

# Categorizing Rhythmic Jumping Motion Using Motion Capture without Markers

Marina Mikami<sup>1</sup>, Noriyuki Kida<sup>2\*</sup>

<sup>1</sup>Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto, Japan

<sup>2</sup>Faculty of Arts and Sciences, Kyoto Institute of Technology, Kyoto, Japan

Email: \*kida@kit.ac.jp

**How to cite this paper:** Mikami, M., & Kida, N. (2023). Categorizing Rhythmic Jumping Motion Using Motion Capture without Markers. *Advances in Physical Education*, 13, 93-105.

<https://doi.org/10.4236/ape.2023.132009>

**Received:** November 15, 2022

**Accepted:** April 7, 2023

**Published:** April 10, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this study, we sought to reveal and evaluate the degree of similarity between individuals and within the same group of individuals when they perform the same movement. The methodology used was that of selecting a panel of participants, 38 normal healthy adults ranging in age from 10 to 30 years. Subsequently, we carried out the experiment by performing the movements while capturing videos using a smartphone. The two-dimensional coordinates of each joint were obtained for videos obtained using marker less motion capture software. The results obtained allowed for a cluster analysis using five time series variables: 1) vertical head movement and 2) lateral limb movement. These time series were normalized by a total time, in order to create a typology of jumping movements. 136 series of movies were obtained. After checking the movies, 115 series revealed errors such as misrecognition. These series were excluded from the analysis. The remaining 21 series had data available for analysis. Cluster analysis was performed on the 21 sets of motion data, which could be classified into seven clusters based on the shape of the dendrogram.

## Keywords

Automation, Remote, Cluster Analysis

## 1. Introduction

The Coronavirus Disease-2019 (Covid-19) pandemic has had a profound impact on human life and habits. Social distancing and lockdown measures implemented to control the spread of the infection have resulted in changes in the realization and supply of various services. General retail, catering, and restaurants were shut

down, and public gatherings were prohibited. Universities suspended classes on campus (Favale et al., 2020; Horita et al., 2021). However, remote learning has several problems, including assessing movements and motor learning proficiency in sports (Jeong et al., 2020). Studies have reported methods of obtaining body coordinates for sports teaching, especially dance, including motion capture using markers and sensors (Shimizu et al., 2005; Chan et al., 2010). However, these motion capture methods require a certain level of knowledge in handling markers and sensors. In addition, attaching markers and sensors to the body limits the subject's movements, making it difficult to move as usual. To solve these problems, Corazza et al. (2006) proposed a motion capture method that does not use markers that can obtain body coordinates from movies. Research on motion capture without using markers has increased in recent years. For example, Cao et al. (2019) presented OpenPose, which can reveal the whole body posture and the position of body parts using Convolutional Neural Network (CNN) techniques from 2D RGB images. Moreover, in the case of dance, a system to classify yoga poses was developed using OpenPose (Yadav et al., 2019). Furthermore, Kim (2017) developed a system to generate skeletal motion data of ballet dancers using multiple RGB and depth sensors. Also, Laggis et al. (2017) reported that a remote folk dance instruction system has been developed using Kinect sensors.

In contrast, few studies have reported remotely assessing dance learning using markerless motion capture. Labuguen et al. (2020) assessed dance proficiency but used multiple devices. Individual identification with gait data using Kinect has been reported (Mori & Kikuchi, 2017; Mori & Kikuchi, 2018). Okugawa et al. (2019) reported that clustering using Openpose and Kmeans methods could evaluate arm movement synchronization in parade dances. It might be possible to apply these methods to assess movement similarities with easily captured 2D images.

This study aimed to automate movement evaluation in remote physical exercise learning. Therefore, we first attempted to evaluate inter-individual similarities of people performing identical movements using videos captured by smartphones. We used rhythmic jumping as the task, because we expected people in remote environments to easily record rhythmic jumping using smartphones. Rhythmic jumping, in which the arms and legs are moved in time with jumps, is used in sports and rehabilitation, and it requires hand and foot coordination to master. Physical education and sports classes also use rhythmic jumping. Studies have reported that rhythmic jumping increased tennis proficiency (Söğüt et al., 2012) and older adult's physical and mental functions (Sugiura et al., 2010).

We obtained body coordinates from movies filmed using the participant's smartphones. We examined the possibility of automatically evaluating inter- and intra-individual similarities based on the coordinate data of five body parts obtained from the videos. We conducted this study during the Covid-19 pandemic when behaviors were restricted.

## 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Equipment

The equipment used for this study consisted of smartphones (iPhone, Apple Inc.), cameras (Ex-100, Casio), software (Pose-Cap, Four Assist, Inc.).

#### 2.1.2. Participants

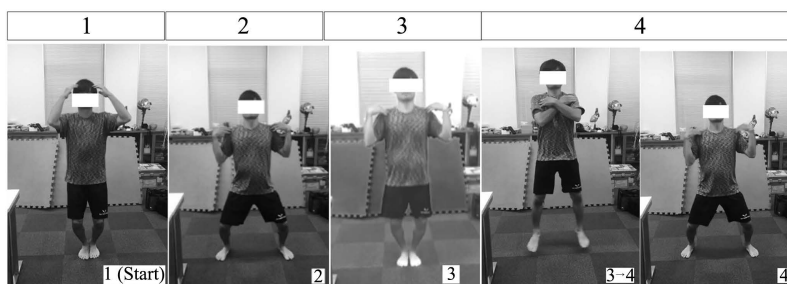
Healthy normal adults in their teens to their thirties ( $N = 38$ ) without any professional dance experience participated in this experiment. We explained the purpose and methods of the study to the participants and obtained their consent to participate.

