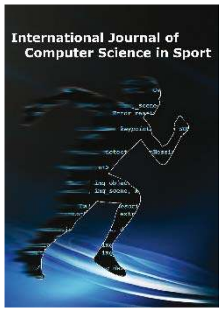
International Journal of Computer Science in Sport

Volume 18, Issue 2 (Special Edition), 2019

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2019-0013



Validation Of Sensor-Based Game Analysis Tools In Tennis

Edelmann-Nusser, A.[1], Raschke, A.[2], Bentz, A.[2], Montenbruck, S.[2], Edelmann-Nusser, J.[3], Lames, M.[2]

[1]Steinbeis Research Center Sports Technology

[2]Technical University of Munich

[3]Otto-von-Guericke University of Magdeburg

Abstract

Three inertial measurement unit (IMU) based tennis sensor systems from BABOLAT (PURE DRIVE PLAY, POP) and HEAD (Tennis Sensor) and a camera-based system (PlaySight) were tested with respect to the question whether the information about the number of strokes by swing type and spin type in training exercises and/or matches and the average as well as the maximum speed of the service per session are reliable. Subsequently, the question whether the mechanical properties of the BABOLAT PURE DRIVE PLAY racket are the same as the mechanical properties of the BABOLAT PURE DRIVE racket without IMU was addressed.

For swing types in standard exercises the results are acceptable for forehand groundstrokes, backhand groundstrokes and services but not for volleys. In a match environment we find inacceptably high errors (>10%) for the number of strokes for forehand and completely inacceptable levels for volley. The wrist-based IMU of BABOLAT POP has not reached an acceptable accuracy at all. For spin types the results are acceptable. The large variances in service speed assessment between devices make it doubtful whether any of them may be used for the control of training processes aiming at increasing the average service speed The mechanical properties of the BABOLAT rackets with and without IMU are quite the same.

KEYWORDS: TENNIS, SENSOR, IMU, VALIDATION, GOLD STANDARD

Introduction

Technological progress leads to the introduction of new game analysis tools or tools to track performance for top-level as well as for leisure sports. For example in tennis, IMU-based technologies were introduced in the consumer market especially in the last five years. The validation of new innovative products is a permanent task of the interdisciplinary field of information technology in sports, because top-level sports as well as leisure sports demand for reliable data. This task has been neglected in the past, because providers do not see a real benefit and customers obviously do not ask for validated products. So, to the best of our knowledge, there is no independent study on the properties of the new tennis diagnostics tools under investigation[1].

At present, different providers offer IMU-based sensor systems for tennis:

- The racket BABOLAT PURE DRIVE PLAY (BABOLAT, 2018a) has an built-in sensor in the racket handle at the butt end (see Figure 1). Due to the fact that it happens quite often during a match that players have to change the racket because of a cracked or broken string or change the racket for worn-off strings, it is necessary that all rackets in use have the same weight, swingweight and recoilweight. Therefore, BABOLAT PURE DRIVE PLAY should have the same mechanical characteristics (weight, swingweight) as the racket BABOLAT PURE DRIVE without a sensor. This is important for customers, as they only have to buy one (more expensive) racket equipped with a sensor.
- BABOLAT POP (BABOLAT, 2018b) has its IMU integrated in a wristband, it is not fixed to the racket. This concept makes it independent of the racket used. The additional weight of BABOLAT POP is about 8g sensor plus 10g wristband for fixing.
- HEAD (HEAD, 2018) offers the HEAD Tennis Sensor powered by ZEPP (see Figure 1). To fix the sensor to the racket at the butt end, a part of the cap of the handle is replaced by the sensor. Therefore, the sensor can be used with many of Head's rackets, but the weight increases slightly (the weight of the sensor is 7g).
- The Smart Tennis Sensor from SONY (SONY, 2018) is similar to the HEAD Tennis Sensor. To fix the sensor to the racket at the butt end, a part of the cap of the handle is replaced by the sensor and a ring adaptor. The sensor can be used with many rackets of different manufacturers. It increases the weight by about 8g.
- ZEPP (ZEPP, 2018) itself offers a sensor that supports many different rackets by the use of three different mounting systems that attach the sensor to the butt end of the handle of the racket. The sensor increases the weight for at least 7g

Manufacturers claim that their sensors provide information about the number of strokes by swing type (forehand, backhand, volley, serve), by spin type (topspin, slice, flat) in training sessions or matches and average as well as maximum speed per session. Beyond that, all three systems with IMUs integrated in the racket handle provide information on the hitting spot on the racket that was not part of this validation. A first approach to validate the hitting spots is suggested by Keany and Reid (2018), who used a VICON-system.

