



What Can Hawk-Eye Data Reveal about Serve Performance in Tennis?

François Rioult, Sami Mecheri, Bruno Mantel, François Kauffmann, Nicolas Benguigui

► To cite this version:

François Rioult, Sami Mecheri, Bruno Mantel, François Kauffmann, Nicolas Benguigui. What Can Hawk-Eye Data Reveal about Serve Performance in Tennis?. MLSA15 - Machine Learning and Data Mining for Sports Analytics workshop (ECML / PKDD 2015), Sep 2015, Porto, Portugal. pp.36-45. hal-01636404

HAL Id: hal-01636404

<https://hal.science/hal-01636404>

Submitted on 16 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

What Can Hawk-Eye Data Reveal about Serve Performance in Tennis?

François Rioult¹, Sami Mecheri², Bruno Mantel², François Kauffmann³, and
Nicolas Benguigui²

¹ CNRS UMR6072 GREYC - Normandy University, F-14032 Caen, France

² EA4260 CesamS - Normandy University, F-14032 Caen, France

³ CNRS UMR6139 LMNO - Normandy University, F-14032 Caen, France

Abstract. In the present study, we aim at showing how some characteristics of the serve summed up in the resulting ball trajectory can determine the efficiency of tennis serves. To that purpose, we analyzed a big set of data collected between 2003 and 2008 at international ATP, WTA and Grand Slam tournaments and corresponding to 84 tournaments, 1729 matches, 262,596 points. Using time-dependent three-dimensional ball trajectory data recorded by the automated ball tracking Hawk-Eye system, we show the relationships that exists between the characteristics of the serve kinematics and impacts on the ground on the gain of the points.

1 Introduction

In recent years, the development of technologies and automatic tracking systems has enabled the capture of ball trajectories during tennis matches. Since 2003, Hawk-Eye vision-based systems have provided ball tracking to assist the players when they think an error of judgments has been made by referees. This system uses a motion capture system with 10 cameras around the court and sophisticated algorithms will calculate the trajectories and impact on the ground of the tennis ball with an accuracy estimated of 3.6 mm at impact.

Despite the high-level accuracy of such tracking systems and the huge amount of kinematic data generated, the use of these systems for quantitative analysis of player performance and scientific analysis is rare and has never been performed on a very big sheer volume of data. To our knowledge, only three studies used Hawk-Eye data to analyze performance. These studies were interested in prediction of shot locations ([1], data volume = matches from the Australian Open men's draw or around 10 000 points), in laterality effect on ball distribution ([2], data volume = 32 matches or 4744 points) and in on-court position effect on groundstroke anticipation ([3], data volume = 38 matches, number of points unspecified).

In the present study, we aim at showing how some characteristics of the serve summed up in the resulting ball trajectory can determine the efficiency of tennis serves. To that purpose, we analyzed a big set of data collected between

2003 and 2008 international at ATP, WTA and Grand Slam tournaments and corresponding to 84 tournaments, 1729 matches, 262,596 points.

The influence of factors such as serve speed, serve location, court-surface and men/women differences on the winning-point rate was assessed in order to provide an extensive insight into efficient serve tendencies in world-class tennis. The positions of serves' impact were also examined in order to provide an accurate description of the serves performed by world-class players during matches. Since the present work is the first to exploit large-scale Hawk-Eye data, a subsidiary objective in these analyses was to demonstrate our method as reliable to analyze serving match strategies by confronting our findings to knowledge emanating from tennis performance analysis studies [4, 5].

We also focused on the unexplored question of the magnus effect intensity in serve trajectories. The spinning of the tennis ball was characterized in the present study directly from kinematic data by the ball axis of rotation and the speed of rotation around this axis. Specifically, the lift coefficient (as an indicator of spin intensity) and the ball axis of rotation (as an indicator of spin nature) were analyzed.