### 2.2. Methods

#### 2.2.1. Jumping Motion and Data Collection

We used a rhythmic jump tempo of 100 bpm, with one series of four beats as the basic unit. The participants' lower limb movements included leg opening and closing movements during the jumps, with the first beat with the legs closed, the second with the legs open, the third with the legs closed, and the fourth with the legs open. The upper limb movements consisted of placing both hands on the head for the first beat, placing both hands on the ipsilateral shoulder for the second beat, maintaining the same position as the second beat for the third beat, and moving both hands from the ipsilateral shoulder to the opposite shoulder on the fourth beat, and quickly returning them to the same shoulder (**Figure 1**).

We made a movie to explain how we wanted the participants to jump. The participants watched the movie and practiced jumping before making the movie for motion analysis. They recorded the movies for motion analysis in their homes using their smartphones' video cameras. We instructed the participants to use their smartphones or video cameras and record at least four rhythmic jump series of four beats each, without directly instructing them on their movements, but showed them the movie we made as a model. Filming the movies, submitting them, and viewing them were done remotely.



**Figure 1.** Rhythmic jumping. The lower limb movements were legs opened and closed jumps. The upper limb movements were: 1) placing both hands on the head for the first beat, 2) placing both hands on the ipsilateral shoulder for the second beat, 3) maintaining the same posture as in the second beat for the third beat, and 4) moving both hands from the ipsilateral shoulder to the opposite shoulder on the fourth beat and quickly returning them to the same shoulder.

### **2.2.2. Processing Movie Data**

We collected the participants' movies and analyzed them using one series of four beats as the basic unit. We set the start of each series as the first beat of the movement with the feet on the ground in a closed position and the hands above the head. We set the end of each cycle as the end of the fourth beat of the movement with the feet in contact with the ground in a closed position and the hands above the shoulders. One series consisted of four beats, which included four jumps, two open and closed leg movements, and one hand crossing.

We obtained two-dimensional coordinates of the joints from the movie data using markerless motion capture software (Pose-Cap, Four Assist, Inc.). The five coordinates, head, left wrist, right wrist, left ankle, and right ankle, were used in the analysis after excluding trials with incomplete data. Then, we developed a typology of jumping movements using cluster analyses of the five coordinates' time-series variables: vertical movement of the head and lateral movement of the limbs normalized by the total time.

## **3. Results**

### **3.1. Motion Data Acquisition**

We obtained 136 series of movies; two participants provided one series, four provided two series, five provided three series, twenty-four provided four series, and three provided five series. The movies showed that 44 series were filmed inadequately; the lower body was not shown in 3 series, and the head and below the knees were sometimes out of range in 41 series. We selected 92 series in which nearly the entire body was filmed to capture the motion without using markers.

Fifteen joint coordinate series that we acquired for motion capture had unrecognizable body parts in over five frames because the subject was outside the camera's range; the head was out of the camera's range during the jumping motion in one series, and the lower body below the ankle was out of the camera's range during the leg opening motion in another series.