[1] While writing this paper, a study of Keany and Reid (2018) was published. They validated ZEPP and BABOLAT PLAY using a VICON system analysing 24 prescribed strokes and counting stroke types of 150 strokes in a match setting.

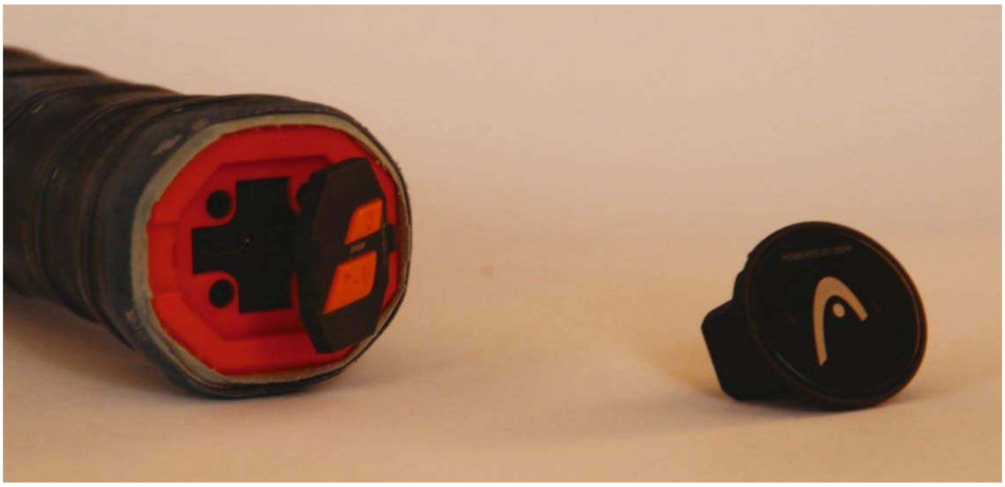


Figure 1. Left side: Handle of a BABOLAT PURE DRIVE PLAY racket with built in sensor.
Right side: HEAD TENNIS SENSOR

Representing a different technological approach for tennis diagnostics, a camera-based analysis tool is on the market, also: PlaySight (PlaySight, 2018). With six to eight fixed cameras ball and players' movements on court are tracked and analysed via image analysis. Swing type, ball speed, spin, in/out are registered for each stroke. For analysis, rallies containing certain events may be retrieved together with any of the six/eight camera perspectives.

The first aim of this paper is to evaluate the IMU-based BABOLAT PURE DRIVE PLAY racket with regard to the question whether the mechanical characteristics are the same as the characteristics of the BABOLAT PURE DRIVE racket. The second aim is to check for the reliability of sensor-based analysis tools. This question is addressed by two studies. A first one, conducted at Otto-von-Guericke University of Magdeburg (OVGU study), investigated whether swing types and spin types are identified correctly by BABOLAT PURE DRIVE PLAY. The second study at Technical University of Munich (TUM study) addressed the reliability of the number of strokes in different swing types and agreement on service speed between four systems: BABOLAT PURE DRIVE PLAY, HEAD Tennis Sensor, BABOLAT POP, and PlaySight.

Methods

Mechanical characteristics

We measured the weight of one BABOLAT PURE DRIVE PLAY racket and one BABOLAT PURE DRIVE racket both with same strings, same racket grips and same overgrips using an electronic scale (Kern Präzisionswaage PCB 1000-2, max. weighing range 1kg, accuracy 20mg).

The swingweight and recoilweight of the two rackets were calculated according to Brody (2002) using equation (1). Equation (1) gives the swingweight in kg*cm². The recoilweight in kg*cm² is calculated according to equation (2) using the parallel-axis theorem. To get the balance points and recoilweight axis of the rackets, a wooden triangle profile was used. Figure 2 shows the geometrical parameters and the experimental setup to quantify the swingweight of a racket: The racket is supported on the top cross string and the time for 60 oscillations was measured.

$$swingweight = m * \left(\frac{P * T^2}{40.28} - \frac{P^2}{1{,}000} + \frac{Y^2}{1{,}000}\right) \quad (1)$$

m: mass of the racket in kg
P: distance in cm from the balance point or recoilweight axis to the rotation axis at the bottom edge of the top cross string used to support the racket (see Figure 2)
Y: The distance BP -10.16cm in cm form the balance point or recoilweight axis to the swingweight axis 10.16cm (4 inches) from the butt end of the handle (see Figure 2)
T: time in seconds for on oscillation