2 Data description

The data analyzed in the present research were made available by the company Hawk-Eye Innovations in the context of a publicly funded research project (TennisServer, ANR-06-BLAN-0413) in which one of us was involved in 2006-2009. For the moment being, the data are not publicly available.

The final stages of the most famous tournaments of the ATP and WTA circuits between 2003 and 2008 are covered by the data. 40 Hz trajectory of the ball and XML information about the points are available. Each file is named after the number of the set, the number of the game, the index of the point, the index of the serve (first or second, there is no file in case of double fault), and the time of the point.

For each point, the XML file gathers the following information (see Figure 1 for an excerpt):

1. the header gives overall information about the point: the server, the receiver, the player who is located on the positive part of the court, the class of the serve (0 for an ace, 1 for a classical one, 2 for a winning serve), the scorer of the point (1 if he/she is the server, -1 otherwise), the duration of the point (in seconds), and the score in the game at the start of the point;
2. the precise coordinates of the serve: who serves, the initial speed, the final speed, the coordinates of the initial impact (at $t = 0$), the coordinates of the bounce;
3. the precise coordinates of each shot.

After cleaning the data, there remains 75,587 points for the women and 187,009 for the men (total: 262,596 points).

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<point valid="true">
  <hawkeye_header>
    <xmldata d="Data"/>
    <server p="CLIJSTERS"/>
    <receiver p="HENINHARDENNE"/>
    <positive p="HENINHARDENNE"/>
    <serve_class c="1"/>
    <scorer s="-1"/>
    <PointDuration w="6.26786"/>
    <score_raw s="1 0"/>
  </hawkeye_header>
  <serve name="CLIJSTERS" player="1" speed="46.46" speedEnd="31.87">
    <coord t="0" x="1.47" y="-11.89" z="2.70"/>
    <coord bounce="true" t="0.49025" x="-3.05" y="6.17" z="0.033"/>
  </serve>
  <shot speed="31.3742" speedEnd="19.8407">
    <coord t="0.8" x="-4.15587" y="11.766" z="1.06128"/>
    <coord bounce="true" t="1.57191" x="0.24" y="-6.26" z="0.033"/>
  </shot>
  ...
</point>

```

Fig. 1: Sample of XML data.

3 Linear magnus model

In this section, we aimed to model the kinetics of a spinning tennis ball by estimating unknown parameters from reconstructed trajectories, using the R software [6]. Our analysis revealed that the Hawk-Eye reconstructed trajectories are using a third degree polynomial in each components (x, y, z) .

In contrast with previous studies which obtained spin rates by manually counting the number of revolutions of the ball from high-speed video cameras recordings (e.g., maximum serve spin rates values of 3529 rpm in Wimbledon qualifications reported by [7] and of 4300 rpm in Davis Cup reported by [8]), we used reconstructed ball trajectories and characterized the spinning of the tennis ball by its axis of rotation ω and the speed of rotation around this axis ω .

We use the model proposed in [9] to simulate the tennis ball trajectories. The tennis ball is considered as a mass point at position $\mathbf{X}(t) = (x(t), y(t), z(t))$ with mass m , diameter d and is influenced by three forces :

