JOURNAL OF
EMERGING INVESTIGATORS

Article

Analysis of professional and amateur tennis serves using computer pose detection

Lisa Baily[1], Jonathan Lai[2]
[1] The American School in Japan, Tokyo, Japan
[2] Tokyo Coding Club, Tokyo, Japan

SUMMARY

Incorporating computer vision and machine learning into sports is an effective and efficient way to improve players' skills. It is often quite challenging to know specifically where a player can fix their performance just by playing or looking at videos. Advances in computer vision allow one to convert generic video imagery of a person to generate limb coordinates as a function of time and apply that to analyze the tennis serves of professionals, semi-professionals, and amateur players. We hypothesized that the overall movement of different players' serves is similar despite their skill level, but the key difference is the timing during the serve at which their racket makes contact with the ball. We analyzed video imagery of 23 different tennis players serving, with each player having 4 to 5 serves, and generated x-y coordinates of the players' limbs throughout the duration of the serve. We then analyzed the full dynamics of each players' serve and identified key differences between players of varying skill level. Based only on the different timing between skill levels, we were able to determine the skill level of players from a generic video. Identifying clear differences between professional and amateur athletes can greatly improve the focus of training and technique.

INTRODUCTION

Professionals and semi-professionals in virtually all sports, including tennis, use video data to improve their skills. With advances in computer vision, some computer vision applications can now score the players' performances in areas specific to their sport, allowing them to make changes to their strategies. Tennis is a sport where an increasing amount of data analytics has been applied; one famous technology now widely used is PlaySight, which uses cameras from multiple angles and can provide video data to support players during their practice (1). Several researchers have analyzed tennis video data to study the importance of timing in tennis. In a 2010 analysis, Landlinger and co-workers focused on the timing of the forehand – a common tennis stroke with a player's dominant hand that relies heavily on the rotation of the torso or trunk and positioning to generate racket speed – and analyzing differences among different skill levels (2). They concluded that the timing of the maximum angular pelvis and trunk rotations – essentially how much a player rotates their body when swinging - was a key difference between the elite group and the high-performance players. Their findings recommended that coaches focus more on improving the timing of the pelvis and trunk rotation of players to refine their forehand stroke further. A 2017 paper by Filipcic *et al.* also focused on timing in tennis but looked specifically at the timing of the split step – a brief hopping movement that tennis players use to re-center their balance in order to react to the movement of the ball (3, 2). They found that the players' response times of their split step changed significantly depending on the game situation.

While previous analyses used physical markers or basic computer vision to identify the ball and body locations, more recently, researchers have started using machine-learning-based pose estimation, which allows for more accurate and detailed analysis of limb dynamics in tennis videos. One of the first uses of pose estimation in tennis was in 2018, when Kurose et al. analyzed the probability of shot success based on machine-learning-generated pose estimates (4). Baily et al. recently looked at the distinguishing stroke differences between professionals and amateurs also using pose estimation (5). While this study showed that pros, semi-pros, and amateurs differ significantly from each other, it also focused on what factors exactly make those differences using numerical data derived from videos of players hitting serves (5).

RESULTS

The objective of our research was to find whether the timing of hitting the ball in a serve is a key differentiator between the skill level of amateurs, semi-professionals, and professionals by looking at the overall movement of each of the players' serves.

For the following data, we analyzed the overall service motion of three groups of players - pro, semi-pro, and amateur (10 professional players, 10 semi-professional players, and three amateur players). To start, we looked at two different skill-leveled players' overall movements by tracking a specific limb's x and y coordinates throughout a serve motion (**Figure 1**). It became apparent that the overall movements of players were similar regardless of skill level, but what made the difference was the timing at which the player made contact with the ball within the swing. The overall movements of players' serves after adjusting the timing were, in fact, very similar despite the differences in the skill levels (**Figure 2**).

This overall movement between different players, regardless of skill level, was similar but a key difference between skill levels was the timing of hitting the ball.

After collecting data from each individual group, we compared results in different combinations to identify any significant differences. These timing differences for all the pairs of serves were grouped into six different sets: Pro2Pro, Pro2Ama, Pro2Semi, Ama2Ama, Semi2Semi, and Ama2Semi. For instance, when the timing difference of Amateur #1 and Professional #8 are compared, the data went under the Pro2Ama group, and so on. In total, there were 15 possible combinations for each, which is every possible comparison involving the three skill level groups (pro, semi-pro, amateur). For example, we observed a significant difference between the arm limb positions of amateurs and semi-pros (pairwise Welch's *t*-test, p-value < 0.05). The conclusion was that all positioning and movement of individual limbs are significant for some pairings (p-value < 0.05), and our pairwise Welch's *t*-tests found that each combination of skill levels had many limbs with significant differences (p-value < 0.003). This gives good support that from timing differences alone, significant differences can be used to determine a player's skill level.