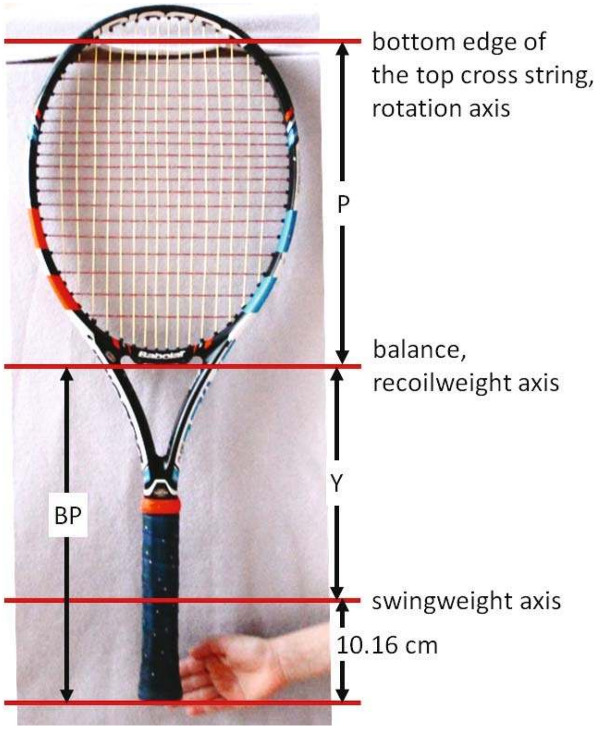


Figure 2. Experimental setup to quantify the swingweight and recoilweight according to Brody (2002) . The racket is supported vertically by resting top cross string on two horizontal rods about 15 cm apart. Pushing the handle allows the racket to swing back and forth as a pendulum. To get the time T of one oscillation, the time for 60 oscillations was measured. To compute the swingweight and recoilweight the geometric parameters P and BP must be measured.

$$recoilweight = swingweight - \left(\frac{m * Y^2}{1{,}000}\right) \quad (2)$$

m: mass of the racket in kg
Y: The distance BP-10.16cm in cm form the balance point or recoilweight axis to the swing-weight axis 10.16cm (4 inches) from the butt end of the handle (see Figure 2)

Methods OVGU study

Sample, exercises, data aquisition

Four skilled tournament tennis players (age[yrs]/sex[m/f]/weight[kg]/height[m]: Player 1: 13/m/52/1.63, player 2: 16/f/61/1.68, player 3: 52/m/82/1.80, player 4: 53/m/76/1.75), playing in top provincial amateur leagues, accomplished five tasks.

1. 20 forehand groundstrokes from the baseline (another player passed the ball to the player from the opposite baseline center mark)

2. 20 backhand groundstrokes from the baseline (another player passed the ball to the player from the opposite baseline center mark)
3. 20 volleys from a distance of about 3m to the net (another player passed the ball to the player from the opposite baseline center mark)
4. 12 services
5. Five minutes of playing against one of the other players, starting with a passed ball from one player from the opposite basline center mark to the other player. The aim was not to score but to control the ball and keep the ball in play.

Each task of each player was recorded by the BABOLAT PURE DRIVE PLAY racket as one separate session and was additionally recorded by video for validation purposes. All players made a total of 598 strokes.

Game analysis, data evaluation

For the first four tasks the real swing types and numbers of strokes in the video were compared with the swing types and numbers of strokes that were acquired by the BABOLAT PURE DRIVE PLAY racket. In task 5 the numbers of forehand strokes, backhand strokes, volleys, strokes with topspin, strokes with slice and flat strokes were counted from video recordings and compared to the BABOLAT PURE DRIVE PLAY racket data. Tasks 1-3 (exercises consisting of a known number of known stroke types) allowed a differentiation between the following errors:

1. number of mismatched strokes (e.g. a forehand stroke is tracked as a backhand stroke)
2. number of missing strokes (i.e. less strokes tracked than really done)
3. number of phantom strokes (i.e. more strokes tracked than really done)

The counts of these different numbers were sampled for each player and then summed up for the four players. So, the overall results may contain phantom strokes as well as missing ones.

Methods TUM Study

Sample, exercises, data aquisition

8 tournament players (10-18 yrs; 4 female, 4 male) from Bavarian Tennis Federation played four matches in best-of-three mode. During each match one player used a BABOLAT PURE DRIVE PLAY racket and wore BABOLAT POP, the other player use a racket equipped with the HEAD Tennis Sensor. All matches took place at a court equipped with PlaySight. A total of 2,098 strokes were subject to comparisons.

