Master's Thesis

Toward The Perfect Stroke: A Multimodal Approach for Table Tennis Stroke Evaluation

Panyawut Sri-iesaranusorn
Program of Information Science and Engineering
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Kazushi Ikeda
Mathematical Informatics Lab. (Division of Information Science)

Submitted on July 30, 2021

A Master's Thesis
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Panyawut Sri-iesaranusorn

Thesis Committee:

Supervisor	Kazushi Ikeda
	(Professor, Division of Information Science)
	Keiichi Yasumoto
	(Professor, Division of Information Science)
	Junichiro Yoshimoto
	(Associate Professor, Division of Information Science)
	Makoto Fukushima
	(Assistant Professor, Division of Information Science)
	Chie Hieida
	(Assistant Professor, Division of Information Science)

Toward The Perfect Stroke: A Multimodal Approach for Table Tennis Stroke Evaluation*

Panyawut Sri-iesaranusorn

Abstract

Developing a consistent stroke is a challenge and even more so for non-professional Table tennis players. To build consistent proper strokes for beginner players, there is a need to understand the stroke differences between standard and beginner players. In terms of table tennis applications, prior works used a video-based method, or accelerometer sensor embedded in a table tennis racket, or infrared (IR) depth sensor for evaluating the stroke. However, there are certain challenges in these methods such as having insufficient data to analyse a complete stroke, time-consuming and costly data collection, as well as using non-prevalent equipment. Hence, to improve the beginner player's performance, an ubiquitous way through readily accessible commercial devices for stroke evaluation is essential. In this study, to achieve such a goal, we (i) recorded videos and accelerator signals of standard and beginner players using consumer-grade products, and (ii) analysed the stroke consistency of both standard and beginner players. The results show the significant differences in the strokes between both kinds of players through the multimodal approach. Also, we found the significantly strong correlation between the stroke consistency and the hitting score for the forehand stroke. These findings motivate us to further examine the improvement of beginner players by instructing procedural knowledge of a standard player's stroke, and implement applications for the motor-skill instruction.

Keywords:

Table tennis, Stroke detection, Motion Capture, Joint Kinematics, Sports Analytics

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, July 30, 2021.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

Table tennis is one of the fastest ball games in the world, requiring precise reaction time and strokes to achieve different kinds of spins and tactics [1, 2]. The nature of this game requires players to consistently build good habits in their forms as they develop their muscle memories for their strokes. Given this nature, proper form building and stroke mechanics are vital not only for the competitive aspects of the game but also for its enjoyment – especially for casual players.

Building consistent stroke is a challenge, especially for beginner players. Although professional coaching is usually needed in order to assess stroke technique and quality, it requires time and monetary investment [3]. Hence, developing an ubiquitous way of evaluating table tennis stroke through readily accessible commercial devices would help beginner players to improve their performance, requiring a small amount of their efforts.

Recent advancements in mobile computing technology have given birth to a plethora of wearable devices that have become popular especially for physically active groups of people that aim to track their fitness through running [4]. Given the success of fitness trackers for daily physical activities, the applications of such devices specifically targeted for a particular sport are still in their infancy. For example, several researches and applications have been done for tennis such as sensor-based swing and stroke classification as a tool for personal feedback and coaching [5–7]. Another similar research has also been done with regards to the baseball pitching [8], badminton stroke classification [9, 10], and the golf swing as well [11].

In terms of table tennis applications, prior studies used video-based methods and ball tracking as a way to give feedback to players [12]. A recent study used an accelerometer sensor embedded in a table tennis racket to classify strokes made by beginner players [13]. However, there are certain challenges in these methods such as the data insufficient to analyse a complete stroke and have it replicated by players at a given time. Previous work also used an infrared (IR) depth sensor to detect the in-corrected played strokes, but the method is time-consuming and costly, requiring a complicated sensor set up and custom proprietary software

[14]. Hence, to improve the beginner player's skill, an ubiquitous solution using prevalent equipment for stroke evaluation is essential.

1.2 Related works

Recently, mobile computing technology plays an important role in sports analysis. Due to their convenient and comfortable usage, wearable devices have become popular and are selected for physically sports groups of people. McCann et al. [4] introduced smart clothes and wearable technology which is a relatively novel and emerging area of interdisciplinary research within the fashion, textile, electronics, and related industries. The outcomes of this wearable technology are used to not only offer instructions for daily activities such as running and fitness but also support personal control over our quality of life, health, and well-being. With the successful outcome of this technology, the applications of specific targets and particular sports are investigated and implemented to enhance the player's level in their society.

Several kinds of research and applications have been done for sensor- and video-based sports analysis. Connaghan et al. [5] explored tennis stroke recognition using a wireless inertial measuring unit (WIMU) attached to a player's forearm during a competitive match. The stroke was classified into three main actions, forehand, backhand and serves, using accelerometer, gyroscope, and magnetometer sensors which are embedded into the WIMU. Ebner et al. [6], and Shah et. al. [7] used the IMU sensor and the traditional machine learning such as Support Vector Machine (SVM), or K-Nearest Neighbors (KNN) to detect and classify the tennis stroke. Lapinski et al. [8] proposed similar research which is a distributed wearable, wireless sensor system for evaluating professional baseball pitchers, and batters They implemented the SportSemble node which is the system into a node dedicated to wearable monitoring of athletic gesture, and specifically used it for measuring baseball players. For the experiment setup, they attached the inertial measurement units to the chest, upper arm, forearm, wrist, and waist. Lin et al. [9, 10] designed and implemented a mobile application a the badminton stroke classification using Random Forest. Egret et al. [11] performed the kinematic analysis of the golf swing for experienced golfers. They examined the differences between men and women golfers using the kinematic data obtained by the optoelectronic system with five cameras operating at 50 frames per second. Numerous sports have been investigated through the sensors- and video-based applications, especially stroke or swing analysis, however, there is limited study for table tennis.

For the table tennis applications, there are prior studies that focused on stroke analysis. Chen et al. [12] measured the quality of the broadcasting table tennis video. They used the video-based data to extract the high-level information of player, table, and ball positions, then derived the ball and player trajectories using Bayesian decision framework with Kalman filtering algorithm. Due to the video source from the broadcasting competitive game, the analysis is lacking the implication for the beginner or novice players. Blank et al. [13] presented the sensor-based detection and classification system for table tennis stroke. The data was collected from 10 participants with 8 basic table tennis stroke types by the inertial sensors attached to table tennis rackets. The stroke was detected by using an algorithm consisting of energy calculation, high-pass filtering, and threshold-based peak detection, and the stroke was classified by applying the Support Vector Machine (RBF kernel). However, because of the sensor embedded in the table tennis racket, the limitation of this work is the imbalance of the racket which affects the player's stroke motion directly. Additionally, the challenges of these methods are the data sufficient to analyse a complete stroke, and have it replicated by players at a given time. Not so far, Habiba et al. [14] published the first research that addresses the usage of IR depth camera on the table tennis player to detect and classify the strokes played in order to enhance the training experience. They tracked the position of skeleton joints, and collected the stroke data from the overall 10 players. To classify the player's strokes, they used a DTW-based algorithm, and achieve the best results compared to KNN, SVM, and Naive Bayes. Nevertheless, their method is time-consuming and costly, requiring a complicated sensor setup and custom proprietary software. Furthermore, although all aforementioned researches motivate us that the analysis of table tennis games can be analysed from the sensor- or video-based equipment, their researches used non-prevalent equipment which is difficult to reach for end-user. Therefore, investigating the table tennis stroke based on prevalent equipment is required, especially for implementation to the real application for end-user.

1.3 Objective

In this study, the research goal is to find the factor for improving the beginner players through investigating the difference in stroke consistency between the standard and the beginner players. We recorded videos and accelerator signals of two players group using prevalent equipment such as an Apple Watch, or GoPro8 action camera. For the video-based data, we use DeepLabCut to extract the arm's positions [15]. Then, we remove the noise signal from both extracted arm's position and accelerometer signals using Kalman filtering algorithm. Lastly, we evaluated the stroke consistency of both standard and beginner players by Dynamic Time Wrapping (DTW) with Root Mean Square Error (RMSE).

2. Methodology

2.1 Dataset

To investigate the stroke based on the player's skill, we selected two groups of table tennis players for this study. The first group is composed of standard players who have playing experience of more than 5 years, or participated in prior table tennis competitions according to the criteria proposed by [16]. The second group is composed of beginner or casual players who fall short of the requirement for the first group. The participants consist of **three standard**, and **seven beginner players**.