We also compared the sum of absolute contact timing differences across all limbs and both x- and y-coordinates for different skill levels to each other and summarized the distribution (**Figure 3**). For example, the summary of the timing differences of all professionals to each other is referred to as Pro2Pro. The Semi2Semi group compares semi-professionals to semi-professionals, and finally, the Ama2Ama group represents the comparison between an amateur and other amateurs (**Figure 3**). Comparing Pro2Pro, Semi2Semi, and Ama2Ama, the timing difference is smallest between a professional and other professionals, followed by a semi-professional compared to other semi-professionals (**Figure 3**). The contact timing difference is greatest amongst amateurs compared to each other. The mean ranged in each of the categories, with the Ama2Ama mean being the greatest (mean = 168.90), Pro2Pro mean being the smallest in value (mean = 122. 46), and Semi2Semi mean being the middle value between Ama2Ama and Pro2Pro (mean = 149.45). As smaller mean values represent more similarity among players, this intuitive result supports our approach as it showed more consistency at higher skill levels. Although the smallest standard deviation (std) was not Pro2Pro, as it was Semi2Semi's (std = 60.40), the greatest value for standard deviation was Ama2Ama (std = 74.97), which showed a consistent result with the mean values.

Finally, we examined the efficacy of our approach using these swing timing distinctions to classify players of unknown skill levels. In one example where an unknown player whose real skill level is a professional, comparison to an amateur shows a distribution of the right elbow vertical timing centered around five frames of difference. The unknown player's comparison to semi-pro shows a distribution centered around ten frames of difference with an additional outlier before

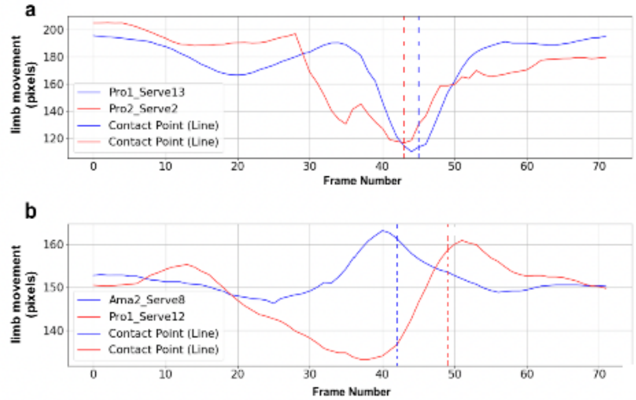

Figure 1. Pose graphs for different skill level players. a) Movement of the right elbow in the vertical direction compared between two pro players during a serve before alignment. The blue dotted line represents the vertical position of the right elbow for Pro1 while the red dotted line represents the same for Pro2. The vertical dashed lines represent where the player's racket makes contact with the ball during the serve. **b)** Movement of the left shoulder in the vertical direction for one pro player and one amateur player during a serve before alignment. (Ama2 and Pro1 comparison) Sizes are normalized so that each player is 70 pixels in width and 100 pixels in height. The vertical dashed lines represent where the player's racket makes contact with the ball during the serve.

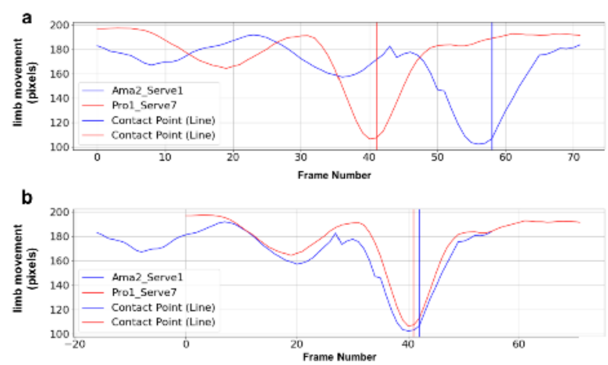

Figure 2. Pose graphs comparing players before and after alignment. a) Movement of the right elbow in the vertical direction for one pro player and one amateur player during a serve before alignment (Pro1 and Ama2 comparison). The vertical lines represent where the player's racket makes contact with the ball during the serve. **b)** the same movements after alignment. The difference between the amateur (blue) ball contact point in the original and shifted graphs represents the time difference we measure (Pro1 and Ama2 comparison). Sizes are normalized so that each player is 70 pixels in width and 100 pixels in height. The vertical lines represent where the player's racket makes contact with the ball during the serve.