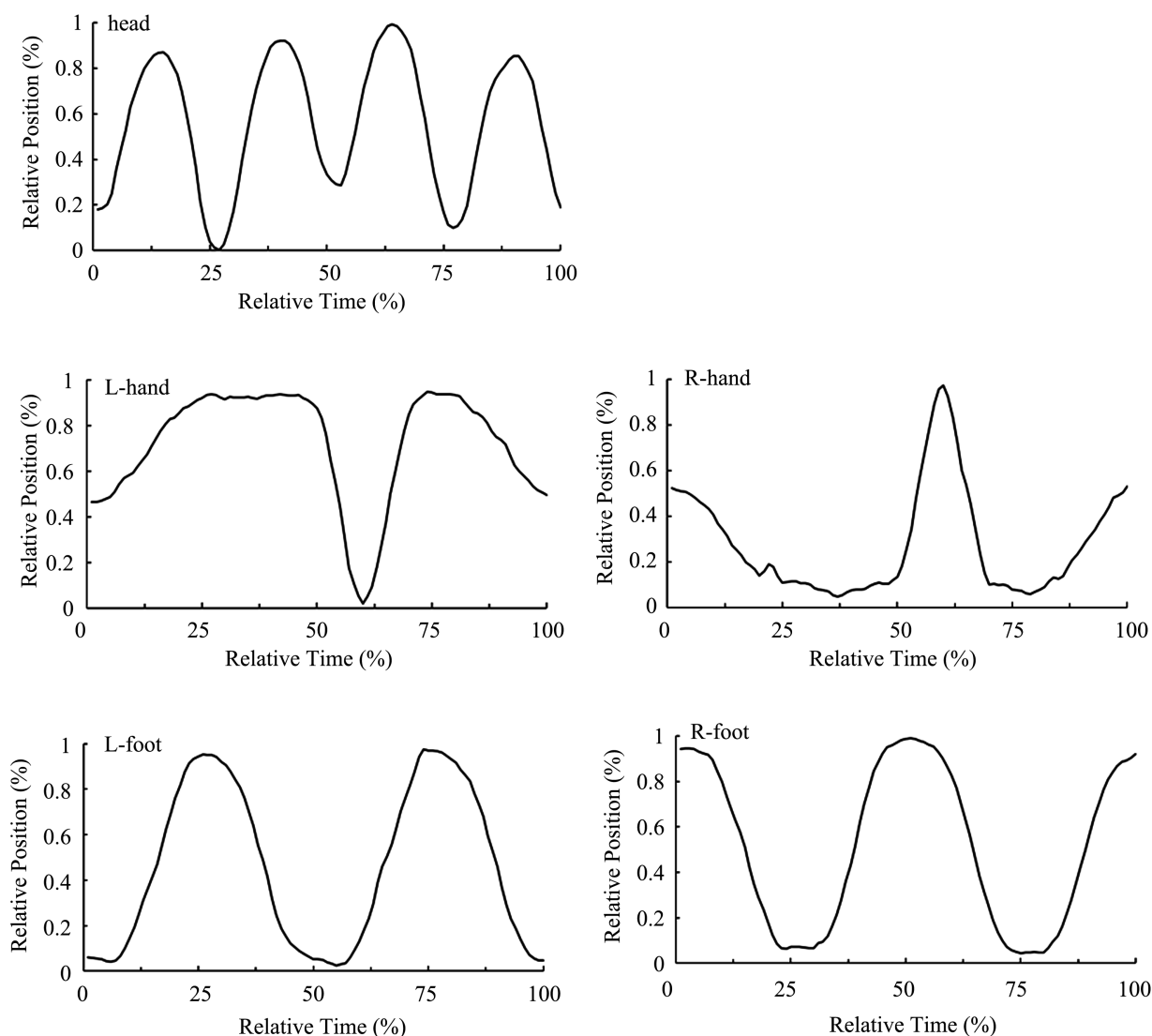
Next, we examined the misrecognition of joint coordinates. The movements and numerical values of the coordinate data differed in 30 series. In 16 of these series, the filming method was problematic. Specifically, the subject was not filmed from the front or was filmed from too far, resulting in inadequate image quality, which made it challenging to obtain the coordinates. In nine series, the area from the ankle to the bottom of the foot was out of the camera's range when the foot opened. The camera did not acquire the hand coordinates in five series; the hand-crossing motion was blurred because it was too fast, resulting in inaccuracies.

In 16 movies, the coordinates were captured; but they were different from the actual movement. The hand motion data of ten series were contaminated by noise, five were contaminated by hand and head movement noise, and one was contaminated by foot movement noise. Nine of these ten contaminated series

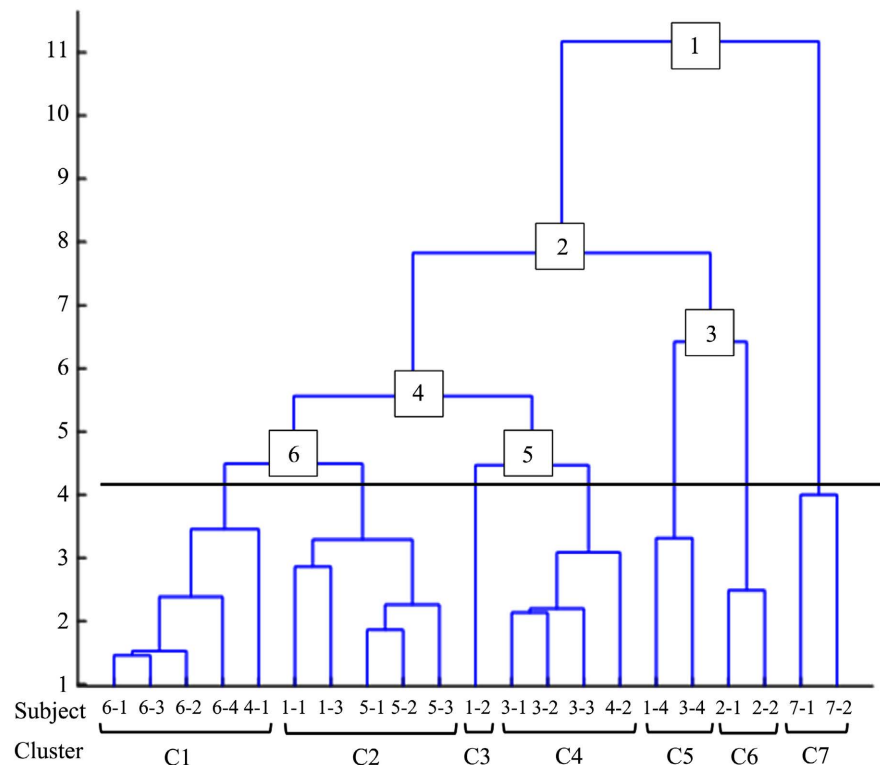
were of hand movements, suggesting that the speed of the hand movements prevented recording the coordinates accurately. Finally, there were subtle misidentifications in ten series. We excluded 56 series from the analysis due to errors such as misidentifications and analyzed 21 series of data comprising 7 movies: 3 movies with 4 series, 1 with 3 series, and 3 with 2 series. **Figure 2** shows the average values of the 21 series.