Table 1 shows the demographic information for the participants in this study. The experiments were performed using a total of 8 male and 2 female participant who are 22-28 years old. To maintain consistency, the participants were restricted to right-handed players. We recruited the participants through the google form, and obtained a written consent form for recording their data. Note that the data collected for this study was approved by the Ethics Review Committee for Human Studies at the Nara Institute of Science and Technology.

2.2 Experiment Setup

The main actions for table tennis are the forehand and backhand, to return the ball after the bounce on different playing surfaces over the table. Thus, we focused on the stroke of receiving the ball from the opponent by using forehand and backhand. Similar to a prior study [17], we set up a ball-shooting robot at the center of one side of the table, and let the players stand on the opposite side. The shooting angle is adjusted to facilitate the actions (forehand or backhand). Figure 1 demonstrated the procedure for this experiment. At the beginning of the experiment, the participants stand in the center area and were allowed to freely move laterally to either side, called the initial position. Then, they are asked to perform the stroke to return the ball before going back to the initial position to prepare for the next stroke.

Figure 2 shows the setup dimensions for the forehand experiment. The participants were instructed to return the ball from the robot (R) to the target area.

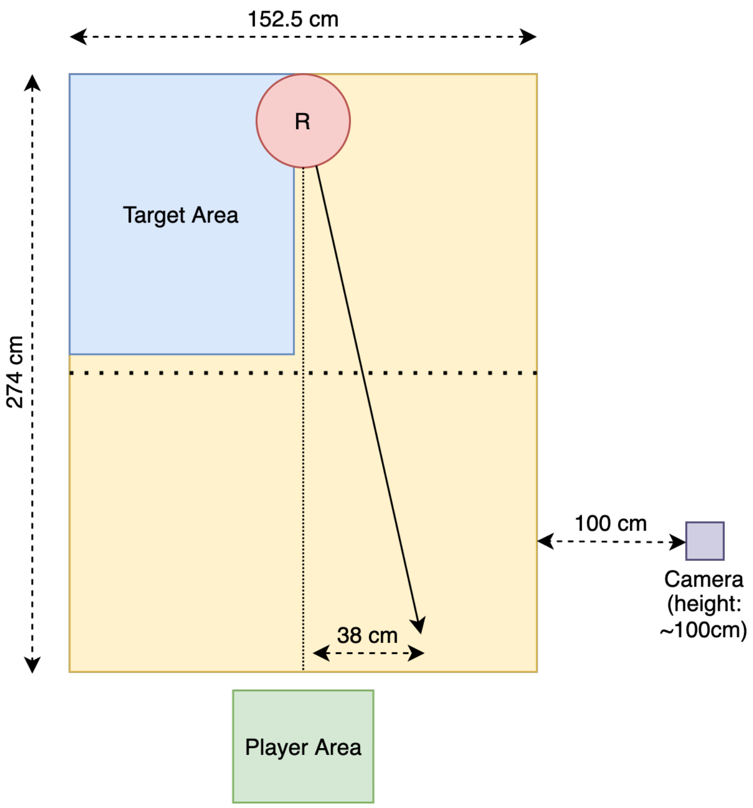
Table 1: Participant's demographic information. The dominant hand of all participants was the right hand

Participant	Gender	Age	Year of Experience
Standard#1	Male	24	7
Standard#2	Female	27	7
Standard#3	Male	22	5
Beginner#1	Male	25	3
Beginner#2	Female	26	1
Beginner#3	Male	24	1
Beginner#4	Male	24	4
Beginner#5	Male	28	3
Beginner#6	Male	23	4
Beginner#7	Male	25	1

Figure 1: Experiment procedures

If the ball is hit and returned to the target area, a score is counted. In the case of the backhand, the setup was simply mirrored horizontally. In this study, the participant consecutively performed 10 forehand strokes, and 10 backhand strokes.

In this study, we used UNIX NX2845 robo-star table tennis practice as the table tennis ball machine, and recorded the participants data using an Apple Watch 5 and a GoPro8 action camera. We tracked the motion of the wrist using a 3-axis accelerometer data logger from the Apple Watch. As illustrated in Figure 3, we tracked four positions: (1) the shoulder joint, (2) the elbow joint, (3) the wrist joint, and (4) the center of the racket. These positions were considered as representative points to determine the form of stroke at the time the player hit the ball. The camera was set to perform video capture of 10 consecutive stroke events for both forehand and backhand over 35 seconds. We configured the device capture rate at 60 frames per second with a 640x360 video resolution. Furthermore, to ensure that player movements are properly captured, the focal length of the camera was set to 19 mm to avoid visual distortion and placed approximately 100 cm away from the table.

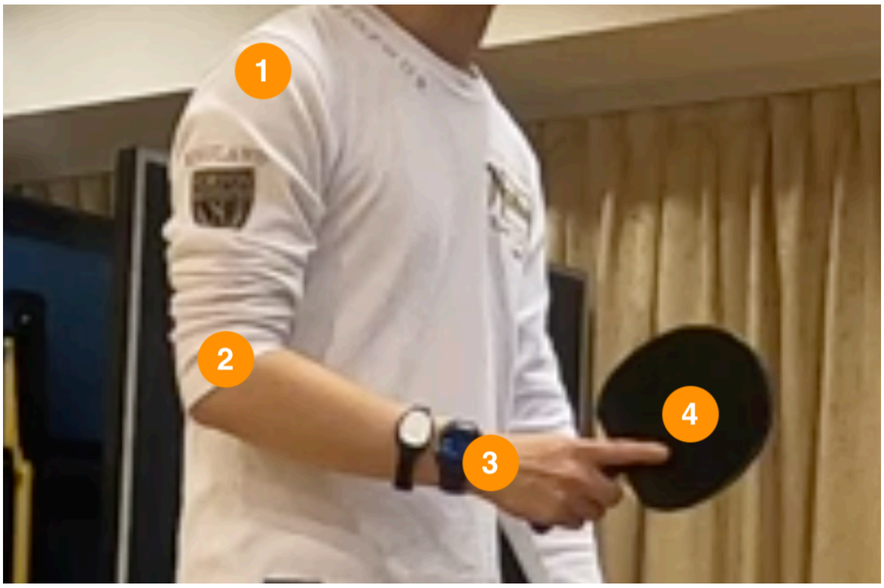


(a) Experiment table

(b) Gopro8 action camera

(c) Table tennis ball machine (R)

Figure 2: Experiment table and equipment setup.



(a) Four tracking positions: (1) the shoulder joint, (2) the elbow joint, (3) the wrist joint, and (4) the center of the racket.

(b) Apple Watch 5

Figure 3: Accessories for participants and four tracking points

2.3 Accelorometer and Preprocessing

To measure the consistency of each stroke, we analysed the x, y, and z-axis of an accelerometer signal data gathered using an Apple Watch 5 with a sampling time of 50Hz. Kalman filtering algorithm [18] was performed on the accelerometer data to filter out noise in the signal for each of the axis. The key idea of this filtering technique is to produce estimates of hidden variables based on inaccurate and uncertain measurements [19, 20]. Given a sequence of signal x from an accelerometer, the denoised signal $\widehat{x}$ can be calculated as follows:

$$\widehat{x}_t = K_t \cdot z_t + (1 - K_t) \cdot \widehat{x}_{t-1} \tag{1}$$

where z_t is the measurement value which we assume that it is not perfectly sure of these values, x_{t-1} is the estimate of the signal on the previous state, and K_t is called "Kalman Gain" which is the weight between z_t and x_{t-1}, and need to be optimized for each each consequent state. This allows us to remove jitters in data without eliminating some signal frequencies when compared to a moving average noise reduction approach as illustrated in Figure 4. In this study, we used the Kalman filtering algorithm implemented in pykalman library [21].

Here, we describe the procedure of Kalman filtering algorithm, which is a linear optimal state estimation method. To use this filtering, our problem needs to be satisfied the conditions that the signal value x and measurement value z must be calculated by using a linear equation as follows:

$$x_t = Ax_{t-1} + w_{t-1} \tag{2}$$

$$z_t = Hx_t + v_t \tag{3}$$

where x_t is a signal value called the state vector, z_t is a measurement value called the observation vector, u_t is a system control vector that is infrequent. A, and H are the state transition matrix, and the observation matrix, respectively. The two aforementioned parameters (A, and H) are used to represent the dynamics of the system, and are initialized as the identity matrix for this study. w_{t-1} is the process noise vector that is assumed to be zero-mean Gaussian with the covariance Q. v_t the observation noise vector that is assumed to be zero-mean Gaussian with the covariance R.