PlaySight's 8 video cameras served as footage for the gold standard, video-based observation. Percent agreement between two independent observers was 98.4% (n=499 strokes)

Game analysis, data evaluation

The number of strokes per swing type (Service, Forehand, Backhand, Volley) was counted per game and player. Comparisons were conducted on match base and overall. Percent agreement with the gold standard was reported.

The average and the maximum service speed reported by the four systems is given per match and player and the results of the IMU-tools were set into relation with PlaySight's results.

Statistical procedure

As the IMU-based tools offer statistics only for one complete session, reliability cannot be assessed on a stroke-by-stroke basis as the methodological standard for testing instrumental consistency would require (Lames, 1994). The alternative, taking each stroke as a training session, is not feasible as this would inflate the number of sessions and require a manipulation of the tools during a rally. As a consequence, only summative information of a session may be compared giving rise to possible overestimation of agreement, because of possible compensation of mismatches.

Therefore, in the TUM study and in task 5 of the OVGU study we can not identify the real number of mismatched strokes, we can only identify the number of at least mismatched strokes. The following example explains the definition of “at least mismatched strokes” in task 5 (rallies without service for 5 min): Assume that in one session 50 forehand strokes and 40 backhand strokes were done, 5 forehand strokes are mismatched as 5 backhand strokes and 3 backhand strokes are mismatched as 3 forehand strokes. 1 forehand stroke is not tracked. This results in 50 - 5 + 3 - 1 = 47 tracked forehand strokes and 40 + 5 – 3 = 42 tracked backhand strokes. The total number of tracked strokes is 47 + 42= 89, one stroke less than really done. I.e. one stroke is missing, this missing stroke is a forehand stroke. Because 50 forehand strokes were done and 47 were tracked and one stroke is missing two forehand strokes are at least mismatched (50 = 47 tracked +1 missing + 2 at least mismatched). At least two backhand strokes are mismatched because 42 were tracked but only 40 were done. Totally 2+2=4 strokes are at least mismatched, one stroke is missing.

Results

Mechanical characteristics

Table 1 shows the mechanical characteristics and geometric parameters of the BABOLAT PURE DRIVE PLAY racket compared to the BABOLAT PURE DRIVE racket. The differences are very small and do not have an effect on playing tennis: It can be assumed that the effect of different string tensions switching from one racket to a freshly restinged one in match has much more impact on the "feeling" of the rackets.

Table 1. Mechanical characteristics and geometric parameters of the BABOLAT PURE DRIVE PLAY racket and the BABOLAT PURE DRIVE racket.

	BABOLAT PURE DRIVE PLAY	BABOLAT PURE DRIVE
Mass	321 g	323 g
Time for 60 oscillations	82.15 s	82.05 s
T (one oscillation)	1.369 s	1.368 s
P	32.0 cm	32.1 cm
Y	22.5 cm	22.4 cm
swingweight	312.2 kg*cm²	311.5 kg*cm²
recoilweight	149.2 kg*cm²	148.9 kg*cm²

Results OVGU study

Figure 3a shows the results for the tasks 1 to 3. For each swing type the strokes were summed up for all four players. The results for the forehand and backhand strokes are acceptable: 78 of 80 forehand strokes are tracked, only two strokes are mismatched. In the backhand stroke task 80 backhand strokes were tracked, but three of these strokes were phantom strokes of one player (23 backhand strokes were tracked although he did only 20) and three of the 80 strokes are mismatched strokes of another player (17 backhand strokes and three forehand strokes were tracked although he did 20 backhand strokes), i.e. the two errors cancel each other. From the 80 volleys only 50 were tracked as volleys, 26 were mismatched to forehand or backhand strokes and 4 strokes are missing.