- the weight force $\mathbf{G} = m\mathbf{g}$ with $\mathbf{g} = (0, 0, -g)$
- the drag force $\mathbf{D} = -D_L(v, \omega) \frac{\mathbf{V}}{v}$ with $D_L(v, \omega) = C_D(v, \omega) \frac{1}{2} \frac{\pi d^2}{4} \rho v^2$
- the magnus force $\mathbf{M} = M_L(v, \omega) \frac{\omega}{\omega} \wedge \frac{\mathbf{V}}{v}$ with $M_L(v, \omega) = C_L(v, \omega) \frac{1}{2} \frac{\pi d^2}{4} \rho v^2$

We introduce physical characteristics of a reference tennis ball and atmospheric conditions:

- a reference diameter $d_0 = 0.067m$
- a reference density of the air $\rho_0 = 1.29$
- a reference mass $m_0 = 0.0577kg$

If we write $\alpha = \frac{\pi d^2 \rho}{8m}$ and $\alpha_0 = \frac{\pi d_0^2 \rho_0}{8m_0}$ then the equation becomes:

$$\frac{d^2 \mathbf{X}(t)}{dt^2} + 9.81 \mathbf{g} = C_G \mathbf{g} - C_D \frac{\alpha}{\alpha_0} (\alpha_0 v \mathbf{V}) + C_L \frac{\alpha}{\alpha_0} \frac{\boldsymbol{\omega}}{\omega} \wedge (\alpha_0 v \mathbf{V}).$$

The scalar coefficient C_G may be interpreted as a variation between normal gravity $9.81m/s^2$ and the observed gravity during the match. Variability of this coefficient may be due to latitude variation $\pm 0.03m/s^2$, gravitational anomalies, sampling frequency or model errors.

If we define the modified drag coefficient as $C'_d = C_D \frac{\alpha}{\alpha_0}$ and the modified lift coefficient as $C'_L = C_L \frac{\alpha}{\alpha_0}$ then the equation becomes:

$$\frac{d^2 \mathbf{X}(t)}{dt^2} + 9.81 \mathbf{g} = C_G \mathbf{g} + C'_D (-\alpha_0 v \mathbf{V}) + C'_L \frac{\boldsymbol{\omega}}{\omega} \wedge (\alpha_0 v \mathbf{V}) \quad (1)$$

The main assumptions of this magnus linear model is that the modified drag and lift coefficients are constant throughout an arc.

In Equation 1, the vectors $\frac{d^2 \mathbf{X}(t)}{dt^2} + 9.81 \mathbf{g}$ and $\alpha_0 v \mathbf{V}$ can be estimated with the model for speed and acceleration. The four unknown coefficients are $C_G, C'_D, C'_L \frac{\omega_x}{\omega}, C'_L \frac{\omega_y}{\omega}, C'_L \frac{\omega_z}{\omega}$, appear linearly in the equation and therefore may be estimated with a linear model.

Modified drag and lift coefficients C'_D, C'_L depend on properties of the roughness of the ball's surface, on velocity and on spinning. For a tennis ball which has the characteristic of the reference ball we have $C'_D = C_D$ and $C'_L = C_L$. The factor $\frac{\alpha}{\alpha_0}$ is a correction factor which only depends on the cross sectional area and the mass of the tennis ball in comparison to a reference tennis ball.

Alam [10] has estimated the drag coefficient C_D in the absence of any spin to lie between 0.5 and 1.2 for the tennis ball. At lower velocity, the mean value was 0.90 ± 0.15 , whereas at higher velocity the mean value was 0.6 ± 0.025 .

Goodwill [11] has studied the lift coefficient C_L of a tennis ball in a wind tunnel in different conditions, as a function of $S = \frac{d/2\omega}{v}$. Drag coefficients were varying from 0.65 ± 0.01 for low value of S i.e. $S \leq 0.3$ and raised to 0.69 ± 0.01 for higher S values. [11, 12, 13] found lift coefficient from 0.02 to 0.3.