Kalman filtering algorithm consists of two stages: prediction and update. At the prediction state, the objective is to calculate a prior state estimate $\hat{x}_t^-$ which means the rough estimate before the update stage, and a prior estimate covariance P_t^-, which is used in update stage. These two priors can be calculated as follows:

$$\hat{x}_t^- = A\hat{x}_{t-1} \tag{4}$$

$$P_t^- = AP_{t-1}A^T + Q \tag{5}$$

where A, and Q are the state transition matrix, and the process noise covariance matrix, respectively. P_t is the error covariance that obtains from the update state. In this stage, the prior state estimate (Eq. 4) is evolved from the updated previous updated state estimate, and the prior estimate covariance (Eq. 5) encrypts the error covariance that the filter assumes the estimated error has.

For the update state, the objective is to find a posterior state estimate $\hat{x}_t$ which is the estimate of x at time t, a posterior estimate covariance P_t which is necessary for the prediction step at time $t+1$, as well as the Kalman Gain K_t that we mentioned in Equation 1. These three values can be calculated as follows:

$$K_t = P_t^- H^T(HP_t^- H^T + R)^{-1} \tag{6}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - H\hat{x}_t^-) \tag{7}$$

$$P_t = (I - K_t H)P_t^- \tag{8}$$

where H, R, and I are the observation matrix, the observation noise covariance matrix, and the identity matrix, respectively. We noted that it is required to know the estimate of x_0, and P_0 to start the process, and the set of parameters (A, H, Q, and R) is estimated by using Expectation–Maximization (EM) algorithm.. After we gathered all the information for the prediction and update stages, we can iterate through the estimates. The previous estimates will be the input for the current state. In the end, we obtain the denoised signal $\hat{x}_t$.

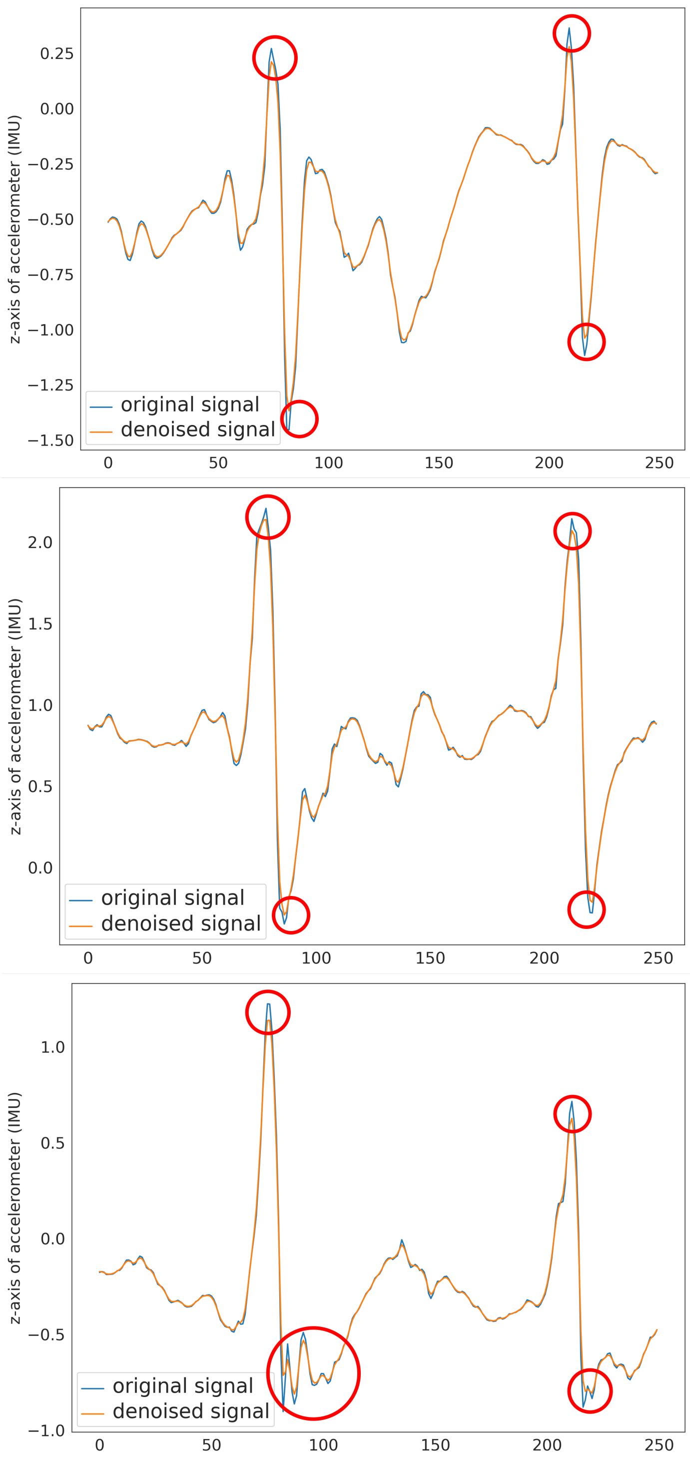


Figure 4: The comparison between before and after applying Kalman filtering to the signal from x, y, and z axis of accelorometer.

2.4 Video-Based Semi-Automatic Detection of Arm's Position

We used DeepLabCut, a supervised CNN-based tool for animal posture quantification, to analyze the players' stroke from the videos [15]. As shown in Fig. 5, this tool allows us to automatically annotate the positions of objects in entire videos, once their positions are manually provided in a few of the video's frames. Since DeepLabCut is based on ResNet50, a supervised learning model, it requires a sample of our videos for training the model. To create the training dataset, we randomly selected 10 frames per video, a total of 200 frames for training, and manually annotated the positions of three parts of the arm and the table tennis racket. Once trained, the tool then was used to estimate the positions throughout the remaining frames. In the end, we obtained the sequences of x and y coordinate of four tracking positions for each participant. We treated the results of the arm's position as the time-series data same with the x, y, and z-axis of accelerometer signal.

This model consisted of two components, pretrained ResNet50 [22], and deconvolutional layers [23]. For the first component, as illustrated in Figure 6, the ResNet50 is a convolutional neural network (CNN) that is 50 layers, and consisted of 5 stages each with a convolution and identity block [24, 25]. This model used the max pooling layers [26, 27] along with the rectified linear unit (ReLU) activation function [28], and batch normalization [29]. As the "skip connection" mechanism in this component, this model is able to resolve the vanishing gradient problem [30]. For the second component, the deconvolution layers are used for up-sampling the visual information and producing the spatial probability densities. By a stochastic gradient descent [31], the model is trained more than 200,000 iterations with setting the learning rate as 0.0005 and weight decay as 0.0001. The training time is taken up to 8 hours, and the loss is less than 0.001. We noted that our model used the pre-trained weight from the ResNet50 for initialization.

For the setup of computational resources, the pose estimation model in this study was implemented, trained, tuned, and evaluated on the Google Colaboratory platform with Intel(R) Xeon(R) CPU @ 2.30GHz CPU, 12 GB of RAM, and Tesla P100-PCIE-16GB GPU. The model was run on Python 3.6.8, Keras 2.2.5, and Tensorflow 1.14.0.