30 frames. We observe the best agreement between the unknown skill level player and the pro distribution, so our approach classifies the unknown player as a professional based on their serve (**Figure 4c**). Three histograms of the right elbow's y-axis timing difference between an unknown player and amateurs, an unknown player to pros, and an unknown player to semi-pros were created (**Figure 5**). Additionally, the histogram of the three groups of three histograms was

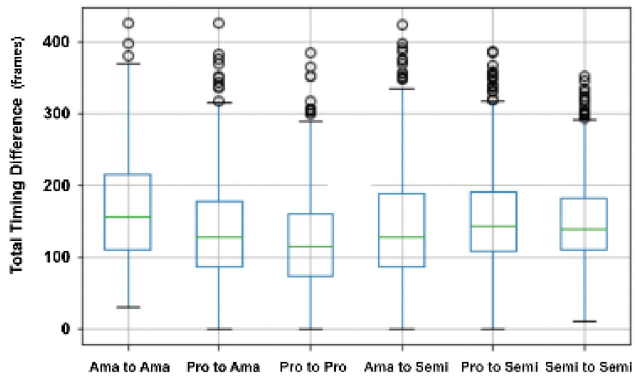

Figure 3. Distribution of the summed time differences for every limb for each comparison between skill level groups. The green bar represents the median of the total timing differences. The upper and lower quartiles of the data are represented by the box. The whisker lines extending out from the box show the variability outside the upper and lower quartiles excluding the circles representing the outliers.

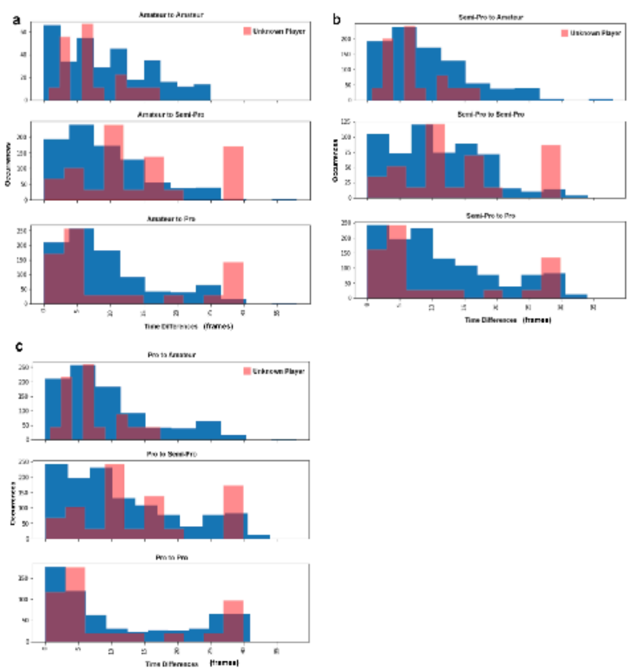

Figure 4. Overlapping Histograms of Two Different Comparisons Histogram comparison by overlapping graphs for the vertical axis of the right elbow. Red histograms represent the unknown professional player. The blue histograms represent the time differences comparisons to amateur, semi-pro, and pro players for a) amateurs, b) semi-pros, and c) pros.

produced, which helped us identify whether the unknown player in Figure 5 is a pro, a semi-pro, or an amateur (**Figure 4a-c**). By comparing the histograms of the unknown player's distribution versus different skill levels and comparing that to the collection of histograms of identified skill levels, it should be able to determine the unknown player's skill level (**Figure 4a-c and 6**). If the player is a pro, we expect the unknown player's histograms to be similar to the histogram patterns of known skill levels for professionals (**Figure 4c and 6**). If the player is a semi-pro, the unknown player's histogram should be like the histograms of a semi-pro, and finally, if the player is an amateur, the unknown player's histogram should be similar to the three histograms corresponding to that of amateurs (**Figure 4a-b**). In the case of the unknown player for the right elbow y-axis shown, the pattern best matches with the histograms of a professional, which corresponds to the unknown player being a professional (**Figure 5**).