We found that 80 % of points' trajectories had a global R^2 greater than 0.97, meaning that for this subset of points, the linear combination of these three estimated components $C_G \mathbf{g}, C'_D (\alpha_0 v \mathbf{V}), C'_L (\frac{\boldsymbol{\omega}}{\omega} \wedge (\alpha_0 v \mathbf{V}))$ provided a good approximation of $\frac{d^2 \mathbf{X}(t)}{dt^2} + 9.81 \mathbf{g}$.

4 Results

4.1 Winning probability for server

The results of Table 1 shed light on the fact that the serve is a redoubtable shot for winning points in tennis. It provided servers with the opportunity to accumulate a high percentage of winning points, particularly from the first serve

(69.46%⁴). This advantage of the server over the receiver confirms the results of [5, 14] as well as [4] who reported 67.3% wins and 53.8% wins on second serves on clay (66.28% and 52.24% in the present study). Unsurprisingly, the court surface also had a significant influence on winning rate.

surface	serve	win	lose
CLAY	first serve	66.28	33.72
	second serve	52.24	47.76
GRASS	first serve	71.19	28.81
	second serve	54.85	45.15
HARD	first serve	68.34	31.66
	second serve	52.68	47.32
INDOORS	first serve	72.01	27.99
	second serve	53.03	46.97

gender	serve	win	lose
women	first serve	62.85	37.15
	second serve	49.43	50.57
men	first serve	71.00	29.00
	second serve	54.18	45.82

Table 1: Probability for the server to win the point as a function of serve rank (first, second) and court surface. Table 2: Probability for the server to win the point as a function of serve rank (first, second) and gender.

4.2 Impact of the gender

The analysis (see Table 2) revealed that men won significantly more points when serving than women both on first and second serves. Other research efforts have also noted gender differences in winning percentages on serve [15, 5, 16]. This result could be mainly explained by the difference in speed of serves across men and women.

4.3 Impact of serve speed

The results of Figure 1 indicate a significant effect of serve speed on winning percentage on serve. These results are in agreement with the findings of [15] who have noted a significant relationship between the serve speed and the probability of winning the point. They found that, serve speed was negatively correlated with the proportion of serves that fell inside the serve box. Also, the proportion of points won when the serve was in was positively correlated with the serve speed for both the first and second serves in Grand Slam tournaments [15]. Therefore, hitting a “hard” first serve is a winning serve strategy to win a high percentage of points [12]. This strategy increases the time constraints on receivers by reducing the time available for executing their shot.

⁴ this is the average of each probability to win on first serve over the surfaces in Table 1.

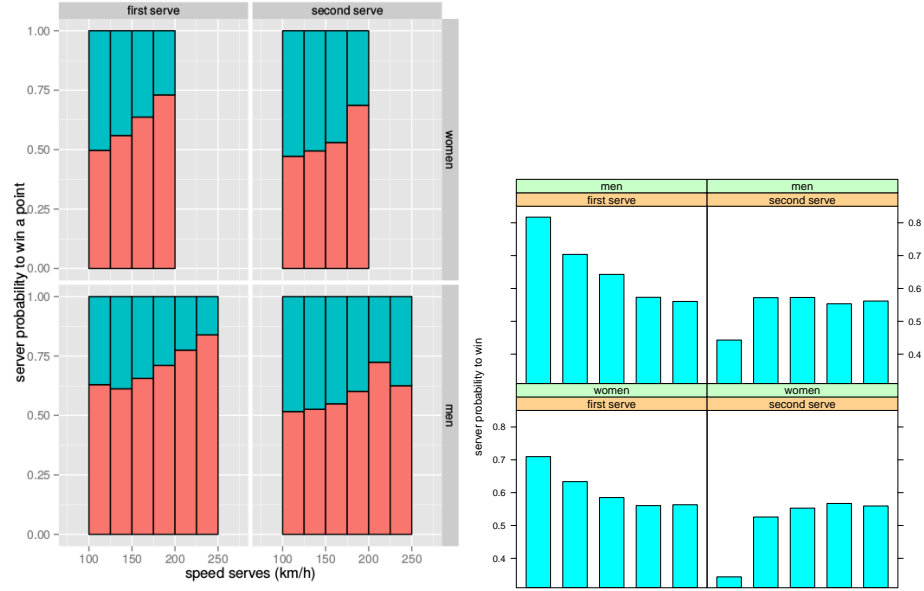


Fig. 2: Probability to win a point w.r.t. the speed of the serve (lose on top, win Fig. 3: Probability to win according to on bottom). the number of strokes.

4.4 Impact of the number of strokes