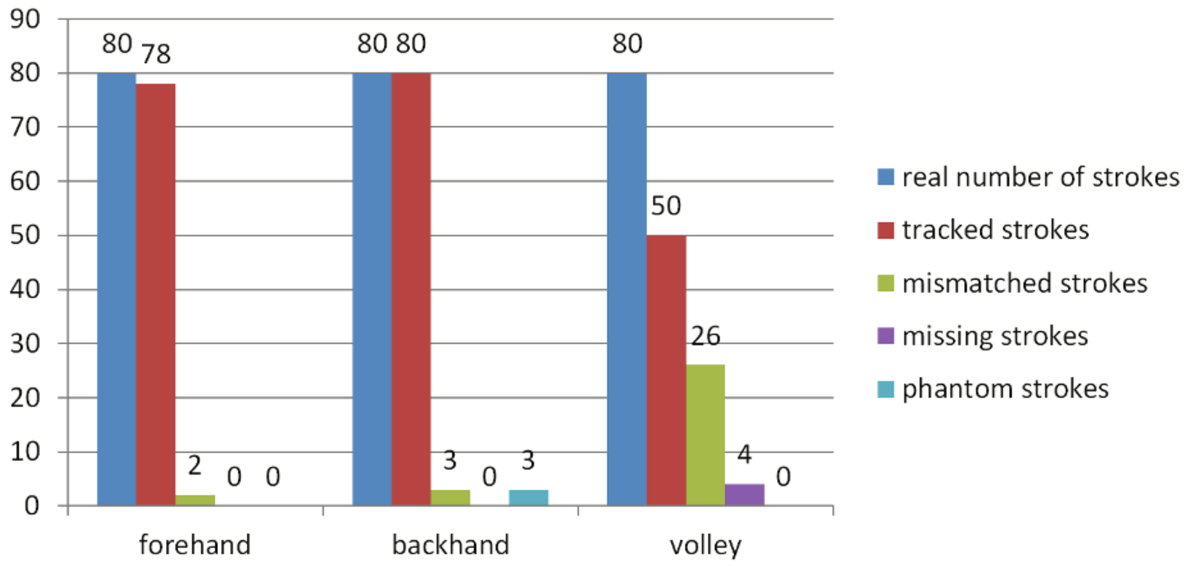


Figure 3a. Results for forehand groundstrokes, backhand groundstrokes and volleys (tasks 1 – 3) summed up for all four players.

In task 4 (12 services) all 12 services of all four players were tracked.

Figure 3b shows the results for task 5. The results are similar to the ones in Figure 3a: They are acceptable for the forehand and backhand strokes but there is a large number of errors in the volleys: 13 volley were played, 9 volleys were tracked, but 16 errors (8+6+2, see Figure 3b) appeared. I.e. the number of errors is larger than the number of volleys the four player did (see Table 2). Further on, only three of the tracked volleys were truly volleys as Table 2 shows. The other six tracked volleys (player no. 4 in Table 2) were mismatched or phantom strokes.

Table 2. Results for the volleys of the five minutes of playing against one of the other players (task 5). The last column shows the line totals. Note that player no. 4 did not play any volley but six volleys were tracked.

Player	1	2	3	4	Sum
Real number of volleys	6	5	2	0	13
Tracked number of volleys	1	2	0	6	9
At least mismatched number of volleys	1	3	0	4	8
Missing number of volleys	4	0	2	0	6
Phantom volleys	0	0	0	2	2

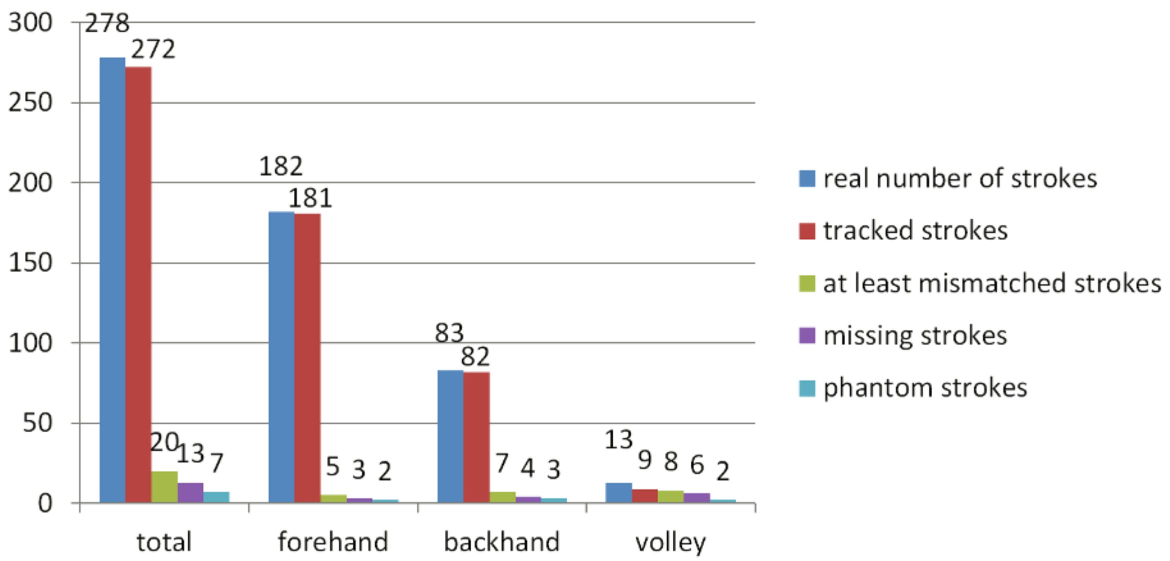


Figure 3b. Results for the swing types of the five minutes of playing against one of the other players (task 5) summed up for all four players.