### 3.2. Classification by Cluster Analysis

We performed cluster analysis on 21 series of motion data and classified them into seven clusters based on the dendrogram's shape (**Figure 3**). We named these Clusters 1 and 2 from left to right, as shown in **Figure 2**. We examined each cluster's motion characteristics based on motion differences of each step, which we used to divide the clusters based on the dendrogram.



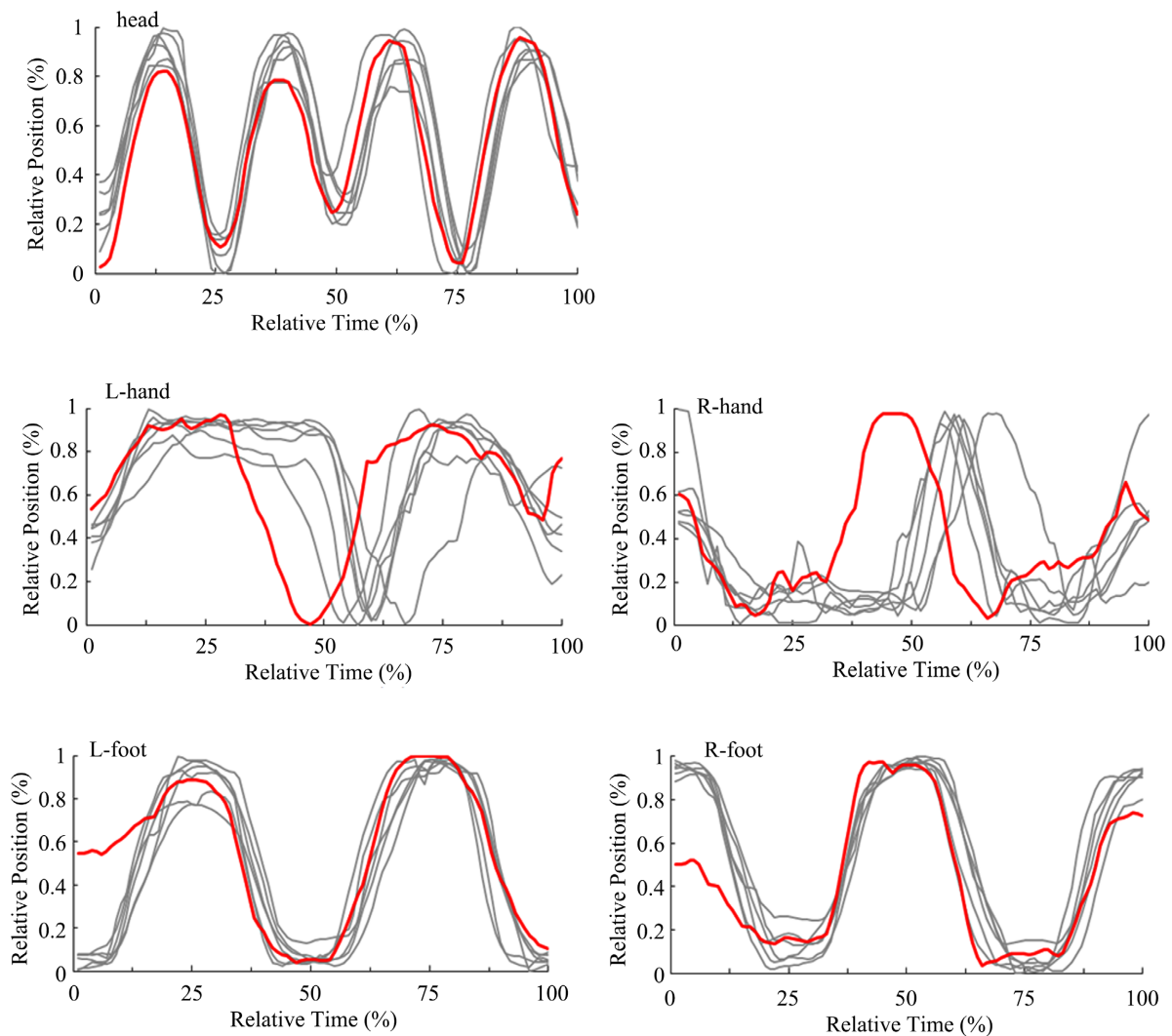
**Figure 2.** Typical sample. The average values of the 21 series. The horizontal axis shows the relative time, and the vertical axis the relative position.