Results of Figure 2 showed a clear negative relationship between the number of recorded shots per rally and first serve's winning percentage in both men and women. These data can be summarized in the following way: the lower the number of strokes per point, the greater the impact of serve on winning the point. This makes sense: the severe spatio-temporal constraints imposed on the receiver when facing a first serve make it difficult to restore the balance in one or two groundstrokes. Consequently, points concluded within a very low number of shots result to outcomes that tend to favor the servers.

However, for second serves, the relationship between the number of recorded shots per rally and the serve winning percentage is much different since the $[0,2]$ category is linked to a very low winning rate (44% in men and 34% in women) while other categories all share similar winning rates (around 43% in men and 45% in women). Since the spatio-temporal constraints on the receiver are considerably lower for second serves, the receiver has the opportunity to initiate a baseline rally in almost all cases. In this configuration, if the point is gained quickly, it cannot be attributed to serve impact (whose speed is very moderate) but to the quality of the return. These results nevertheless remain surprising for their magnitude, and another explanation could be problems in the sub-sets of data used to perform these analyses.

4.5 Serves' impacts positions

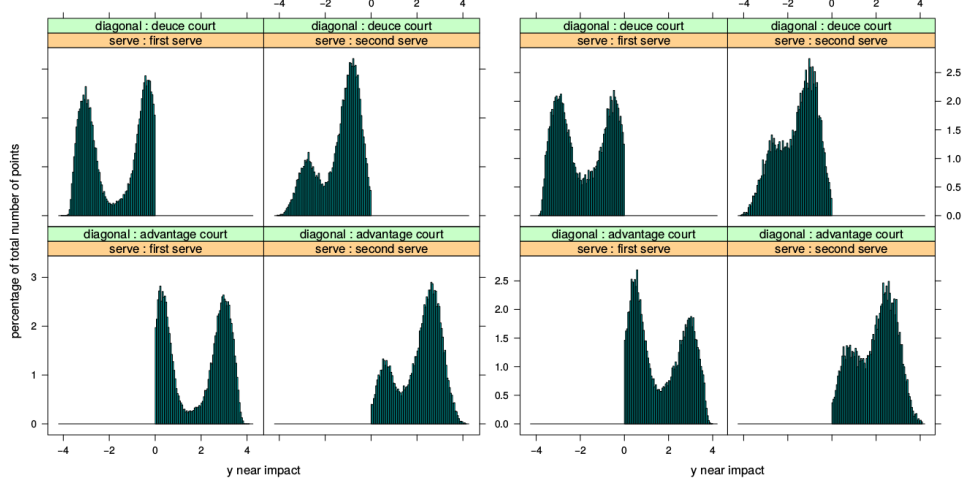


Fig. 4: Distribution of serves' impacts (y-axis) in service boxes for the men (left) and the women (right). Deuce court is located on the left of the server while the advantage court is on his/her right.

The distribution of serves' impacts positions along the y-axis⁵ (Figure 4) revealed that most first serves were directed toward the T (middle) and W (edge) locations in a very similar fashion on both deuce and advantage courts. A similar trend has already been reported in a previous work describing the serve locations of male professionals on hard courts [14]. The first serve is typically at pace [17], and one would expect first serves to be directed more often toward the W and T locations because serving to these locations takes the ball away from the receiver, making it difficult for him/her to return (which is the main goal of first serves [18]).

Interestingly, the distribution of serves' impacts positions along the y-axis revealed differences between deuce (mainly T) and advantage courts (mainly W) for second serves. This finding is in line with [4] which noted that close to 95% of second serves were directed either toward the T in the deuce court (48.0%) or to the wide zone in the advantage court (46%). In other words, on second serve on both sides of the court, professional tennis players serve to the corners of the service box with a specific focus on their opponents' backhand (most of which are right-handed opponents), which is usually considered the weaker side. Our data strongly confirmed that two strategies are employed on second serves, depending on the service box being played. When serving on the deuce side, servers attempt to push back the receivers with a topspin (as demonstrated by C'_L values) toward the T so as to keep them behind the baseline. When serving on the advantage side, players attempt to find more angles by serving wide and with topspin (as demonstrated by C'_L values) to open up the court. In both