The results for the spin types (see Figure 3c) show acceptable values: 278 strokes were done, 272 were tracked and the number of errors was at least (104-97) + (35-33) + (141-140) = 10 respectively 3.6%.

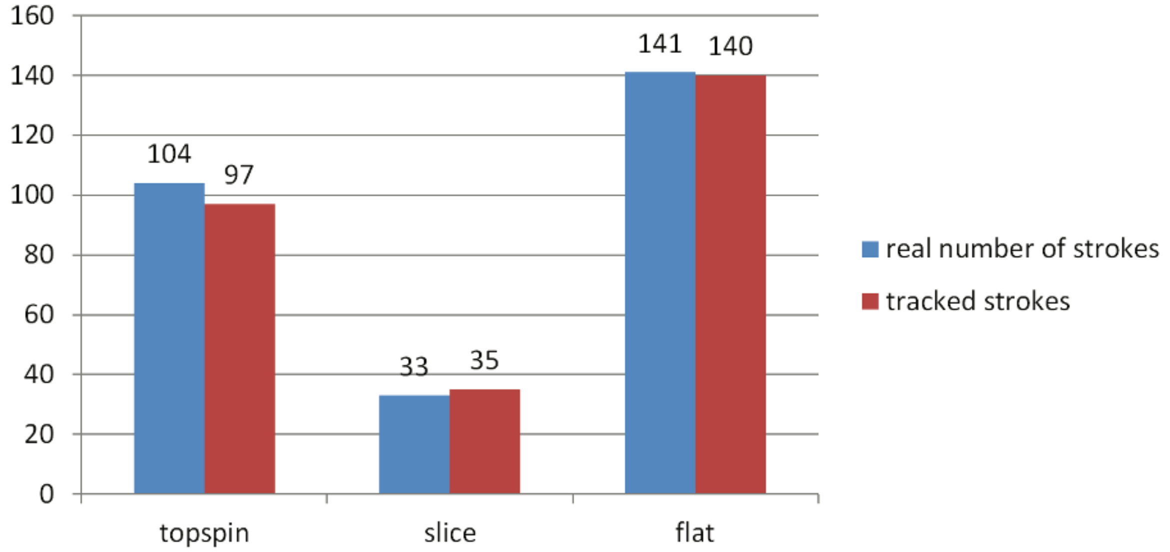


Figure 3c. Results for the spin types of the five minutes of playing against one of the other players (task 5) summed up for all four players.

Results TUM Study

Table 3 shows the results for the number of strokes counted in different stroke classes between the three sensor-based systems and the video-based PlaySight system. These numbers are set in relation to the number of strokes obtained from video-based observation serving here as a gold standard. As mentioned, only summative counts per session are available. Their deviations from the gold standard are calculated and used for specifying the summative percent agreement.

The number of services is recorded with an acceptable error <5% except for BABOLAT POP. The same holds for backhands, although on a slightly lower level (error between 5 and 10%)

with BABOLAT POP performing much better. Forehand strokes do not reach an agreement of 90% with any system. All systems have huge problems in identifying volleys.

Table 3. Number of strokes obtained from PLAYSIGHT, BABOLAT PURE DRIVE PLAY, BABOLAT POP and HEAD Tennis Sensor compared to the gold standard (GS), video-based observation. Dev is the deviation of the systems from GS, %Agr is Percent agreement: 100-100*Dev/GS. Explanation: GS for PlaySight is the sum of GS for BABOLAT devices plus HEAD, because latter devices were worn each by one player per match.

	PlaySight			**BABOLAT PURE**			**BABOLAT POP**			**HEAD Tennis Sensor**		
	GS	**Dev**	**% Agr**	**GS**	**Dev**	**% Agr**	**GS**	**Dev**	**% Agr**	**GS**	**Dev**	**% Agr**
Service	484	22	95.5	249	11	95.6	249	75	69.9	235	7	97.0
Forehand	867	103	88.1	434	74	83.0	434	92	78.8	433	83	80.8
Backhand	618	40	93.5	310	19	93.9	310	33	89.4	308	30	90.3
Volley	21	13	38.1	13	40	-	13	9	30.8	8	94	-

Table 4 shows the results for service speed measurements. Percent deviations of the sensor-based systems are given here with respect to PlaySight, the camera-based system. We find a large variation for average service speed between the systems, differences are ranging from 0 to 19.6%, and even more for maximum speed, where they are ranging from 5.4 to 23.0%.