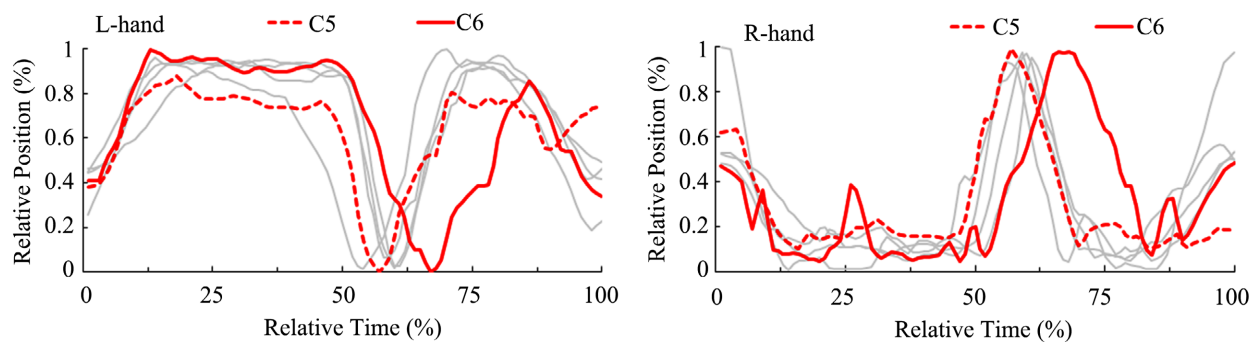
**Figure 3.** Cluster analysis' dendrogram. Cluster analysis was performed on the 21 series of motion data, which were classified into seven clusters based on the dendrogram's shape.

In the first step, we divided the clusters into two sets: Clusters 1 - 6 and Cluster 7. **Figure 4** shows the mean values of each cluster. The horizontal axis shows the relative time, and the vertical axis the relative position. The mean value of Cluster 7 is shown in red, whereas Clusters 1 - 6 are shown in gray. The left and right-hand motions of Cluster 7 were markedly different, with the cross-motions of both hands starting approximately at 45% relative time. In contrast, the other clusters started their cross-motions at approximately 60% relative time. Only Cluster 7 started the cross-motions earlier. The correct move was to touch the shoulders on the same side from the second to the third beat. However, 7 showed the cross-motion of both hands after the second beat making this an independent cluster. The starting position of the feet in Cluster 7 was also different from the other clusters. The first beat was a vertical movement with both feet together, the right foot positioned farthest to the left, and the left foot positioned farthest to the right. However, the feet were slightly open In Cluster 7, which included only one participant who displayed many movement errors.

In the second step, we divided the 19 series into 15 and 4 series, and these 4 series were named Clusters 5 and 6, with 2 series in each cluster, characterized by differences in hand movements. The average of both hand motions in Clusters 1 - 4 is shown in gray, those of Cluster 5 are shown in a red dotted line, and of Cluster 6 in a red dashed line (**Figure 5**). We observed differences between



**Figure 4.** Cluster analysis, Step-1. The red line shows the mean value of Cluster 7, and the gray line those of Clusters 1 - 6.



**Figure 5.** Cluster analysis, Step-2. The gray line shows the averages of Clusters 1 - 4 for both hand motions; the red dotted line those of Cluster 5, and the red line those of Cluster 6.

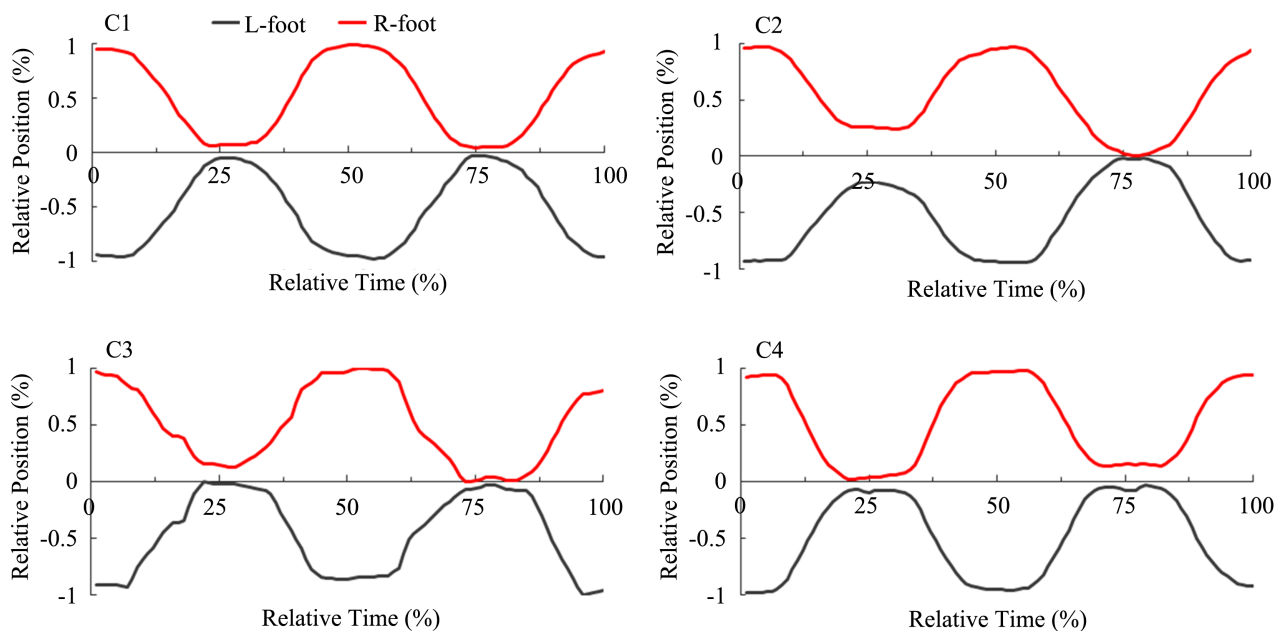
Clusters 5 and 6 in 70% - 85% of the relative time for the left hand. The left hand in Clusters 1 - 4 was 0.9, whereas Cluster 5 was 0.7. Cluster 6 varied from 0 to approximately 0.8. Subjects in Clusters 1 - 4 maintained contact with the ipsila-