Our classification has correctly identified the real skill level of all unknown players, which includes professional, semi-professional, and amateur players (**Table 1**). In addition, the distribution of the results is clear, as the players with the real skill level of pro have the most pros as their estimation, and the other skill levels have the greatest number of classified skilled levels as their corresponding level.

DISCUSSION

From our research, we were able to conclude that the overall movement of the serve is generally similar regardless of skill level, but that one of the deciding factors of determining a player's skill is the timing of when the player hits the ball. The findings in our research support our hypothesis and show that ball strike timing in a tennis serve is a key differentiator between skill levels. There are some variations in the collected numbers in the dataset, but those variations are fairly small. For instance, with the 8th unknown player – who is actually a pro – there were 43 fully detected histograms, with one histogram that was given a 0.5, that was in the correct category of Pro, which had a percent accuracy of 90.62%. Semi-pros were the trickiest, some semi-professionals can be on a level close to a pro or close to an amateur. The real skill level was correctly identified for all the players regardless of level. By identifying the specific patterns in the histograms, it is possible to differentiate between the three skill levels and analyze which category new players would fit into using just that information. This methodology was used by repeating the classification at least four times per player to ensure accuracy in our process. We also showed that from using a video of a tennis serve and the resultant pairwise timing differences compared to other players, we can determine the skill level of an unknown player. Our approach can be used to classify tennis players into skill levels just based on their swing video and timing difference compared to other players but cannot show an easy-to-interpret qualitative difference.

When looking at the results, we can see that the patterns of Pro1 and Pro2 are similar, which makes intuitive sense as it is expected for players of the same skill level to have relatively similar patterns (**Figure 1a**). Not only is the overall movement similar, but the point at which the players' limbs both dip is alike. However, even the pose graph of players of different skill levels, amateur and professional, displays a very closely related pattern. They both reach their peaks at around 160 pixels, and after the peak, both have a similar slope down

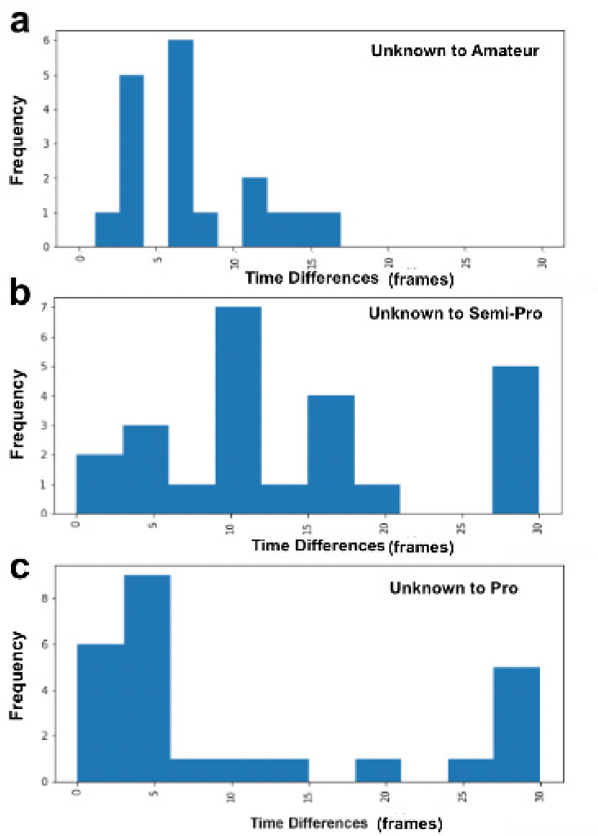

Figure 5. Time Difference Histograms. Histogram of the time differences of one unknown player's right elbow in the vertical direction when compared to our three skill level groups of known players.