Table 4. Average (Ø) and maximum (max) service speed measured with Playsight, BABOLAT PURE DRIVE PLAY, BABOLAT POP and HEAD Tennis Sensor in three matches. Comparison is made by percent deviation (%) from PlaySight results.

		PlaySight		**BABOLAT PURE**				**BABOLAT POP**				**HEAD Tennis Sensor**			
		Ø	**max**	**Ø**	**%**	**max**	**%**	**Ø**	**%**	**max**	**%**	**Ø**	**%**	**max**	**%**
Match 1	Player 1	124	136	124	0.0	165	17.6	109	12.1	165	21.3				
	Player 2	135	175									126	6.7	153	12.6
Match 2	Player 1	138	183	128	7.2	161	12.0	134	2.9	158	13.7				
	Player 2	113	135									117	3.5	139	23.0
Match 3	Player 1	107	130	86	19.6	123	5.4	90	15.9	116	10.8				
	Player 2	95	120									92	3.4	109	9.0

The Bland-Altman plot in Figure 4 illustrates the findings. X-values are the means between the results of PlaySight and the respective device tested. The mean difference for all measurements of average speed (Mean Avg) is 7.7 km/h, which amounts to 6.6% of all recorded values for average speed. The mean error for all measurements of maximum speed (Mean Max) is 2.7 km/h, which is only 1.9% of the overall mean value for maximum speed.

As can be seen already in table 4 the Bland-Altman plot also shows that the large variation of the differences is a bigger problem than average deviation. The width of the 95% confidence intervall for average speed is 29.6 km/h. This amounts to 25.6% of the overall mean value of all recorded average speed values. For maximum speed these figures are 66.7 km/ for the width of the 95% confidence interval corresponding to 46.3% of the overall mean value of recorded maximum speeds.

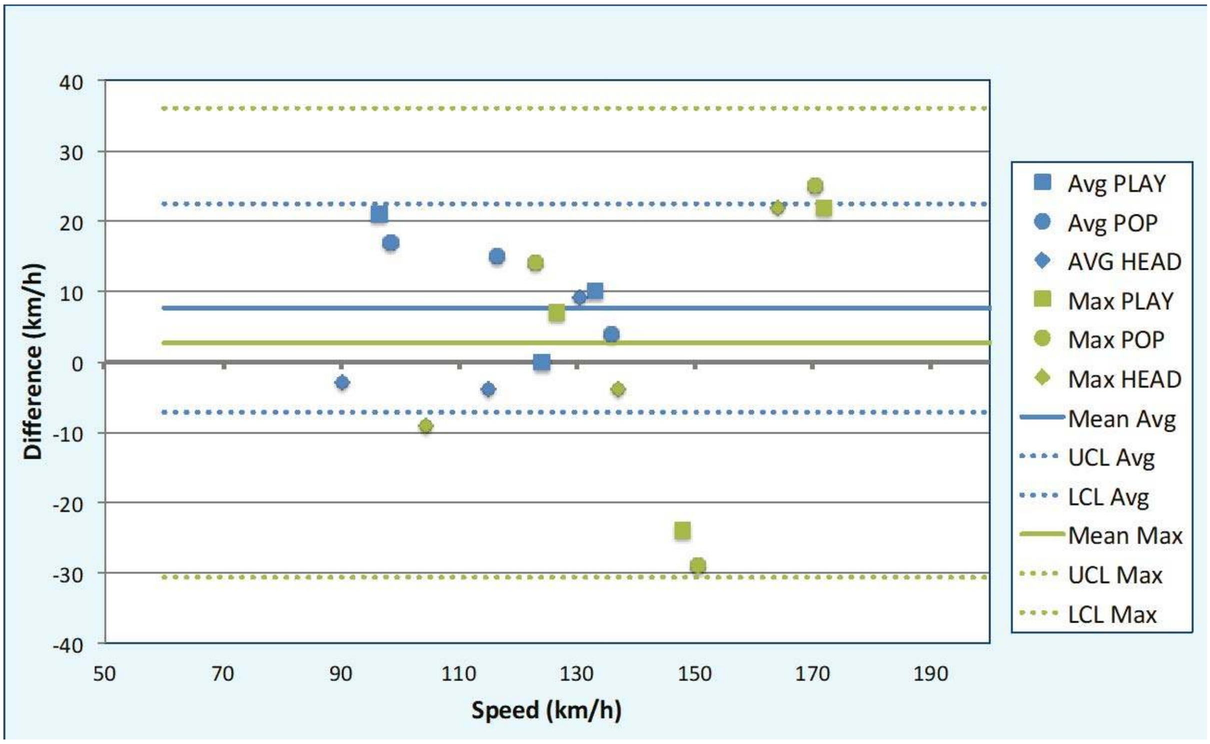


Fig. 4. Bland-Altman plot for average (Avg) and maximum (Max) service speed. The three IMU-based devices are compared to PlaySight, thus the x-value is the mean between PlaySight and the respective tested device. Horizontal lines give mean differences for average and maximum speed as well as respective upper and lower confidence interval.