⁵ the x-axis is oriented along the depth of the court, the y-axis is parallel to the net.

cases, the server’s intention is to dominate the rally from its start by exerting a territorial influence.

4.6 Lift coefficient and rotation axis

The analysis (see Figure 5) revealed clear differences in C'_L as a function of serve ball with C'_L values significantly higher for second serves. This result confirms and extends current knowledge about tennis. Indeed, [4] reported that for first serves, the flat option was the most used (55.7%) while for second serves spin variations were massively used (99.0%). First serve spins are employed to introduce tactical variations but with parsimony since it reduces serve’s speed. However, during second serve, the players’ goal is to limit aggressive and offensive returns. For this reason, as reported by [4], topspin strategy is classically used on the second serve (91.6%) to generate a shoulder or head-high and deep bounce, which prevents the receiver from executing an offensive stroke.

5 Conclusion

The present study has confirmed and extended knowledge about tennis duel by manipulating various performance indicators of the first stroke of each point and assessing its influence on winning-point probabilities. Having demonstrated the validity of our trajectory reconstruction method for tennis performance analysis by replicating the findings of earlier tennis performance studies ([4, 14, 15]), this method could be used to provide coaches and researchers with objective and massive information on serving performance. On this plan, the high proportion of first serves oriented to the two corners of the box revealed that players do not maximize the possibility of varying the direction of the serve. The above results could be made even more meaningful by incorporating serve variability indicators such as entropy to determine the location succession effect on serve winning rate. Future research on this plan is encouraged to disentangle the complexity of situational probability information that is integrated into decisions of expert players in serve-return.

Also, the direct method of spinning determination used in the present paper is highly valuable since it is applicable with no supplementary time costs to all players competing or just training on ball tracking equipped-courts. Obtaining spinning data by this way is desirable since it allows players and coaches to obtain accurate information about their stroke quality from matches and practice that are readily available by avoiding manual analysis which is time consuming. This approach based on 3d-ball tracking data not only offers further works for serve or serve-return performance, but might also help to add further knowledge about players’ fitness level and ground-stroke quality.

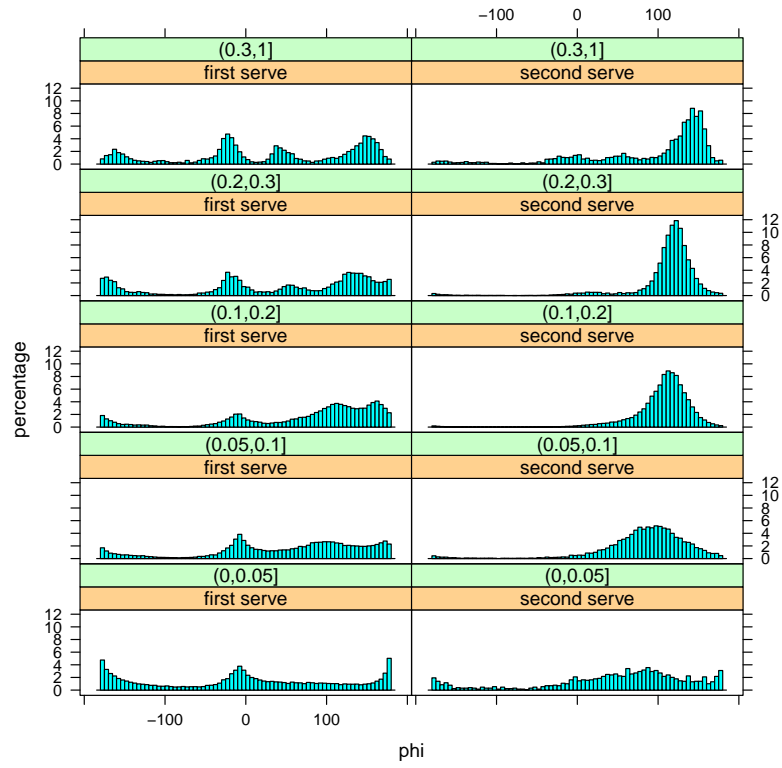


Fig. 5: Distribution of ϕ as a function of serve and modified lift coefficient C_L' for right handedness (ϕ is the angle of the rotation vector in the yz -plane, $\phi = 180^\circ$ means pure lift for right handed players and $\phi = 90^\circ$ means pure slice).