teral shoulder and showed the correct motion. Subjects in Cluster 5 did not show the left hand-to-the-shoulder motion. Subjects in Cluster 6 were slow to initiate the hand-crossing motion. For the left hand, approximately 100%, all but Cluster 5 were below 0.5, whereas Cluster 5 was 0.8. This pattern was also observed for the right-hand motion, which was above 0.5, except for the subjects in Cluster 5, which was 0.2. Since the last motion in a series was nearly identical to the first, we expected their values to be identical at 1% and 100%. However, only Cluster 5 was different, which suggested the possibility of movement errors.

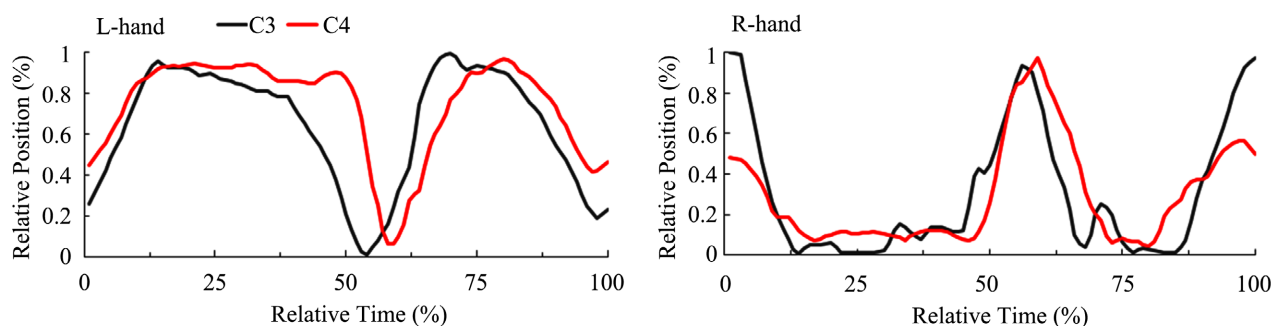
The third step was divided into Clusters 1 and 2 and 3 and 4. We observed differences between left and right foot motion variability in these clusters (**Figure 6**). We transformed the left foot coordinates to a range of  $-1$  to  $1$  to compare the left and right foot motions. When the feet were closed, the left foot was  $-1$ , and the right foot was closer to  $1$ , whereas the right foot was closer to  $0$  when the feet were open. There was no difference in the motion between the left and right legs' open and closed conditions in Clusters 1 and 2, whereas there was a left-right difference in the width of the open leg in the first half of Cluster 3 and the second half of Cluster 4. The legs always move symmetrically. Therefore, when the left leg was  $-1$ , the right leg was  $1$ , and when the left leg was  $0$ , the right leg was also  $0$ . However, in Cluster 3, the left leg was  $0$ , approximately at the 25th percentile, whereas the right leg was  $-0.1$ , indicating that the legs did not spread symmetrically. In Cluster 4, the left leg was  $0$ , approximately at the 75th percentile, whereas the right leg was  $-1$ .

The fourth step divided the group into Clusters 3 and 4. The foot and hand motions of this cluster were different (**Figure 7**). The left and right hands in



**Figure 6.** Cluster analysis, Step-3. Left foot coordinates were transformed into a range of  $-1$  to  $1$  to compare left and right foot motions. When the foot was closed, the left foot was  $-1$ , and the right foot was closer to  $1$ ; when the foot was open, the right foot was closer to  $0$ .





**Figure 7.** Cluster analysis, Step-4. The red line represents the average of Cluster 4, and the black line that of Cluster 3.

Cluster 3 had different starting and ending positions. Cluster 3 consisted of only one series with one participant, which was identified for its unique hand movement. The left hand moved from 40% relative time in Cluster 3, but the right hand moved from 45% relative time and the two hands were unsynchronized. The left hand moved from 50% relative time in Cluster 4, but the right hand moved from 45% relative time and was unsynchronized on either side. Clusters 3 and 4 were detected as having incorrect movements because correct hand movements were from 50% relative time. Participants in Cluster 3 had 0.2 for the left hand at 1% and 100%, but 1 for the right hand. Although their hands were above the head at the 1% and 100% time points, the participants did not align the left and right hands, and the right hand was in the incorrect position.