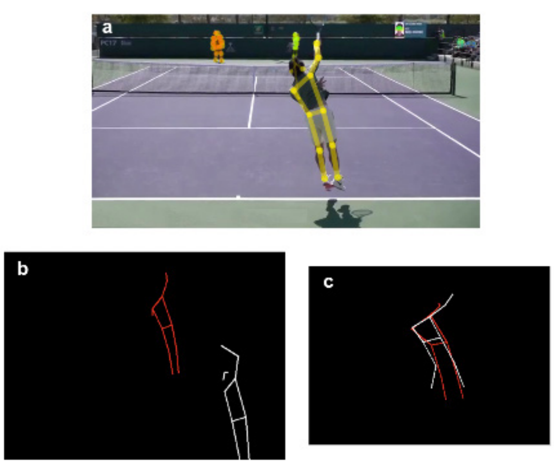

Figure 6. Steps of the Video Analysis. a) Still image of a video with AlphaPose tracking figures overlayed on professional tennis players. Visualization of the comparison between two detected poses from two different videos **b)** before normalization and **c)** after normalization and alignment.

to around 150 pixels (**Figure 1b**). The overall movements are similar, but it is also evident that the difference lies in the timing of hitting the ball. For instance, the contact point for Ama2 is before the peak, but Pro1's contact point is after the peak, meaning the vertical lines indicating the contact point do not align (**Figure 1b**). The alignment of these graphs is done by matching the overall pattern of the graphs with the highest cross-correlation (**Figure 2**). With the alignment, it is possible to calculate the difference between the placement of the contact points. This provides evidence for our hypothesis that the timing of hitting the ball in a serve is what makes the difference between skill levels, while the overall movement of the serve is not notably different.

There were some limitations to our research. One limitation was the limited amount of data collected for the research. Although there were several players per skill group and multiple serves per player, it was still limited and increasing our data size would have further improved our accuracy in our results. One shortcoming of our analysis was that, since we looked at the absolute timing difference rather than the relative timing difference, we could not determine whether players of one skill tend to make contact with the ball either before or after another group. Another factor is that the classification of unknown players to skill levels was manually done by analyzing histograms, which could have caused some bias in interpretation, as we knew beforehand which category the players would fit under. Finding a way to either automate this analysis or assign this portion of the analysis to researchers without prior knowledge would reduce that potential for bias. Most histograms were easy to classify while others were ambiguous, and by only having two people making those estimations of the results there still could have been some inaccuracy.

Finally, the method of cross-correlation of pose graphs between players was not consistent for individual players. Most of the pose graphs were graphs with a fairly accurate cross-correlation processing, and with those, the calculation of the difference in the contact point frames was accurate (**Figure 1a-b and 2a-b**). However, some graphs were not as accurate when maximizing the cross-correlation of the two graphs. This was one of the reasons we created histograms that displayed the total data taken from the pose graphs.

Some potential future implementations of our research may include using our data to create an application that could even suggest how players can change their timing to improve their skill in a more effective way, as it is often hard to recognize the changes players need to make to improve. This could give personalized advice to each player. Another application could be to automate the skill detection component, as our classification of unknown players research was done manually. By automating skill detection, the process will become more time-efficient and consistent and can be used by anyone without having to go through complex steps. This would require the setup (camera angles, camera quality) to be consistent with our approach, but that setup phase is well-documented and not particularly complex.

It is obvious that professionals, amateurs, and semi-professional tennis players are all significantly different in

Player	Real Skill Level of Player	# of Amateur to Player Skill Matches	# of Semi-Pro to Player Skill Matches	# of Pro to Player Skill Matches	Detected Skill Level	% Accuracy
Unknown 1	Semi-Pro	11	19.5	5	Semi-Pro	54.9
Unknown 2	Semi-Pro	3	47	5.5	Semi-Pro	84.7
Unknown 3	Semi-Pro	2	39.5	4.3	Semi-Pro	82.6
Unknown 4	Ama	30	7	2	Ama	76.9
Unknown 5	Ama	39	5	3	Ama	83.0
Unknown 6	Pro	4	6	29	Pro	74.4
Unknown 7	Pro	6	8.5	32.5	Pro	69.2
Unknown 8	Pro	1	3.5	43.5	Pro	90.6

Table 1. Results of the classification of unknown players with their real and detected skill levels along with the positive histogram group match results for each skill level. The numbers in each of the amateur, semi-pro, pro columns are the number of histograms that fit each category (0.5 if the histogram fit into two different skill categories). All *p*-values were significant with the *alpha* value of 0.05.