References

1. Wei, X., Lucey, P., Morgan, S., Sridharan, S.: Predicting shot locations in tennis using spatiotemporal data. In: Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on, IEEE (2013) 1–8
2. Loffing, F., Hagemann, N., Strauss, B.: The serve in professional men’s tennis: Effects of players’ handedness. *International Journal of Performance Analysis in Sport* **9**(2) (2009) 255–274
3. Loffing, F., Hagemann, N.: On-court position influences skilled tennis players’ anticipation of shot outcome. *Journal of sport & exercise psychology* **36**(1) (2014) 14–26
4. Gillet, E., Leroy, D., Thouvenecq, R., Stein, J.F.: A notational analysis of elite tennis serve and serve-return strategies on slow surface. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association* **23**(2) (2009) 532–539
5. O’Donoghue, P., Ingram, B.: A notational analysis of elite tennis strategy. *Journal of sports sciences* **19**(2) (2001) 107–115
6. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2015)
7. Kelley, J., Goodwill, S., Capel-Davies, J., Haake, S.: Ball spin generation at the 2007 Wimbledon qualifying tournament (P110). In: *The Engineering of Sport 7*. Springer (2008) 571–578
8. Goodwill, S., Douglas, J., Miller, S., Haake, S.: Measuring ball spin off a tennis racket. In: *The Engineering of Sport 6*. Springer (2006) 379–384
9. Klvana, F.: Trajectory of a Spinning Tennis Ball. In: *Solving Problems in Scientific Computing Using Maple and Matlab®*. Springer Berlin Heidelberg (1995) 15–23
10. Alam, F., Watkins, S., Subic, A.: The aerodynamic forces on a series of tennis balls. In: *15th Australasian Fluid Mechanics Conference*. (2004)
11. Goodwill, S.R., Chin, S.B., Haake, S.J.: Aerodynamics of spinning and non-spinning tennis balls. *Journal of Wind Engineering and Industrial Aerodynamics* **92**(11) (September 2004) 935–958
12. Brody, H., Cross, R., Lindsey, C.: *The Physics and Technology of Tennis*. Racquet Tech Pub. (2002)
13. Cross, R., Lindsey, C.: Measurements of drag and lift on tennis balls in flight. *Sports Engineering* **17**(2) (November 2013) 89–96
14. Unierzyski, P., Wieczorek, A.: Comparison of tactical solutions and game patterns in the finals of two grand slam tournaments in tennis. *Science and Racket Sports III* (2004) 200–205
15. O’Donoghue, P., Ballantyne, A.: The impact of speed of service in Grand Slam singles tennis. In: *Science and Racket Sports III: The Proceedings of the Eighth International Table Tennis Federation Sports Science Congress and The Third World Congress of Science and Racket Sports*, Routledge (2004) 212
16. Verlinden, M., Ruyskensvelde, J.V., Gorp, B.V., Decker, S.D., Goossens, R., Clarijs, J.P.: Effect of Gender and Tennis Court Surface Properties Upon Strategy in Elite Singles. In: *Leeds, A., Kahn, J.F., Maynard, I.W., eds.: Science and Racket Sports III*. Routledge, London (2004) 163–168
17. Magnus, J.R., Klaassen, F.J.G.M.: On the advantage of serving first in a tennis set: four years at Wimbledon. *Journal of the Royal Statistical Society: Series D (The Statistician)* **48**(2) (1999) 247–256
18. Crespo, M., Miley, D.: *Advanced Coaches Manual*. International Tennis Federation, ITF Ltd, London (1998)