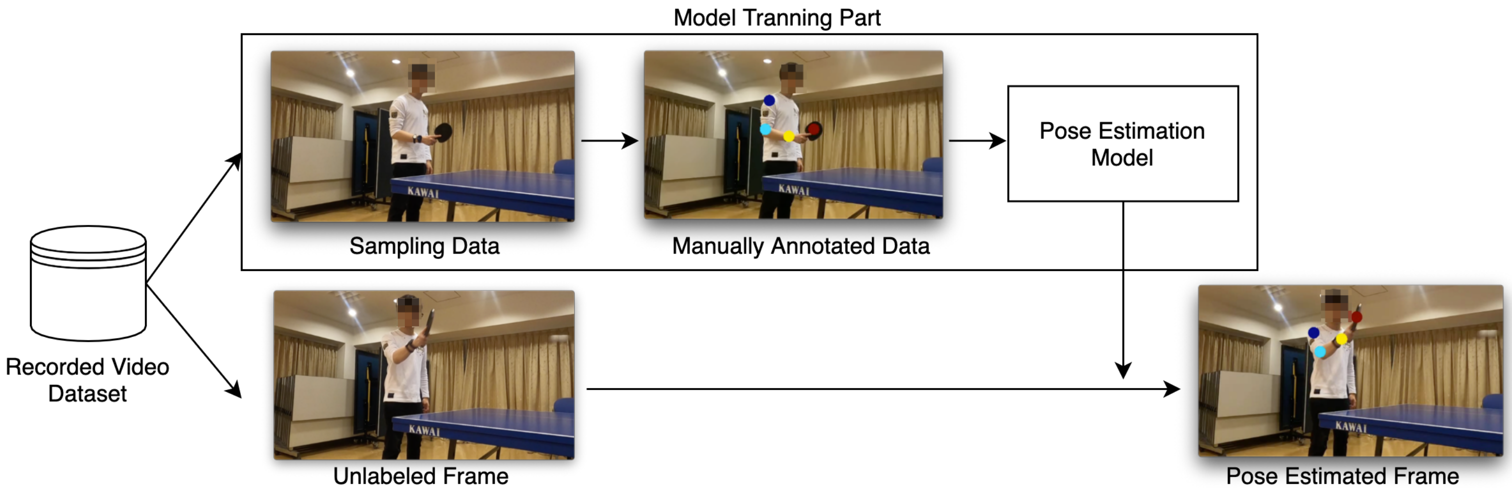


Figure 5: Procedure of Semi-automatic detection via DeepLabcut

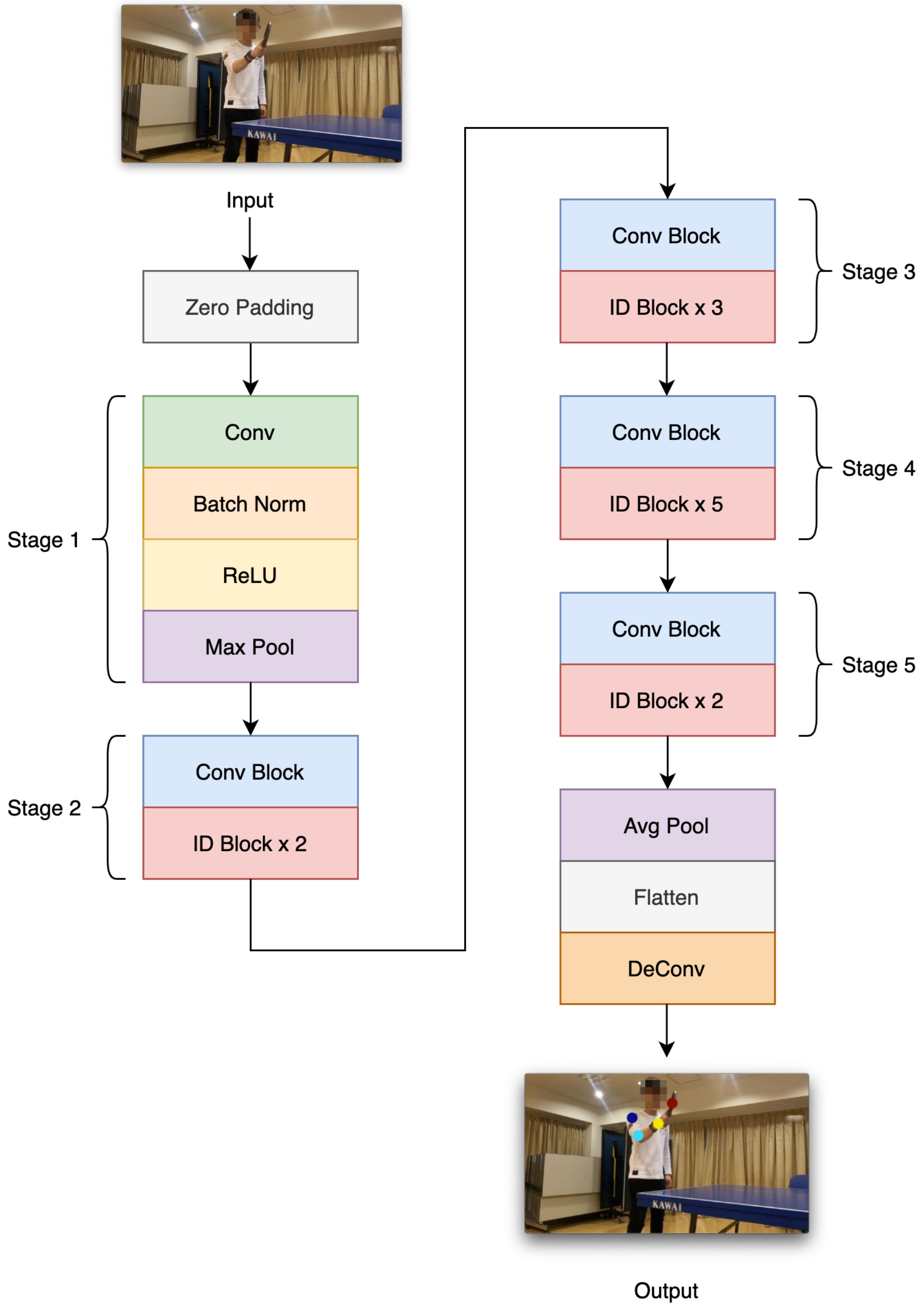


Figure 6: Pose estimation architecture in DeepLabCut. Conv and DeConv are convolutioanl layer, and deconvolutional layer, respectively.

2.5 Dynamic Time Warping for Stroke Consistency Evaluation

We used Dynamic Time Warping (DTW) and Root Mean Square Error (RMSE) on pairs of two strokes from each player, in order to evaluate the consistency between the stroke motion signals from video motion tracking and the Apple Watch accelerometer. Prior works applied this algorithm for analysis on various fields such as basketball [32], and surgery [33]. The use of DTW allows for matching the peaks of the strokes by reducing the effects of time and shifting distortion between the two signals and detects how similar their phase and shapes as described by Algorithm 1.

RMSE was used as the metric to compare the consistency of the stroke motions after using DTW on the signals by measuring the square root of the difference between two compared signals as shown in Equation 9. RMSE can be calculated as follows:

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(a_i - b_i)^2} \tag{9}$$

where a and b are two input sequences, and n is the total number of the feature which denotes to the index of the element of a vector, and equal to 1 for this study. In this context, a sizeable difference in RMSE implies greater inconsistency and erraticism in the stroke motion. For each of the players, we selected the reference signal as signal that produces the lowest averaged RMSE when paired with other strokes. Further note, we discarded the first and last strokes because they contain signals from initial and ending postures during the recording. For each player, we calculated RMSE from the set of signals, and use it as the feature for analysis. The video tracking approach contains eight features, x and y coordinate of four tracking points from the GoPro8 action camera. For the accelerometer tracking approach, there are three features, x-axis, y-axis, and z-axis of the accelerometer from Apple Watch.

Algorithm 1: Dynamic Time Warping Algorithm

Data: s, t: the sequences
Result: d: the minimum distance
$n = s$.length() ;
$m = t$.length() ;
DTW = array[0...n][0...m] ;
for $i \leftarrow 0$ **to** n **do**
  for $j \leftarrow 0$ **to** m **do**
    DTW[i][j] = infinity
  end
end
DTW[0][0] = 0 **for** $i \leftarrow 0$ **to** n **do**
  for $j \leftarrow 0$ **to** m **do**
    cost = RMSE($s[i]$, $t[i]$) ;
    DTW[i][j] = cost + minimum(DTW[$i-1$, j], DTW[i, $j-1$],
    DTW[$i-1$, $j-1$])
  end
end
return DTW[n][m];

2.6 Statistical Testing

To investigate which factor(s) were correlated with the consistency of the stroke motions, we performed statistical comparisons between the following pairs:

- successful v.s. failure results when the beginners tried forehand strokes.
- successful v.s. failure results when the beginners tried backhand strokes.
- standard v.s. beginner players when the players tried forehand strokes.
- standard v.s. beginner players when the players tried backhand strokes.
- successful v.s. failure results in case of forehand strokes.
- successful v.s. failure results in case of backhand strokes.