Limbs (x or y)	*p*-value
Right shoulder x distance	$p < 0.0001$
Right shoulder y distance	$p < 0.0001$
Left shoulder x distance	$p < 0.0001$
Left shoulder y distance	$p < 0.0001$
Right elbow x distance	$p < 0.0001$
Right elbow y distance	$p < 0.0001$
Left elbow x distance	$p < 0.0001$
Left elbow y distance	$p < 0.0001$
Right wrist x distance	$p < 0.0001$
Right wrist y distance	$p < 0.0001$
Left wrist x distance	$p < 0.0001$
Left wrist y distance	$p < 0.0001$
Right hip x distance	$p < 0.0001$
Right hip y distance	$p < 0.0001$
Left hip x distance	$p < 0.0001$
Left hip y distance	$p < 0.0001$

Table 2. ANOVA test for each of the limbs for x and y. Asterisk added for all significant values (n = 4489 comparisons from 67 videos).

skill, but we were able to demonstrate that the contact time of the serve is an important predictor of skill level. Additionally, our data suggest that the overall swing movement is similar among all groups regardless of their level when excluding timing differences, which has not been previously reported. Our paper contributes to our still-growing understanding of sports analytics, even adding some relevance to previous work such as that of Landlinger *et al.* (1). Landlinger *et al.* mention the importance of the timing, specifically in the forehand stroke, while we were able to conclude in a similar sense that the timing is also important for a serve (1). If both forehands and serves are largely determined by the contact timing, all tennis players can examine this sport in a new light.

MATERIALS AND METHODS

In our research, a total of 23 pre-recorded videos of tennis players (10 professional players, ten semi-professional players, and three amateur players) hitting serves were collected, all taken from the back and with similar angles. Most professionals and semi-professionals, both male and female, were taken from YouTube using practice and tournament videos, while amateur videos were taken by recording three junior players and one coach in Tokyo, Japan, with each participant's consent. The majority of the players collected were right-handed players. The professionals included players such as Roger Federer, Maria Sharapova, Kei Nishikori, Rafael Nadal, Fernando Verdasco, Ugo Humbert, and Petra Kvitová. Professionals were classified as higher-ranked players according to their standing in the world rankings (Association of Tennis Professionals (ATP) for men, Women's Tennis Association (WTA) for women), while the semi-professionals were players who are ranked but on the lower end (outside of the top 100).

First, each of the videos was turned into a collection of individual frames. 72 frames were manually chosen, including one full motion of a serve; the same number of frames was used for all players throughout this work. After repeating this process four to five times per player for all videos, a pose estimation algorithm was used called AlphaPose, an open-source system created by Fang *et al.* (7). AlphaPose detects people from a video and creates a file with all the horizontal and vertical, x- and y-, respectively, coordinates of each of the key limb joints, such as the elbows or the shoulders, using an additional algorithm called PoseFlow, created by Xiu *et al.* (7). A wireframe overlay is drawn onto each frame of a video by AlphaPose (**Figure 6a**). Repeating this process over all the frames per serve for all players turned the entire stroke into numerical data. Even though the videos collected were taken from similar angles, the players in the videos still had to be normalized by a fixed scale in both the x- and y-direction. Using the normalized data points, pose graphs representing a serve's overall movement were created in horizontal and vertical dimensions. The pose graphs were a way in which different players' swing motions could be compared, which was done by overlapping the graphs of two chosen players.

The similarities or differences were measured by how much one pose graph needed to shift to match the pattern of the other graph as closely as possible. The pairwise differences between players were analyzed and sorted into groups based on the skill level of the two players being compared. To quantify the results, each classification was given a number. If the group of histograms for one player did not clearly belong to one category (pro, semi-pro, amateur) but could only be narrowed down to one of two categories, a 0.5 rating was assigned for the two closest categories. If the group of histograms fits under one category, it was given a 1 rating. The data used in this classification were all significant in the *t*-tests. Histograms of the timing differences between two players were created for all possible combinations in the data set. The histograms then were used to classify 8 unknown players, whose videos were recorded and data collected using the identical methodology to the initial 23 videos, into one of the three skill level groups (professional, semi-professional, amateur).

Normalization was used to scale the players' figures up or down to make the sizes of all the same, as sizes differed despite them all being taken from a similar angle (**Figure 6b**). A scale in both the x- and y-directions was applied using the shoulder width and torso height, respectively. Additionally, the left-handed players' coordinates were flipped so that they overlapped with the right-handed players' figures. The stick figures after scaling in x- and y-coordinates and flipping on the vertical axis for left-handed players are almost the same in size after the normalization making the comparison ever easier and more accurate (**Figure 6c**).

From the normalized x- and y-coordinates created by the AlphaPose algorithm, unshifted pose graphs that represent the overall movement flow of the player's serve, in separate horizontal and vertical dimensions, were created (**Figure 1, 2 and 3**). This data was manually collected by first establishing when each player made ball contact. Then, the x- or y-coordinate movements of a certain limb in relation to the overall time were laid out. This way, the way in which a player moved their different limbs throughout their serve was determined.