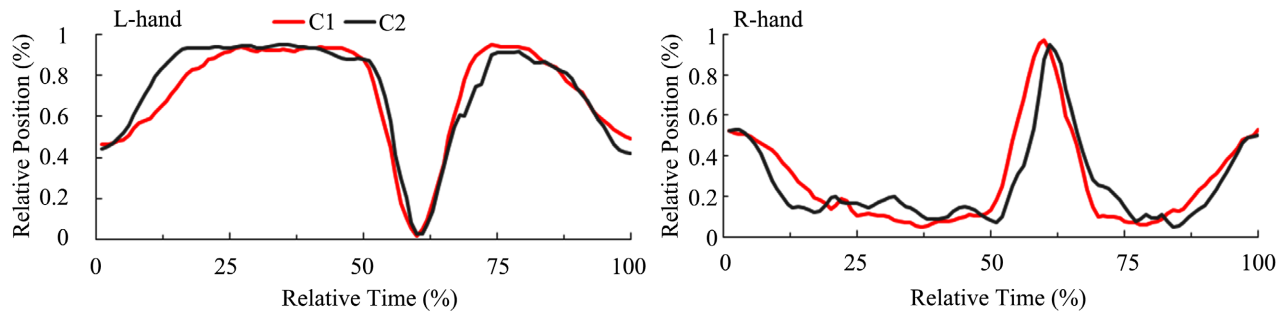
We divided them into Clusters 1 and 2 at the fifth step (**Figure 8**), which showed no hand motion differences between the two groups. In contrast, Cluster 2 had a different foot motion in the first and second half of the step, whereas the foot motion of Cluster 1 was stable. In Cluster 1, both left and right feet were 0 at 25% and 75% relative times, whereas, in Cluster 2, the left foot was  $-0.2$  at 25% relative time, and the right foot was  $0.2$  at 75% relative time. Moreover, both feet were 0 at 75% relative time. We found no difference in the foot opening width between the second and fourth beats, with a narrower foot opening for the second beat and a wider foot opening for the fourth beat.

### 3.3. Relationship between Participants and Clusters

We divided the 21 motions series of the seven participants into seven clusters. **Table 1** shows the numbers of the series and clusters for each participant. Four participants were classified into just one cluster. Two of them were in clusters with only one participant. We classified three into multiple clusters, which had one participant in three clusters and two participants in two clusters.

## 4. Discussion

The participants of this study captured their movements without using markers in a remote environment with their recording devices and uploaded the data. Then, we classified the data using cluster analysis. We obtained 136 series; however, we could not capture the motion of 44 movie series due to recording problems.



**Figure 8.** Cluster analysis, Step-5. The average of Cluster 1 is shown in red, and Cluster 2 is in black.

**Table 1.** Series number and clusters for the participants.

Subject ID	Cluster No.						
	1	2	3	4	5	6	7
1	0	2	1	0	1	0	0
2	0	0	0	0	0	2	0
3	0	0	0	3	1	0	0
4	1	0	0	1	0	0	0
5	0	3	0	0	0	0	0
6	4	0	0	0	0	0	0
7	0	0	0	0	0	0	2

The movement task in this study was a difficult jump that participants performed for the first time. Therefore, we explained the movement to the participants using a movie. However, our explanation might have been insufficient. We suggest explaining the points that participants must consider when recording the motion, in addition to showing a movie with a sample of the desired motion in future studies to ensure usable movies.

Moreover, we could not obtain the participants' coordinates using motion capture in 16 series, presumably due to the speed of the hand-crossing motion. Fast movements become blurred at normal frame rate and shutter speeds. This problem can be solved by filming at high speeds or increasing the shutter speed. However, the participants in this study used their video recording devices, and we did not standardize the equipment's performance. Nevertheless, over half the participants could capture the target movement. We expect that smartphones with better performance in the future will enable acquiring more stable videos.

This study did not solve specific issues related to the video recording environment. However, this study indicated the possibility of remote, automatic motion evaluation by capturing movements in distant places. Researchers have pointed out that the accuracy of acquiring body coordinates in motion capture without markers varies depending on the shooting angle (Takada et al., 2012)

and the distance from the image (Tsai et al., 2020). We expect further research to facilitate remote, automatic motion evaluation of complex movements, including dance. The results of this research could be applied to other areas such as sports skill evaluation and posture evaluation in medical checkups.

This study assessed the similarity degree between and within individuals. We categorized participants 2, 5, 6, and 7 into one cluster and Participants 1, 3, and 4 into multiple clusters. Four participants in one cluster showed little intra-individual variability. Participants 2 and 7 were included in the only clusters in which participants could be identified by markerless motion analysis. The three participants that we classified in multiple clusters had high motion variability. We classified Participant 1 into two Cluster 2 series and one Cluster 3 and 5 series. Also, we classified Participant 3 into three Cluster 4 series and one Cluster 5 series. Cluster 5 differed from the others due to the starting and ending positions of the hand motions and because it included a unique motion. It is concluded that we identified typical motion characteristics among different participants based on these categorizations.

Cluster analysis also facilitates evaluating the mastery of tasks. The task in this experiment was rhythmic jumping, in which a given action was repeated multiple times. Japanese Bon dancing (Bon Odori) also repeats the same action to a specific rhythm. Motion capture without markers might help evaluate intra-individual similarities of people Bon dancing. Such studies could quantify dance movements' phylogeny in different regions of Japan through cluster analysis and facilitate evaluating regional similarities in traditional Bon dances handed down from one region to another. Video images, including past videos, could be used to evaluate traditional dance movements in remote environments, resulting in identifying the succession of traditional culture for preserving intangible cultural assets.