The comparison was implemented by two-sample t-test (significance level = 1.0e-4 without adjusting for multiple comparisons) for each coordinate. We calculated t value of each comparison as follows:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{10}$$

where $\overline{x}_1$ and $\overline{x}_2$ are the means of the two samples, s_1 and s_2 are the standard deviations of the two samples, n_1 and n_2 are the sample sizes. The number of degrees of freedom for the problem is the smaller of n_1 and n_2.

For the relationship between the player's stroke and the hitting score, we defined the **stability error** for representing the player's stroke consistency. We applied Principle Component Analysis (PCA), which is a technique for reducing the dimensionality of the features, and have been used in various works [34–36]. We then selected the first principal component score as the stability error. For example, we reduced the dimensional of the eight features of the video tracking approach into one feature called stability error. This procedure was also used with the features for the accelerometer tracking approach. To evaluate the strength of the correlation between the stability error and the hitting score, we applied Pearson's correlation coefficient and used the criteria according to the previous work [37, 38].

3. Results

3.1 Detection of Arm's Position

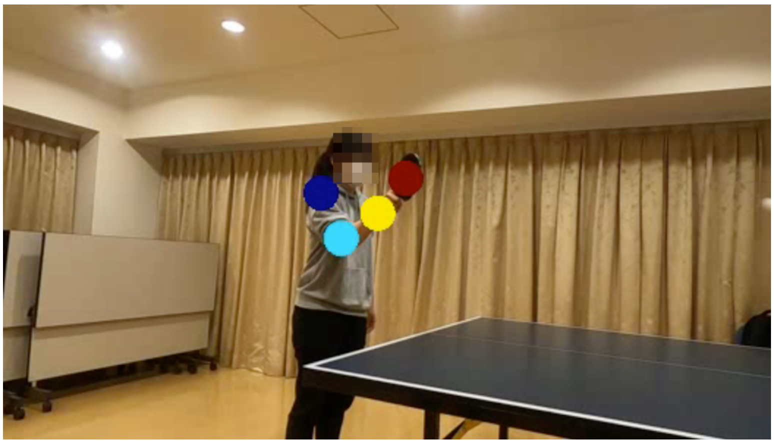
As a preliminary assessment, we analyze the detection of three arm's positions and the center of the racket. The annotation results are shown in Figure 7 for the forehand stroke, and Figure 8 for the backhand stroke. The blue, cyan, yellow, and crimson colors represent (1) the shoulder joint, (2) the elbow joint, (3) the wrist joint, and (4) the center of the racket, respectively. We can see that the annotated arm's positions are precise and are well prepared to be used in the stroke analysis for the participants. With the pose estimation model, we obtained the x, and y coordinates of the three arm's position and the center of the racket of the video-based data in this study. Again, we treated the results of the tracked position as the time-series data same with the x, y, and z-axis of the accelerometer signal for the stroke analysis.

3.2 Stroke Consistency and Statistical Analysis

We analysed the stroke consistency using averaged RMSE of video tracking and accelerometer tracking. Table 2 and 3 shows the averaged and standard deviation of RMSE for video tracking and accelerometer tracking, and the hitting score of each player, for forehand and backhand, respectively. We have a total of eleven features, eight features from video tracking and three features from accelerometer tracking. For a simple analysis, we average the feature depending on the feature group and the dominant hand such as video tracking of forehand, video tracking of backhand, accelerometer-based tracking of forehand, accelerometer-based tracking of backhand. Noted that the results of the video tracking approach are based on the pixel-to-pixel difference of footage from the GoPro8 camera, while the results of the accelerometer tracking are based on the inertial measurement unit (IMU) in the Apple Watch 5.

Considering the forehand results in Table 2, we see that the lowest RMSE of the standard player's group is 162.07 ± 100.36 for video tracking, and 3.39 ± 1.10 for accelerometer tracking. For the beginner player's group, the lowest RMSE is 265.60 ± 124.65 for video tracking, and 7.07 ± 2.05 for accelerometer tracking.

(a)

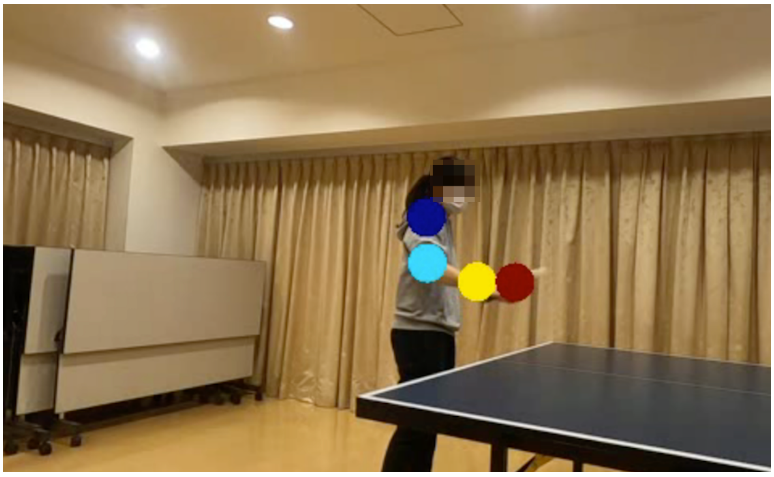


(b)

(c)

Figure 7: The automatically annotation results of the forehand strokes of the standard players

(a)



(b)

(c)

Figure 8: The automatically annotation results of the backhand strokes of the standard players

It can be implied that the stroke consistency of standard players is more efficient than the beginner players. Furthermore, all standard player's group achieve the perfect score with the low RMSE for forehand stroke. In contrast to the forehand results, as shown in Table 3, we see that the lowest RMSE of the standard player's group is 205.12 ± 106.70 for video tracking, and 4.27 ± 0.92 for accelerometer tracking, while, the lowest RMSE of beginner player's group is 87.76 ± 36.17 for video tracking, and 6.42 ± 1.01 for accelerometer tracking. The results show that some beginner players, who achieve the low RMSE for video and accelerometer tracking, can obtain the various hitting score.

For pair-wise comparisons, after we applied two-sample t-test for our features, the coordinates showing significant difference were as follows:

As results of pair-wise comparisons,

1. **standard v.s. beginner players when the players tried forehand strokes**
   - video tracking: x and y coordinate of elbow joint, wrist joint, and the center of the racket.
   - accelerometer tracking: x-axis, y-axis, and z-axis
2. **standard v.s. beginner players when the players tried backhand strokes**
   - accelerometer tracking: y-axis, and z-axis

We noted that the significant level in these results is 1.0e-4 without adjusting for multiple comparisons. These pair-wise comparison results suggest that we can distinguish the forehand stroke between the standard player's group and the beginner players' group through the video and accelerometer tracking. On the other hand, we can use only the accelerometer tracking for backhand stroke. For other comparisons, they did not show a significant difference.

Table 2: Average and standard deviation of RMSE within players forehand strokes

Participant	**Video Tracking (RMSE)**	**Accelerometer Tracking (RMSE)**	**Hitting score**
Standard#1	214.91 ± 73.86	6.41 ± 0.49	10
Standard#2	162.07 ± 100.36	6.93 ± 0.92	10
Standard#3	243.56 ± 175.19	3.39 ± 1.10	10
Beginner#1	663.73 ± 388.00	9.36 ± 3.62	10
Beginner#2	1064.31 ± 868.51	7.98 ± 1.00	7
Beginner#3	401.66 ± 230.38	10.62 ± 3.72	7
Beginner#4	265.60 ± 124.65	9.59 ± 1.00	10
Beginner#5	698.91 ± 319.56	16.96 ± 1.58	6
Beginner#6	605.69 ± 456.54	7.07 ± 2.05	8
Beginner#7	725.51 ± 632.22	21.90 ± 7.33	7

Table 3: Average and standard deviation of RMSE within players backhand strokes

Participant	**Video Tracking (RMSE)**	**Accelerometer Tracking (RMSE)**	**Hitting score**
Standard#1	237.12 ± 114.34	4.27 ± 0.92	10
Standard#2	539.16 ± 475.45	6.76 ± 0.61	9
Standard#3	205.12 ± 106.70	4.34 ± 0.85	7
Beginner#1	87.76 ± 36.17	10.80 ± 3.21	10
Beginner#2	147.92 ± 65.76	9.87 ± 2.31	6
Beginner#3	335.59 ± 207.97	7.45 ± 1.40	8
Beginner#4	391.58 ± 274.08	13.30 ± 1.80	8
Beginner#5	562.62 ± 167.08	11.35 ± 1.85	7
Beginner#6	394.33 ± 428.32	6.42 ± 1.01	9
Beginner#7	291.21 ± 141.69	9.10 ± 4.16	6

