metric.

The animations were modified using Autodesk Maya and Python Scripts to create the modifications as explained above. The segments were exported to Unity where they were recorded from two different viewpoints. For all versions of the same base motion, all aspects of the recording stayed the same: lighting, model, model position and angles, and the same two camera viewpoints and positions.

2.2 Method

The questionnaire was created using Qualtrics and distributed through Amazon Mechanical Turk. Each question showed two animations from two different views for a total of four videos playing of each question, as shown in Figure 5. The videos were combined using Adobe Premiere Pro to make sure the videos played at the exact same time to compare the animations frame by frame.

The distance metric values for the animations used were 2.5, 5, 7.5, and 10. For offsetIndex, the distance metric values were exact, for offsetHand and addedMotion, the distance metric values were +/- .01 while for jitter, which had a random aspect to it, the previously stated values of the distance metrics deviated up to +/- .1.

The variables used for each error remained constant across all base motions, as shown in Table 1.

All animations with a change were compared to the original twice. Once with the original animation on the left and the changed animation on the right, and a second time with the original on the right and the changed animation on the left. For each base motion, there were four questions that asked the participant to compare the original animation to the original animation. This made a total of 144 questions *(4 base motions * 4 offset types * 4 offset distance metric values * 2 + 4 base motions (original compared to original) * 4 = 144 questions)*.

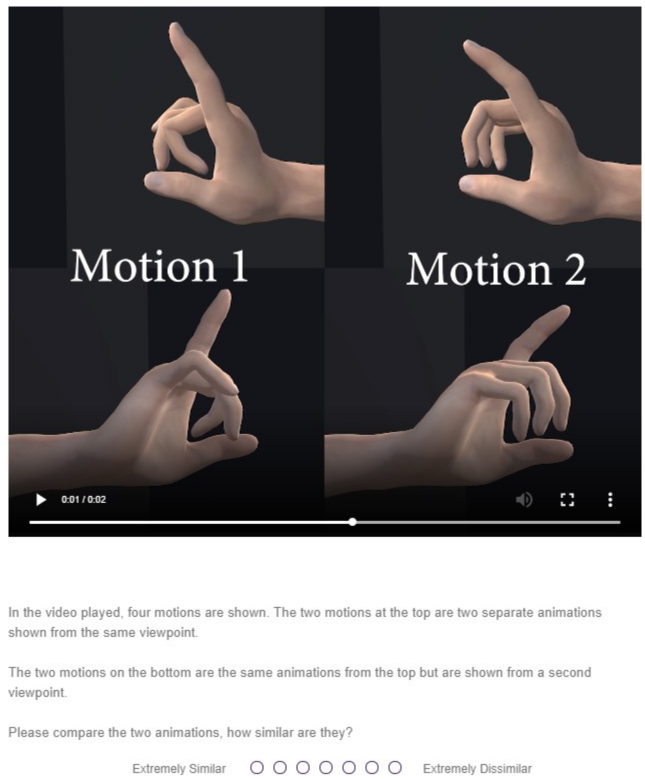

Figure 5: Comparison of two animations and the question asked after each comparison. The single video played two animations, each shown from two views, automatically and continuously. The viewer then ranked how similar the two animations were on a seven-point Likert Scale.

Each participant only answered questions relating to one base motion for a total of 36 questions per participant. We purposely kept the experiment short to ensure our participants would remain alert and active throughout the entire experiment. Each question asked the participant to rank the animations in a version of the seven-point Likert scale: from Extremely Similar to Extremely Dissimilar, as shown in Figure 5.

3 RESULTS

The experiment was administered to 120 participants, and it took an average of 60 minutes to complete. For some, it took 10 minutes while others completed the experiment in two hours. Each participant was given $0.80 in compensation for their time. 42 participants' responses were discarded as they did not pass our quality criteria. For each participant, there were four questions that asked for comparison of the base motion to the base motion, sorted randomly throughout the survey, with one comparison always at the end of the survey. The participants' responses were discarded if the four answers had a wide range on the 1-7 Likert scale, or if the answers were above 4. This procedure left a total of 78 responses to the survey; 18 participants compared animations with the base motion Attention; 22 participants compared animations with the base motion Idle; 20 participants compared animations with the base motion Small; 17 participants compared animations with the base motion Shrug.

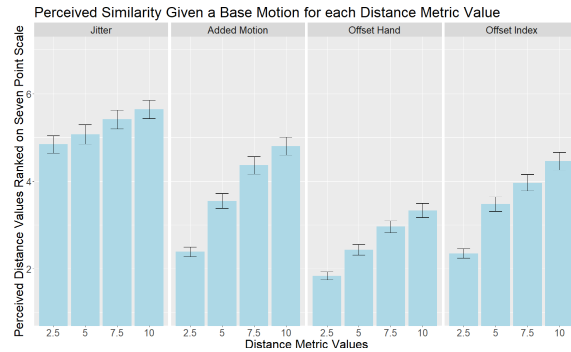

Figure 6: Graphs showing the mean responses given for each Distance Metric separated by Base Motion

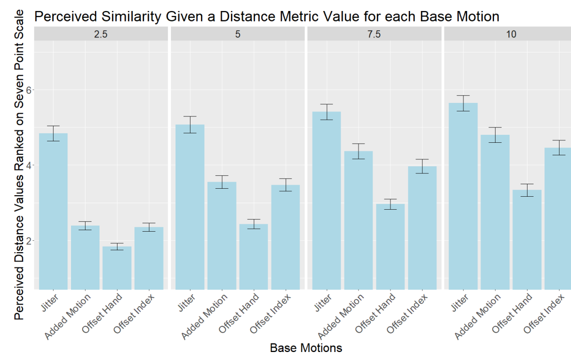

Figure 7: Graphs showing the mean responses given for each Base Motion separated by Distance Metric

Based on these results, our hypothesis was supported showing that the distance metric based on the Root Mean Square Error is not accurate to the perception of the viewer. While comparing the same type of error added to all motions but in different degrees, as shown in Figure 6, there are definite differences perceived by the viewer. As we expected, larger distance metric values lead to larger perceived distances within each error type, which validates our precedent. However, as shown in Figure 7, across different base motions and error types, the same distance metric value is not perceived as having the same level of similarity to the original base motion.

These results show that while using distance metrics to search for images in a limited range of motions might work to a better degree, especially for large databases of motions, which are typically used, this distance metric is not accurate in finding similar motions.

4 Future Work

Continuing this line of investigation, future works could include a wider range of motions, longer motions, or motions that take into account more than just the hands, such as looking at the entire body. One future experiment could be determining which angles or the number of angles that need to be affected or to what degree do they need to be affect for a viewer to perceive a motion as not being similar.

ACKNOWLEDGMENTS

The material is based upon work supported by the National Science Foundation under Grant Number IIS-1652210.

REFERENCES

[1] Sophie Jörg, Jessica Hodgins, and Alla Safonova Data-driven Finger Motion Synthesis for Gesturing Characters *ACM Transactions on Graphics (SIGGRAPH Asia)*, November 2012, volume 31, issue 6.

[2] Jeff K. T. Tang, Howard Leung, Taku Komura, and Hubert P. H. Shum Emulating human perception of motion similarity. *Wiley InterScience*, August 2008, online.

[3] Pejsa, T., and I.S. Pandzic State of the Art in Example-Based Motion Synthesis for Virtual Characters in Interactive Applications. *Computer Graphics Forum*, February 2010, volume 29, issue 1.

[4] Sophie Jörg, Jessica Hodgins, and Carol O'Sullivan The Perception of Finger Motions. *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, 23 July 2010