## 5. Conclusion

In this study, we sought to reveal and evaluate the degree of similarity between individuals and within the same group of individuals when they perform the same movement. We carried out the experiment by performing the movements while capturing videos using a smartphone. The two-dimensional coordinates of each joint were obtained for videos obtained using marker-less motion capture software. Cluster analysis was performed on the 21 sets of motion data, which could be classified into seven clusters based on the shape of the dendrogram. We were able to evaluate the similarity of movements and divide them into categories using a simple method. The findings are expected to be applied to other fields, such as sports skill evaluation and posture evaluation for health checkups.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 172-186. <https://doi.org/10.1109/TPAMI.2019.2929257>
- Chan, J. C., Leung, H., Tang, J. K., & Komura, T. (2010). A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Transactions on Learning Technologies*, 4, 187-195. <https://doi.org/10.1109/TLT.2010.27>
- Corazza, S., Muendermann, L., Chaudhari, A. M., Demattio, T., Cobelli, C., & Andriacchi, T. P. (2006). A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach. *Annals of Biomedical Engineering*, 34, 1019-1029. <https://doi.org/10.1007/s10439-006-9122-8>
- Favale, T., Soro, F., Trevisan, M., Drago, I., & Mellia, M. (2020). Campus Traffic and e-Learning during Covid-19 Pandemic. *Computer Networks*, 176, Article ID: 107290. <https://doi.org/10.1016/j.comnet.2020.107290>
- Horita, R., Nishio, A., & Yamamoto, M. (2021). The Effect of Remote Learning on the Mental Health of First Year University Students in Japan. *Psychiatry Research*, 295, Article ID: 113561. <https://doi.org/10.1016/j.psychres.2020.113561>
- Jeong, H. C., & So, W. Y. (2020). Difficulties of Online Physical Education Classes in Middle and High School and an Efficient Operation Plan to Address Them. *International Journal of Environmental Research and Public Health*, 17, 7279. <https://doi.org/10.3390/ijerph17197279>
- Kim, Y. (2017). Dance Motion Capture and Composition Using Multiple RGB and Depth Sensors. *International Journal of Distributed Sensor Networks*, 13, Article ID: 1550147717696083. <https://doi.org/10.1177/1550147717696083>
- Labuguen, R. T., Negrete, S. B., Kogami, T., Ingco, W. E. M., & Shibata, T. (2020). Performance Evaluation of Markerless 3D Skeleton Pose Estimates with Pop Dance Motion Sequence. In *2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)* (pp. 1-7). IEEE. <https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306581>
- Laggis, A., Doulamis, N., Protopapadakis, E., & Georgopoulos, A. (2017). A Low-Cost Markerless Tracking System for Trajectory Interpretation. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W3, 413-418. <https://doi.org/10.5194/isprs-archives-XLII-2-W3-413-2017>
- Mori, T., & Kikuchi, H. (2017). Proposal of a Personal Identification and Tracking Method Using Depth Sensor Gait Features. In *Computer Security Symposium 2017* (pp. 972-979). Information Processing Society of Japan. (In Japanese)
- Mori, T., & Kikuchi, H. (2018). Person Identification Method Based on DTW Distance of Gait Sequence and Its Evaluation of Robustness against Obstacles. In *Multimedia, Distributed, Cooperative, and Mobile Symposium 2018* (pp. 672-680). Information Processing Society of Japan.
- Okugawa, Y., Kubo, M., Sato, H., & Viet, B. D. (2019). Evaluation for the Synchronization of the Parade with OpenPose. *Journal of Robotics, Networking and Artificial Life*, 6, 162-166. <https://doi.org/10.2991/jrnal.k.191203.001>
- Shimizu, K., Woong, C., & Hachimura, K. (2005). Research on Interaction System of Tele-Dance with Motion Capture and Network. *Jinmonkom*, 2005, 173-178. (In Japanese)

- Söğüt, M., Kirazci, S., & Korkusuz, F. (2012). The Effects of Rhythm Training on Tennis Performance. *Journal of Human Kinetics*, 33, 123-132.
- Sugiura, Y., Sakurai, H., Wada, H., Sakakura, T., & Kanda, Y. (2010). Effect on Physical and Mental Function of a Group Rhythm Exercise for Elderly Persons Certified under the Less Severe Grades of Long-Term Care Insurance. *Rigakuryoho Kagaku*, 25, 257-264. (In Japanese)
- Takada, K., Kitasuga, T., & Aritsugi, M. (2012). A Consideration of an Individual Identification Method by Gait Using a Markerless Motion Capture Device. *Entertainment Computing: Technologies and Applications*, 2012, 1-7. (In Japanese)
- Tsai, Y. S., Hsu, L. H., Hsieh, Y. Z., & Lin, S. S. (2020). The Real-Time Depth Estimation for an Occluded Person Based on a Single Image and OpenPose Method. *Mathematics*, 8, 1333. <https://doi.org/10.3390/math8081333>
- Yadav, S. K., Singh, A., Gupta, A., & Raheja, J. L. (2019). Real-Time Yoga Recognition Using Deep Learning. *Neural Computing and Applications*, 31, 9349-9361. <https://doi.org/10.1007/s00521-019-04232-7>