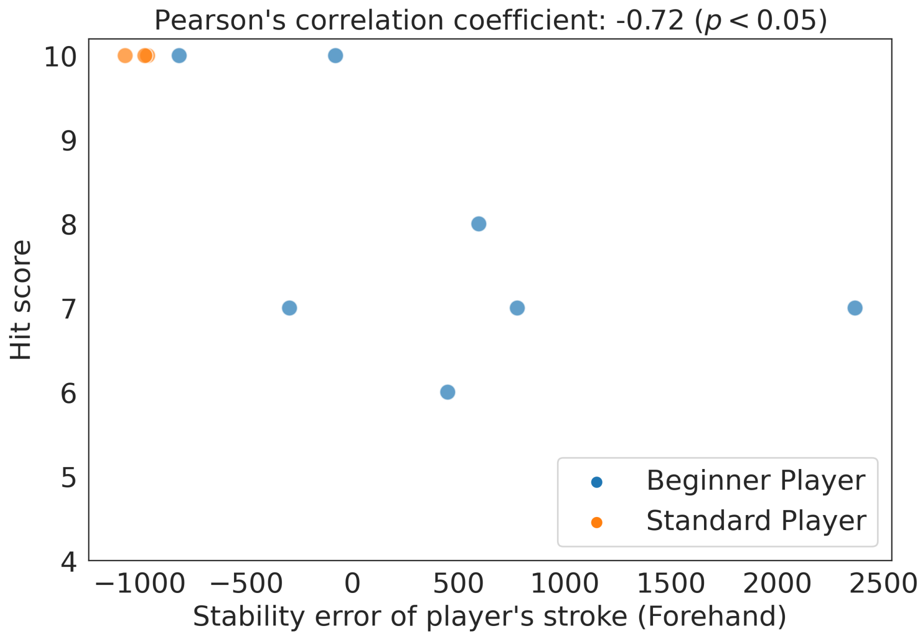
3.3 Player's Stroke and Hitting Score

To examine the relationship between the player's stroke and hitting score, we analysed the stability error and the hitting score. Figure 9 and 10 show the relationship for video and accelerometer tracking approach, respectively. The x-axis is the stability error of the player's stroke which obtain by applying PCA to the feature for each approach. The y-axis is the hitting score. The blue point represents the beginner player, while the yellow point represents the standard player.

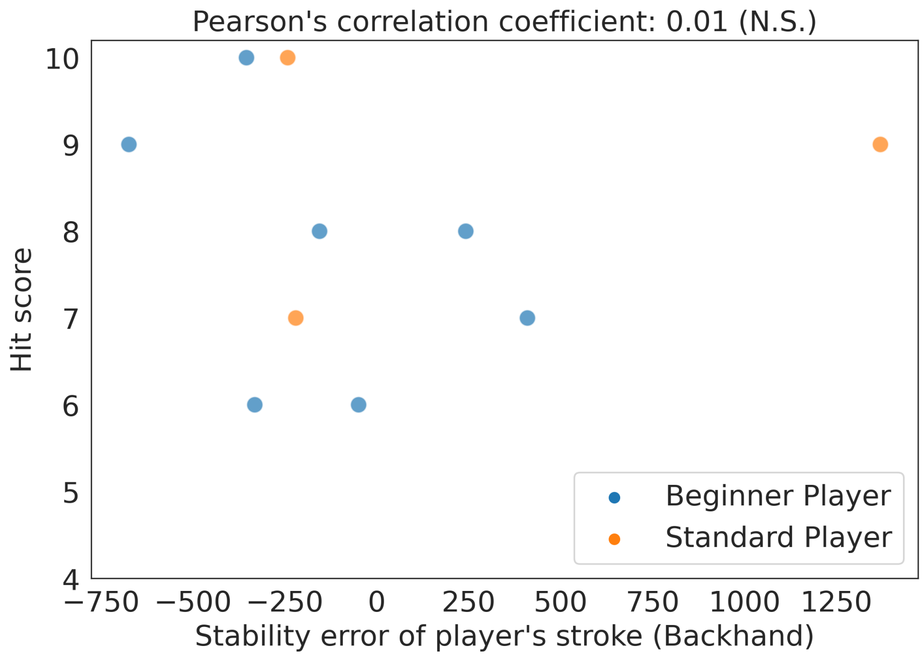
For the results of the video tracking approach as shown in Figure 9a, we can see that there is an obvious difference in stability error between standard and beginner players for forehand stroke. We noted that there are overlapped two points of standard player's group in this visualization. On the other hand, Figure 9b suggests that there are unclear differences for backhand. Based on the result from Pearson's correlation, we see that there are strong negative and negligible correlations for the forehand and the backhand strokes, respectively.

Similar to the video tracking approach, the forehand result of the accelerometer also shows the difference between two player's groups through the stability error as shown in 10a. For the backhand stroke, the differences seem to be more clear comparing to the video tracking approach. The interpretation results from Pearson's correlation suggest that there are moderate negative and negligible correlations for the forehand and the backhand strokes, respectively.

Based on these results, it is obvious that there is a difference of stability error between standard and beginner players for forehand results for video and accelerometer tracking approaches. These imply that stability error can be an efficient candidate for evaluating a player's skill, especially for right-handed players. On the other hand, there are unclear differences between the two players groups for the backhand strokes.

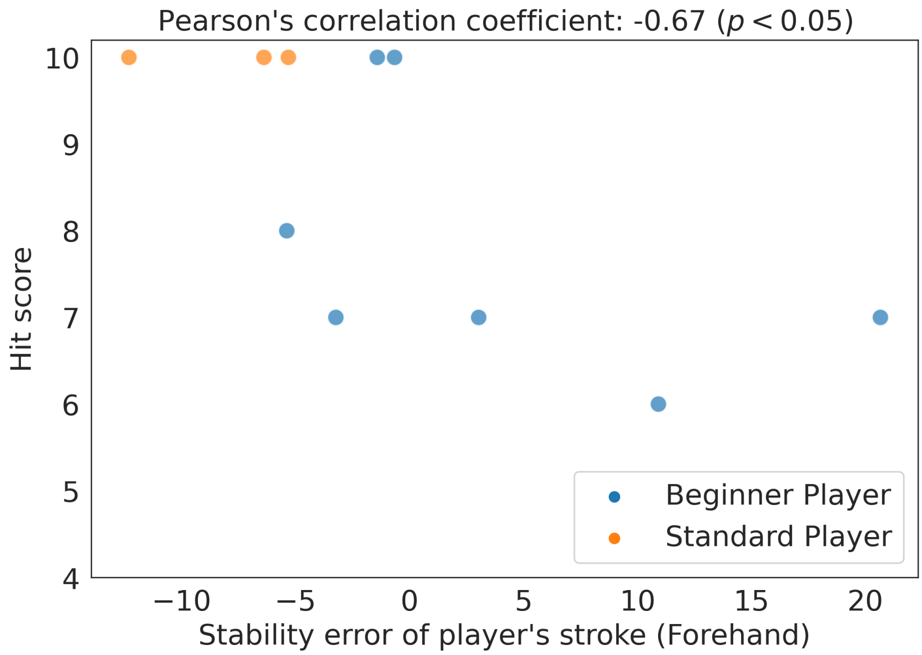


(a)