Initially, the resulting plots of the two compared players were not aligned, so we aligned them by maximizing the cross-correlation of the two data series to align the swing patterns (**Figure 1a-b**). To find the highest correlation, all different combinations of the serves were tested. Essentially, this means that the timing of the two players was aligned to maximize the similarity of the swing pattern over the whole swing, leading to more overlap between the graphs (**Figure 2b**). It is important to note that aligning the two videos caused the number of frames that can be compared together to fall below 72 frames. This would have affected the accuracy of our results since our model was built on analyzing 72 total frames of movement. To solve this, the first and last value of the frames were extended out so that the other frames could be compared even with limited data after that point.

We then needed to determine a fixed parameter for the timing to standardize our analysis across data sets. First, the amount of timing alignment required was considered. Second, the timing for each player was set by using the time the ball hits the racket. Combining the two, the resulting difference in the point of ball contact time between two players was defined as the time difference between the two players during their swing.

The next focus was the pairwise timing differences of hitting the ball during the swing. For each pro, semi-pro, and amateur player, there were multiple serves analyzed for each player. There was a total of 116 serves in our sample set. A pairwise comparison of each serve to all other serves using the methodology described above was created for each serve.
For each serve pair, the absolute timing difference in ball strike was calculated between the two serves. Then, these timing differences for all the pairs of serves were grouped into six different sets: Pro2Pro, Pro2Ama, Pro2Semi, Ama2Ama, Semi2Semi, and Ama2Semi. For instance, when the timing difference of Amateur #1 and Professional #8 are compared, the data went under the Pro2Ama group, and so on.

From the pairwise timing differences of the ball strikes, histograms for the six different sets of comparisons were created. Each histogram showed how much the poses of the limbs had to be aligned to maximize the cross-correlation across all the serve pairs in each set. The specific patterns of each histogram were identified, whether it was the distribution, the shift in the data, or the locations of the peaks.

In an application of the generated distributions, the histograms were used to see if they can identify an unknown player's skill level based only on the video of the serve and the x-y position data collected from the videos. This method of cross-correlation was not perfect, and some graphs did not align as well as others, so taking the average lessened the variance in the data. An ANOVA test and pairwise Welch's *t*-tests with Bonferroni correction on python pandas were used to see if there are significant differences between the different pairwise skill level groups in each category of the limbs (**Table 2**).

Received: August 17, 2021
Accepted: February 10, 2022
Published: November 07, 2022

REFERENCES

1. "Tennis SmartCourt," PlaySight, 18 Nov. 2021, https://playsight.com/our-sports/tennis/.
2. Landlinger, Johannes, et al. "Kinematic differences of elite and high-performance tennis players in the cross court and down the line forehand." Sports Bio Mechanics, vol. 9, no. 4, Dec. 2010, pp. 280-295, doi:10.1080/14763141.2010.535841.
3. Filipcic, Ales, et al. "Split-Step Timing of Professional and Junior Tennis Players." Journal of Human Kinetics, vol. 55, Jan. 2017, pp. 97-105, doi:10.1515/hukin-2017-0009.

4. Kurose, Ryunosuke, et al. "Player pose analysis in tennis video based on pose estimation." 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, Jan. 7-9, 2018, pp. 1-4, doi:10.1109/IWAIT.2018.8369762.
5. Baily, Lisa, et al. "Stroke Comparison between Professional Tennis Players and Amateur Players using Advanced Computer Vision." Proceedings of the 8th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2020), Online Streaming Conference, Nov. 5-6, 2020, pp. 44-52, doi:10.5220/0010145800440052.
6. Fang, Hao-Shu, et al. "RMPE: Regional Multi-person Pose Estimation," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2353-2362, doi:10.1109/ICCV.2017.256.
7. Reno, Vito, et al. "Convolutional Neural Networks Based Ball Detection in Tennis Games." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, Jun. 18-22, 2018, pp. 1871-1877, doi:10.1109/CVPRW.2018.00228.
8. Xiu, Yuliang, et al. "Pose Flow: Efficient Online Pose Tracking," British Machine Vision Conference 2018 (BMVC), Newcastle upon Tyne, United Kingdom, Sep. 3-6, 2018, pp. 53.