Discussion

There are no remarkable differences in the mechanical properties: Neither the mass nor the swingweight or recoilweight of the BABOLAT PURE DRIVE PLAY racket are increased compared to the BABOLAT PURE DRIVE racket without an IMU.

In the standard exercise sessions of the OVGU study (tasks 1-4) the results for the swing types are acceptable for forehand groundstrokes, backhand groundstrokes and services but not for the volleys. In task 5, the task that is more similar to a match, the total number of real strokes and tracked strokes seems to be much better than it really is (see Figure 3b): Although the difference between real strokes (278) and tracked strokes (272) is only 6 respectively 2.1% the summed up number of errors is much larger with 40 (20+13+7, see Figure 3b) respectively 14.4%. This corresponds to the results of the TUM study that show that in a match environment we find inacceptable high errors (>10%) for the number of strokes for forehand and completely inacceptable levels for volley. The wrist-based IMU of BABOLAT POP has not reached an acceptable accuracy at all. Maybe this is due to missing/more inaccurate information on racket orientation of its wrist-based sensor compared to the racket-based ones of the competitors.

The large variances in service speed assessment between devices make it doubtful whether any of them may be used for the control of training processes aiming at increasing the average service speed, as improvements here may be assumed to be less than 10% at least for a practically relevant duration of an intervention of 4-8 weeks. This holds especially when strict criteria like the width of the 95% confidence interval for the error from Bland-Altman plot are applied.

Comparing OVGU and TUM underlines that the usage of sensor-based tools is more reliable in standard exercise settings compared to extended match settings. On the other hand, playing standardized exercises with controlled number and type of strokes is a comparatively easy condition with information on number and stroke type already given. They perform worse in match settings where they could provide additional diagnostic information.

A limitation of the studies - besides that only summative counts were available for stroke counts - was that no gold standard for speed measurements and hitting spot measurements was available leaving the validation of these parameters for the future with an appropriate design. Whereas for speed measurements, some (validated?) gold standards exist, this is still an open problem for hitting spots. The first approach by Keany and Reid (2018) with a VICON-system using 4 markers for the racket and 2 for the ball may only be seen as a first step in this direction.

Conclusions

The tested IMU-systems can not fulfill the promise of the providers that a tennis player can break down his strokes by swing type and provide a reliable report of his training and his matches for the different swing types. The situation seems to be better with spin types. The recorded average and maximum service speed values show large differences between the systems giving rise to doubts in their usefulness for practical training purposes.

These disappointing results on the accuracy of diagnostic devices for tennis underline the necessity of methodologically sound validation studies. As has become obvious, validation studies pose heavy problems, for example in finding appropriate gold standards or in applying the spectrum of appropriate test situations (exercises vs. matches). Moreover, it is important to conduct independent studies as manufacturers seem not to be interested in critical testing.

The demand for validation of diagnostic devices will hopefully increase in future, when in top level sports the awareness of existence and meaning of measurement errors will be more widespread and in leisure sports people will be less distracted by secondary functions of these devices, e.g. being symbol for a modern lifestyle or creating impact in social media.

References

BABOLAT. (2018a). Retrieved November 27, 2018, from http://en.babolatplay.com/play

BABOLAT. (2018b). Retrieved November 27, 2018, from http://en.babolatplay.com/pop

Brody, H., Cross, R. & Lindsey, C. (2002). *The Physics and Technology of Tennis*. Solana Beach, Calif. : Racquet Tech Pub.

HEAD. (2018). Retrieved November 27, 2018, from https://www.head.com/us-US/sensor/

Keany, E.M. & Reid, M. (2018). Quantifying hitting activity in tennis with racket sensors: new dawn or false dawn? *Sports Biomechanics*. https://doi.org/10.1080/14763141.2018.1535619.

Lames, M. (1994). *Systematische Spielbeobachtung* (Systematic Game Observation). Münster: Philippka.

PlaySight (2018). Retrieved December 1st, 2018, from https://playsight.com/our-sports/tennis/

SONY, 2018. Retrieved November 27, 2018, from https://www.sony.de/electronics/smart-geraete/sse-tn1w

ZEPP, 2018. Retrieved November 27, 2018, from http://www.zepp.com/en-us/