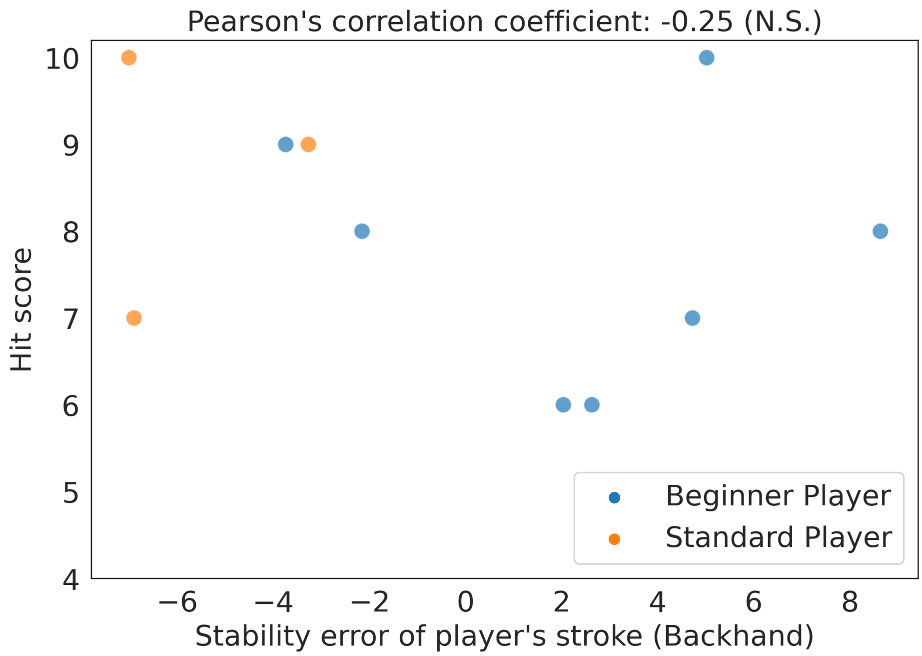


(b)

Figure 9: Scatter plot between stability error of player's stroke and their hitting score of video tracking approach



(a)



(b)

Figure 10: Scatter plot between stability error of player's stroke and their hitting score of accelerometer tracking approach

4. Discussion and Conclusion

In this work, we present a multimodal approach for table tennis stroke evaluation. We recorded the videos and accelerator signals of a total 10 players of standard and beginner players through the Apple Watch 5 and Gopro8 action camera. We then applied Kalman filtering algorithm to remove the noise from the accelorometer, and extracted the arm's position using DeepLabCut. Lastly, we evaluate the table tennis stroke consistency through Dynamic Time Wrapping with RMSE, and perform statistical testing analysis. Our results suggest that there is a difference in stroke consistency between standard and beginner players based on the multimodal approaches, especially for the forehand stroke. The statistical analysis reveals that the significant differences of the forehand strokes between the two players groups are the x and y coordinate of elbow joint, wrist joint, and the center of the racket from video tracking, and x, y, and z-axis from accelerometer tracking. On the other hand, only the y- and z-axis from accelerometer tracking are significant differences for the backhand stroke. For the relationship between the stability error and the hitting score, we see the strong negative and moderate correlation of the forehand stroke for both tracking approaches. These findings suggest that the stability error can be an efficient candidate for evaluating a player's skill, especially for the forehand stroke of right-handed players.

As the results show that some beginner players, who achieve the low RMSE for video and accelerometer tracking, can obtain the various hitting score for the backhand strokes. Additionally. for the relationship between the stability error and the hitting score, there are unclear differences between the two players groups for backhand. We suspect that it is possibly affected by the camera position setup, the slight movement for the backhand stroke, the player's dominant hand, and so on. Therefore, we conclude that any patterns in the backhand strokes are unclear to our current study, and more investigation is required.

There are three main limitations that affect the accuracy and reliability of our results. The first is that the result of using DeepLabCut is affected by the camera position, especially for the backhand. Regarding the full stroke motion for the forehand of the right-handed player, we set the camera position to see the entire stroke. With the full stroke motion, we can analyse the forehand stroke accurately. On the other hand, as shown in Figure 11, when the player performs the backhand, their left hand can be intercepted by their body. This problem might be a cause of inaccurate analysis for the backhand stroke. To resolve this issue, we aim to place another camera to capture the entire player's stroke for the backhand. The second is Dynamic Time Warping (DTW) discards the timing information. The nature of Dynamic Time Warping (DTW) is to match the peak of the strokes by reducing the effects of time and shifting distortion. While DTW is able to capture how consistent is the stroke motion, the stroke duration information is not included in the analysis results. Hence, measuring additional insight such as speed of the stroke can be considered given that stroke consistency, hit timing, and hand-eye coordination movements are affected by speed not only in table tennis but in other sports as well [39]. The last limitation is the small size of participants and the stroke types in this study. As the hitting score represents the success or failure strokes, we can see that our dataset contains a small amount of fail strokes. Mainly, we are lacking the failure strokes from standard players for forehand. Also, due to the technical difficulty, we cannot perform the rally situation. Therefore, these analyses should be validated by increasing the number of participants and strokes types for all player groups in further studies.

With our evaluation procedure, we can monitor the beginner player's skill by using the stroke consistency through the prevalent equipment. Regarding the training of beginner players, one possible way is to use the stroke analysis of the standard player as a template of proper stroke for beginner players to follow. Since our data consists of the arm's tracked position and accelerometer tracked signal, we can build the stroke template from this data. Afterward, the beginner players can use the template to improve their strokes to be closer to the standard player. From an engineering standpoint, we aim to implement these applications on both smartwatch and mobile phone as a monitoring and training device for convenience and accessibility purposes in future directions of this study.

(a) The example of the common backhand stroke

(b) The example of the backhand stroke when the participant's body intercept their left hand

Figure 11: The example of the backhand stroke

Acknowledgements

Firstly, I would like to thank my supervisors, Prof. Kazushi Ikeda, Prof. Keiichi Yasumoto, Assoc. Prof. Junichiro Yoshimoto, Assoc. Prof. Takatomi Kubo, Asst. Prof. Makoto Fukushima, and Asst. Prof. Chie Hieida. They gave me a lot of useful comments and recommendations in Colloquiam B and our Lab seminar. I have improved myself, including technical knowledge, problem-solving, research, and so on. In particular, I am appreciate the helps and supports from Prof. Ikeda and Assoc. Prof. Yoshimoto. They gave me not only the research advice but also the opportunities such as participating the conferences. These experiences will absolutely help me in the future after graduation from the University. I would like to thank the secretaries, Ms. Aya Tanimoto and Ms. Sachiko Yasuki. They helped us with supports of the research and the documents tasks. Also, I would like to thank our Lab members. In particular, PhD. Bryan Lao, he always gave me advice kindly after my Lab seminar. I was able to understand what are good or bad in my presentation because of his comments and recommendations. Lastly, I would also like to thank my family. Thank you very much for watching over my challenges so far. I will keep on trying.

References

[1] Adrian Lees. Science and the major racket sports: a review. *Journal of Sports Sciences*, Vol. 21, No. 9, pp. 707–732, 2003.

[2] Irene Faber, Marije Elferink-Gemser, Frits Oosterveld, and Maria Nijhuis-van der Sanden. Revision of two test items of the dutch motor skills assessment measuring ball control in young table tennis players: a reproducibility and validity study. *Annals of Research in Sport and Physical Activity*, Vol. 5, pp. 53–69, 01 2015.

[3] Miran Kondric, Alessandro Zagatto, and Damir Sekulic. The physiological demands of table tennis: A review. *Journal of sports science & medicine*, Vol. 12, pp. 362–70, 10 2013.

[4] J. McCann, David Bryson, Mikko Malmivaara, Richard Hurford, Fatema Saifee, Laura RCA, Dr Kane, Adam Martin, Peter Lam, Gao Min, Sirpa Morsky, Xiang Dong, Ida Agnusdei, A. Taylor, Dr Treadaway, Mike Timmins, Sally Underwood, Johannes Birringer, Michele Danjoux, and Will Stahl. *Smart Clothes and Wearable Technology*. 02 2009.

[5] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna. Multi-sensor classification of tennis strokes. In *SENSORS, 2011 IEEE*, pp. 1437–1440, 2011.

[6] Christopher J. Ebner and Rainhard Dieter Findling. Tennis stroke classification: Comparing wrist and racket as imu sensor position. In *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*, p. 74â 83, 2019.

[7] Hitesh Shah, Prakash Chokalingam, Balamanohar Paluri, Nalin Pradeep, and Balasubramanian Raman. Automated stroke classification in tennis. In Mohamed Kamel and Aurélio Campilho, editors, *Image Analysis and Recognition*, pp. 1128–1137, 2007.

[8] Michael Lapinski, Eric Berkson, Thomas Gill, Mike Reinold, and Joseph

Paradiso. A distributed wearable, wireless sensor system for evaluating professional baseball pitchers and batters. pp. 131–138, 09 2009.

[9] Juyi Lin, Chia-Wei Chang, TsÃ¬-UÃ Ik, and Yu-Chee Tseng. Sensor-based badminton stroke classification by machine learning methods. In *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pp. 94–100, 2020.

[10] Juyi Lin, Chia-Wei Chang, Chih-Hao Wang, Hong-Chuan Chi, Chih-Wei Yi, Yu-Chee Tseng, and Chih-Chuan Wang. Design and implement a mobile badminton stroke classification system. In *2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pp. 235–238, 2017.

[11] Claire Egret, B Nicolle, F Dujardin, J Weber, and Chollet Didier. Kinematic analysis of the golf swing in men and women experienced golfers. *International journal of sports medicine*, Vol. 27, pp. 463–7, 07 2006.

[12] W. Chen and Y. Zhang. Tracking ball and players with applications to highlight ranking of broadcasting table tennis video. In *Computational Engineering in Systems Applications*, Vol. 2, pp. 1896–1903, 2006.

[13] Peter Blank, Julian Hoßbach, Dominik Schuldhaus, and Bjoern M. Eskofier. Sensor-based stroke detection and stroke type classification in table tennis. In *ACM International Symposium on Wearable Computers*, p. 93â 100, New York, NY, USA, 2015.

[14] Habiba Hegazy, Mohamed Abdelsalam, Moustafa Hussien, Seif Elmosalamy, Yomna M.I Hassan, Ayman M. Nabil, and Ayman Atia. Online detection and classification of in-corrected played strokes in table tennis using ir depth camera. *Procedia Computer Science*, Vol. 170, pp. 555 – 562, 2020.

[15] Alexander Mathis, Pranav Mamidanna, Kevin Cury, Taiga Abe, Venkatesh Murthy, Mackenzie Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, Vol. 21, , 09 2018.

[16] Alessandro Zagatto and Claudio Gobatto. Relationship between anaerobic parameters provided from maod and critical power model in specific table tennis test. *International journal of sports medicine*, Vol. 33, pp. 613–20, 05 2012.

[17] Taisa Belli, Milton Shoiti Misuta, Pedro Paulo Ribeiro de Moura, Thomas dos Santos Tavares, RenÃª Augusto Ribeiro, Yura Yuka Sato dos Santos, Karine Jacon Sarro, and Larissa Rafaela Galatti. Reproducibility and validity of a stroke effectiveness test in table tennis based on the temporal game structure. *Frontiers in Psychology*, Vol. 10, p. 427, 2019.

[18] J. Woods and C. Radewan. Kalman filtering in two dimensions. *IEEE Transactions on Information Theory*, Vol. 23, No. 4, pp. 473–482, 1977.

[19] Youngjoo Kim and Hyochoong Bang. *Introduction to Kalman Filter and Its Applications*. 11 2018.

[20] Qiang Li, Ranyang Li, Kaifan Ji, and Wei Dai. Kalman filter and its application. In *International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, pp. 74–77, 2015.

[21] pykalman, the dead-simple kalman filter, kalman smoother, and em library for python, 2015. Available at `https://github.com/pykalman/pykalman/`, version 0.9.2.

[22] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[23] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[24] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, 2017.

[25] Rikiya Yamashita, Mizuho Nishio, Richard Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, Vol. 9, , 06 2018.

[26] H. Gholamalinezhad and H. Khosravi. Pooling methods in deep neural networks, a review. *ArXiv*, Vol. abs/2009.07485, , 2020.

[27] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, Vol. 29, pp. 1–98, 06 2017.

[28] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv:1811.03378*, 2018.

[29] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37, pp. 448–456, 07 2015.

[30] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 6, pp. 107–116, 04 1998.

[31] Sebastian Ruder. An overview of gradient descent optimization algorithms. *ArXiv*, Vol. abs/1609.04747, , 2016.

[32] Xinyao hu, Shaorong Mo, and Xingda Qu. Basketball activity classification based on upper body kinematics and dynamic time warping. *International Journal of Sports Medicine*, Vol. 41, , 01 2020.

[33] Jingyu Jiang, Yuan Xing, Shuxin Wang, and Ke Liang. Evaluation of robotic surgery skills using dynamic time warping. *Computer methods and programs in biomedicine*, Vol. 152, pp. 71–83, 2017.

[34] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society*

A: Mathematical, Physical and Engineering Sciences, Vol. 374, p. 20150202, 2016.

[35] Shruti Sehgal, Harpreet Singh, Mohit Agarwal, V. Bhasker, and Shantanu. Data analysis using principal component analysis. In *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, pp. 45–48, 2014.

[36] Andrzej MaÄ kiewicz and Waldemar Ratajczak. Principal components analysis (pca). *Computers & Geosciences*, Vol. 19, No. 3, pp. 303–342, 1993.

[37] Patrick Schober, Christa Boer, and Lothar Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, Vol. 126, p. 1, 02 2018.

[38] Mavuto Mukaka. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*, Vol. 24, pp. 69–71, 09 2012.

[39] Gabriel Delgado-GarcÃa, Jos Vanrenterghem, Alejandro MuÃ±oz-GarcÃa, Alejandro Molina-Molina, and VÃctor M. Soto-Hermoso. Does stroke performance in amateur tennis players depend on functional power generating capacity? *The Journal of Sports Medicine and Physical Fitness*, Vol. 59, , 2019.

Publication and Conference List

1. **Panyawut Sri-iesaranusorn**, Felan Carlo Garcia, Francis Tiausas, Supatsara Wattanakriengkrai, Kazushi Ikeda, and Junichiro Yoshimoto, "Toward the Perfect Stroke: A Multimodal Approach for Table Tennis Stroke Evaluation", International Conference on Mobile Computing and Ubiquitous Networking (ICMU) 2021, accepted

2. **Panyawut Sri-iesaranusorn**, Attawit Chaiyaroj, Chatchai Buekban, Songphon Dumnin, Ronachai Pongthornseri, Chusak Thanawattano, and Decho Surangsrirat, "Classification of 41 Hand and Wrist Movements via Surface Electronyogram using Deep Neural Network," Frontiers in Bioengineering and Biotechnology, June, 2021

3. **Panyawut Sri-iesaranusorn**, Raula Gaikovina Kula, Takashi Ishio, "Does Code Review Promote Conformance? A Study of OpenStack Patches," Mining Software Repositories Conference (MSR), May, 2021

4. Raksit Raksasat, **Panyawut Sri-iesaranusorn**, J. Pemcharoen, P. Laiwarin, S. Buntoung, S. Janjai, E. Boontaveeyuwat, P. Asawanonda, Sira Sriswasdi and Ekapol Chuangsuwanich, "Accurate Surface Ultraviolet Radiation Forecasting for Clinical Applications with Deep Neural Network," Scientific Reports, March, 2021

5. Peerapon Vateeekul, **Panyawut Sri-iesaranusorn**, Pawit Aiemvaravutigul, Adsadawut Chanakitkarnchok, and Kullatida Rojviboonchai, "Recurrent Neural-Based Vehicle Demand Forecasting and Relocation Optimization for Car-Sharing System: A Real Use Case in Thailand," Journal of Advanced Transportation, Jan, 2021

6. **Panyawut Sri-iesaranusorn**, Felan Carlo Garcia, Francis Tiausas, Supatsara Wattanakriengkrai, Kazushi Ikeda, and Junichiro Yoshimoto, "Evaluating Table Tennis Stroke based on Arm Movement Tracking using DeepLabCut", SSI2020, Online, November, 2020.

7. **Panyawut Sri-iesaranusorn**, Supatsara Wattanakriengkrai , Teyon Son, Takeru Tanaka, Christopher Wiraatmaja, Takashi Ishio, and Raula Gaikov-

ina Kula, ”Kode_Stylers: Author Identification through Naturalness of Code: An Ensemble Approach.” Forum for Information Retrieval Evaluation (FIRE), 2020

8. **Panyawut Sri-iesaranusorn**, Saeka Shimochi, Naoki Ono, Emrah Yatkin, Hidehiro Iida, Kazushi Ikeda, and Junichiro Yoshimoto, “Quantification Analysis of Behavioral Changes after Sciatic Nerve Ligation in Rats,” Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA-ASC), vol.in press, Dec, 2020

9. **Panyawut Sri-iesaranusorn**, Saeka Shimochi, Naoki Ono, Emrah Yatkin, Hidehiro Iida, Kazushi Ikeda, and Junichiro Yoshimoto, “A preliminary study on quantification of behavioral changes after sciatic nerve ligation in rats,” Winter Workshop on Mechanism of Brain and Mind, Jan, 2020